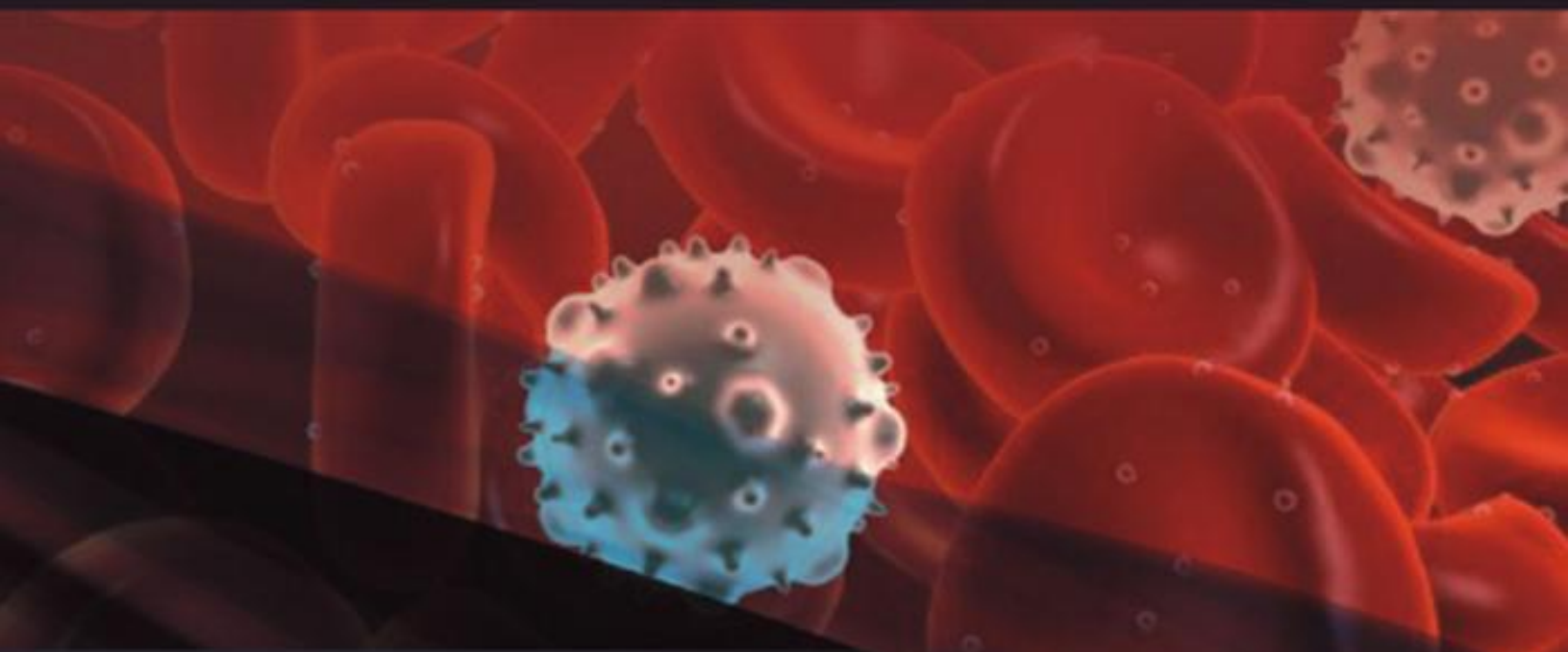
ELSEVIER

Medical Physiology

For Undergraduate Students



Indu Khurana

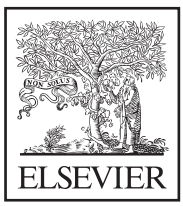
Medical Physiology

for Undergraduate Students

Medical Physiology

for Undergraduate Students

Indu Khurana
Senior Professor, Department of Physiology,
Postgraduate Institute of Medical Sciences,
University of Health Sciences,
Rohtak, India



ELSEVIER
A division of
Reed Elsevier India Private Limited
New Delhi

Medical Physiology for Undergraduate Students, 1/e
Indu Khurana

ELSEVIER
A division of
Reed Elsevier India Private Limited

Mosby, Saunders, Churchill Livingstone, Butterworth-Heinemann and Hanley & Belfus are the Health Science imprints of Elsevier.

ISBN: 978-81-312-2805-0

Published by Elsevier, a division of Reed Elsevier India Private Limited.

Registered Office: 622, Indraprakash Building, 21 Barakhamba Road, New Delhi-110 001.
Corporate Office: 14th Floor, Building No. 10B, DLF Cyber City, Phase-II, Gurgaon-122 002, Haryana, India.

Managing Editor: Shabina Nasim
Copy Editor: Shrayosee Dutta
Manager – Publishing Operations: Sunil Kumar
Manager – Production: NC Pant
Production Executive: Arvind Booni
Cover Designer: Raman Kumar

Typeset by Olympus Infotech Pvt. Ltd., Chennai, India (www.olympus.co.in).

Printed and bound at

To
the teachers and residents in physiology for their endeavour to dissipate and acquire knowledge

My parents and teachers for their blessings

My children, Aruj and Arushi, for their patience and tolerance shown to loss of many precious moments

My husband, Dr AK Khurana, for his understanding, encouragement and invaluable guidance

“This page intentionally left blank"

Preface

Medical Physiology for Undergraduate Students provides a thorough exposition of the text in such a balanced way that the undergraduate medical students can easily cover the syllabus during the one-year period provided for the first professional in the revised curriculum by Medical Council of India, and also find adequate material while preparing for various post-graduate entrance tests. The text on core and applied aspects of human physiology has been skilfully intermingled for the students to apply their learning in clinical situations.

The subject matter on various systems has been arranged into twelve sections and each section has been subdivided into various chapters. Metabolism and Nutrition has purposely not been included in this book; since, it is mainly covered in biochemistry books and there is no point in repeating the information here.

The *salient features* of this book which make it an indispensable tool for undergraduate medical students are:

- Each section begins with brief overview highlighting the topics covered. The text is then organized in such a way that the students can easily understand, retain and reproduce it.
- Various levels of headings, subheadings, bold face and italics given in the text will be helpful for a quick revision of the subject.
- Relevant functional anatomy given in the beginning of each chapter is quite useful in conceptualizing the subject.
- The text is illustrated with plenty of diagrams. The illustrations mostly include clear line diagrams providing vivid and lucid details.
- To further enhance the lucidity of the book each section is presented in a different colour format and the text and the figures are presented in four colour.
- In order to emphasize the clinical significance of physiology, the relevant applied aspect has been covered adequately in each chapter.
- Tables and flowcharts given in abundance will help in quick comprehension of the text.

No venture of this kind is possible without the blessings of teachers and parents to whom I shall remain ever indebted. I owe this arduous task to the love and patience of my children Aruj, Bhawna and Arushi. It would have not been possible for me to complete this task without the unconditioned help of my inspirator Dr AK Khurana, Sr Prof and Head, RIO, Postgraduate Institute of Medical Sciences (PGIMS), Rohtak, who has not only guided me during each and every step in the preparation of this book but also authored the chapter on 'Sense of Vision'.

I am extremely grateful to the faculty members and resident doctors of Department of Physiology, PGIMS, Rohtak for their generous help (especially Dr Manjeet and Dr Jai).

I wish to place on record my deep appreciation for my sincere friend Dr Sushma Sood, Sr Prof and Head, Department of Physiology, PGIMS, Rohtak for her invaluable guidance and support. I also acknowledge with gratitude the constant encouragement and conducive working atmosphere provided by Dr CS Dhull, Director, Pt BD Sharma, PGIMS, Rohtak and Dr SS Sangwan, Vice-chancellor, University of Health Sciences, Rohtak.

It is my special pleasure to acknowledge with gratitude the most assured, co-operation and skill from the staff of Elsevier, A division of Reed Elsevier India Pvt. Ltd., especially Ms Shabina Nasim, Managing Editor, and Shrayosee Dutta, Copy Editor.

For a volume like this it is not possible to be entirely free from human errors, some inaccuracies, ambiguities and typographical mistakes. Feedback and suggestions from the teachers and the students for further improvement of the book will be welcomed and dully acknowledged.

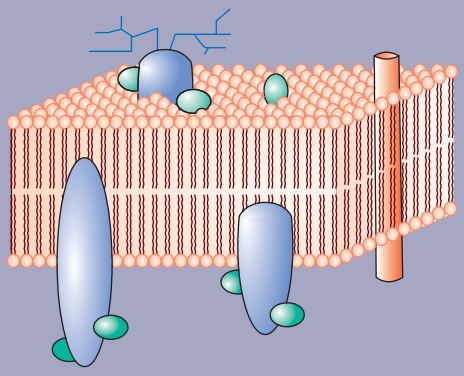
Indu Khurana

“This page intentionally left blank"

Contents

SECTION 1

General Physiology



Physiology, in simple terms, refers to the study of normal functioning of the living structures. The *human physiology* is concerned with the way the various systems of the human body function and the way each contributes to the functions of the body as a whole. In other words, the human physiology is concerned with specific characteristics and mechanisms of the human body that make it a living being and the mechanisms which help in adaptation and homeostasis which are the two fundamental features of life.

The *general physiology* envisages the general concepts and principles that are basic to the functions of all the systems. As we know, the fundamental unit of human body is a cell, therefore, this section includes a short review of fundamental aspect of the cell physiology. Before studying the general biophysiological processes and the cell physiology, it will be worthwhile to have a brief knowledge about the functional organization, composition and internal environment of the human body.

“This page intentionally left blank"

Chapter 1.1

Functional Organization, Composition and Internal Environment of Human Body

FUNCTIONAL ORGANIZATION OF THE HUMAN BODY

- Skin and its appendages
- Skeletal system
- Muscle system
- Nervous system
- Cardiovascular system
- Respiratory system
- Digestive system
- Excretory system
- Reproductive system
- Endocrine system
- Blood and immune system

BODY COMPOSITION

- **Total body water**
- **Body electrolytes**

INTERNAL ENVIRONMENT AND HOMEOSTASIS

- **Internal environment**
- **Homeostasis**

FUNCTIONAL ORGANIZATION OF THE HUMAN BODY

The human body is actually a social order of about 100 trillion cells organized into different functional structures, some of which are called organs, some organs combinedly form a system. For convenience of description, the human body can be considered to be functionally organized into various systems.

1. Skin and its appendages

Skin is the outermost covering of the human body. Its appendages include hairs, nails, sebaceous glands and sweat glands. The skin performs following important functions:

- It acts as a physical barrier against entry of microorganisms and other substances.
- It prevents loss of water from the body.
- It is a very important sensory organ containing receptors for touch and related sensations.
- It plays an important role in regulating body temperature.

2. Skeletal system

The basic framework of the body is provided by a large number of bones that collectively form the *skeleton*. At joints, the bones are united to each other by fibrous bands called *ligaments*. In addition to the bones and joints, the skeletal system also includes the *cartilages* present in the body.

3. Muscle system

Overlying and usually attached to the bones are various muscles. Muscles are composed of many elongated cells called *muscle fibres* which are able to contract and relax. Three distinct types of muscles can be identified which are skeletal muscles, smooth muscles and cardiac muscles.

4. Nervous system

The specialized cells that constitute the functional units of the nervous system are called neurons. The *nervous* system may be divided into: (i) the *central nervous system*, made up of brain and spinal cord and (ii) the *peripheral nervous system*, consisting of the peripheral nerves and the ganglia associated with them. The nerves supplying the body wall and limbs are often called *cerebrospinal* nerves. The nerves supplying the viscera, along with the parts of the brain and spinal cord related to them, constitute the *autonomic nervous system*. The autonomic nervous system is subdivided into two major parts: the *sympathetic* and the *parasympathetic* nervous system.

5. Cardiovascular system

The cardiovascular system consists of the *heart* and the *blood vessels*. The blood vessels that take blood from the

heart to various tissues are called *arteries.* The smallest arteries are called *arterioles.* Arterioles open into a network of *capillaries* that perfuse the tissues. Exchange of various substances between the blood and the tissues take place through the walls of capillaries. In some situations, capillaries are replaced by slightly different vessels called *sinusoids.* Blood from capillaries (or from sinusoids) is collected by small *venules* which join to form veins. The veins return blood to the heart.

6. Respiratory system

The respiratory system consists of the lungs and the passages through which air reaches them. The passages are nasal cavities, the pharynx, the trachea, the bronchi and their intrapulmonary continuations.

7. Digestive system

The digestive or the alimentary system includes all those structures that are concerned with eating, and with digestion and absorption of food. The system consists of an alimentary canal which includes the oral cavity, pharynx, oesophagus, stomach, small intestine and large intestine. Other structures included in the digestive system are the liver, the gall bladder and the pancreas.

8. Excretory system

Excretion is the removal of waste products of metabolism from the body. *Egestion* (or defaecation) is the removal of undigested food from the gut and is not regarded as excretion because the material taken into the gut through the mouth is not made by the body itself. The organs forming excretory system are the kidney, the ureters, the bladder and the urethra.

9. Reproductive system

Reproduction is the production of a new generation of individuals of the same species. It involves the transmission of genetic material from one generation to the next. *The male reproductive organs* are the testis, the epididymis, the ductus deferens, the seminal vesicles, the prostate, the male urethra and the penis. *The female reproductive organs* are the ovaries, uterine tubes, the uterus, the vagina, the external genitalia and the mammary glands.

10. Endocrine system

Endocrine tissue is made up essentially of cells that produce secretions which are poured directly into blood called *hormones.* Some organs are entirely endocrine in function. They are referred to as *endocrine glands* (or *ductless glands*) e.g. the hypophysis cerebri (pituitary gland), the pineal gland, the thyroid gland, the parathyroid glands and the suprarenal (adrenal) glands. Groups of endocrine cells may be present in the organs that have other functions. These include the *islets of Langerhans* of pancreas, the interstitial cells of the testis, the follicles and corpora lutea of the ovaries. Hormones are also produced by some cells in the kidney, the thymus and the placenta.

11. Blood and immune system

Blood is regarded as a modified connective tissue because the cellular elements in it are separated by a considerable amount of 'intercellular substance' and because some of the cells in it have close affinities to cells in general connective tissue.

Circulating blood normally contains three main types of cells which perform their respective physiologic functions: (i) the red cells (*erythrocytes*) are largely concerned with oxygen transport, (ii) the white cells (*leucocytes*) play various roles in the body defence against infection and tissue injury and (iii) platelets (*thrombocytes*) which are primarily involved in maintaining the integrity of blood vessels and in preventing blood loss. Detailed physiology of each organ system is considered in the relevant chapters.

BODY COMPOSITION

The normal body in an average adult male is composed of *water* (60%), *minerals* (7%), *protein* and related substances (18%), and *fat* (15%). The water, denoted by the term total body water (TBW), and the electrolytes need special emphasis.

TOTAL BODY WATER

Water is the principal and essential constituent of the human body. The total body water is about 10% less in a normal young adult female (average 50%) than that in an average adult male (60%) due to relatively greater amount of adipose tissue in the females. In both sexes the value tends to decrease with age.

THE BODY FLUID COMPARTMENTS

The total body water is distributed into two main compartments of the body fluids separated from each other by membranes freely permeable to water (Fig. 1.1-1, Table 1.1-1):

1. Intracellular fluid compartment

The intracellular fluid (ICF) compartment comprises about 40% of the body weight, the bulk of which is contained in the muscles.

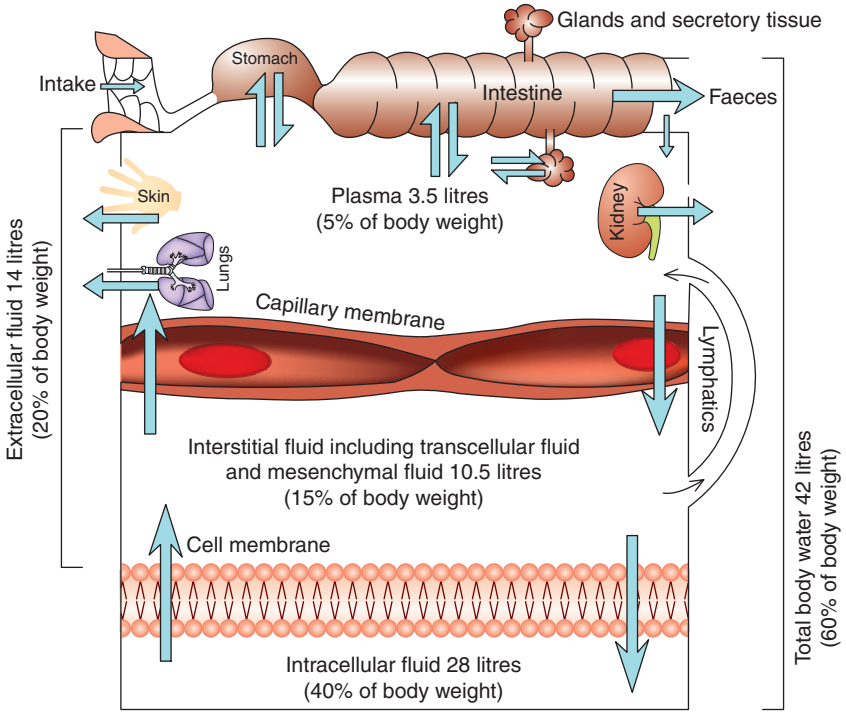


Fig. 1.1-1 Distribution of total body water in different compartments. Arrows indicate fluid movement.

Table 1.1-1 Distribution of total body water in a normal 70 kg person

Compartment	Volume (L)	Body weight (%)	Body water (%)
Total body water (TBW)	42	60	100
Intracellular fluid (ICF)	28	40	67
Extracellular fluid (ECF)	14	20	33
• Plasma (25% of ECF)	3.5	5	8
• Interstitial fluid, transcellular fluid and mesenchymal tissue fluid (75% of ECF)	10.5	15	25

2. Extracellular fluid compartment

The extracellular fluid (ECF) compartment constitutes about 20% of the body weight. The ECF compartment comprises following:

(i) Plasma. It is the fluid portion of the blood (intravascular fluid) and comprises about 5% of the body weight (i.e. 25% of the ECF). On an average out of 5 L of total blood volume 3.5 L is plasma.

(ii) Interstitial fluid including lymph. It constitutes the major portion (about 3/4) of the ECF. The composition of interstitial fluid is the same as that of plasma except it has little protein. Thus, interstitial fluid is an ultrafiltrate of plasma.

(iii) Transcellular fluid. It is the fluid contained in the secretions of the secretory cells and cavities of the body e.g. saliva, sweat, cerebrospinal fluid, intraocular fluids (aqueous humour and vitreous humour), pericardial fluid, bile, fluid present between the layers (pleura, peritoneum and synovial membrane), lacrimal fluid and luminal fluids of the gut, thyroid and cochlea.

Transcellular fluid volume is relatively small, about 1.5% of the body weight, i.e. 15 mL/kg body weight (about 1 L in a person of 70 kg).

(iv) Mesenchymal tissue fluid. The mesenchymal tissues such as dense connective tissue, cartilage and bones contain about 6% of the body water.

The interstitial fluid, transcellular fluid and mesenchymal tissue fluid combinedly form the 75% of ECF.

The normal distribution of total body water in the fluid compartments is kept constant by two opposing sets of forces: osmotic and hydrostatic pressure.

MEASUREMENT OF BODY FLUID VOLUMES

Theoretically, it is possible to measure the volume of each fluid component by injecting a substance (indicator) that will stay in only one compartment (provided the concentration of the substance in the body fluid and the amount removed by excretion and metabolism can be accurately measured) as:

$$V = \frac{A_1 - A_2}{C}$$ where,

V = Volume of fluid compartment,

A_1 = Amount of indicator injected in the fluid,
A_2 = Amount of indicator removed by excretion and metabolism, and
C = Concentration of the indicator in the fluid.

For example, if 150 mg of sucrose (A_1) is injected into a 70 kg man, 10 mg sucrose (A_2) has been excreted or metabolized and the concentration of plasma sucrose (C) measured is 0.01 mg/mL; then the volume distribution of sucrose is

$$\frac{150\,\text{mg} - 10\,\text{mg}}{0.01\,\text{mg/mL}} = 14{,}000\,\text{mL}$$

Prerequisites for accurate body fluid measurement

Though the formula described above for measuring the body fluid volume appears simple, the material injected (indicator) should have following characteristics:

- It should be non-toxic.
- It must mix evenly throughout the compartment being measured.
- It should be relatively easy to measure its concentration.
- It must have no effect of its own on the distribution of water or other substances in the body.
- Either it must be unchanged by the body during the mixing period or the amount changed (excreted and/or metabolized) must be known.

This method of measuring body fluids is called ***'indicator dilution method'*** and can be used to measure the volume of different compartments of the body fluid by using the suitable indicator/marker which will get distributed in that particular compartment as follows:

1. Measurement of total body water volume

The volume of TBW can be measured by injecting a marker which will be evenly distributed in all the compartments of body fluid. Such markers include:

- Deuterium oxide (D_2O),
- Tritium oxide, and
- Aminopyrine.

The volume of the TBW can be calculated from the values of the concentration of the marker in the plasma.

2. Measurement of extracellular fluid volume

The volume of ECF can be measured by injecting those marker substances which cannot enter the cells but can freely pass through the capillary membrane, and thus can distribute evenly in all the compartments of ECF. Such substances include:

- *Radioactive substances* like sodium, chloride ($36Cl^-$ and $38Cl^-$), bromide ($82Br^-$), sulphate and thiosulphate; and
- *Non-metabolizable saccharides* like inulin, mannitol and sucrose.

Most accurate method of measuring the volume of ECF is by using *inulin* (polysaccharide, MW 5200). The values of ECF volume are calculated from the values of concentration of *inulin* in the plasma since it makes an important component of the ECF.

3. Measurement of plasma volume

The plasma volume can be measured by injecting those markers which bind strongly with the plasma protein and either do not diffuse or diffuse only in small quantities into the interstitium. These substances are:

- Radioactive iodine – ^{131}I, and
- The dye Evan's blue – T-1824.

The plasma volume can also be calculated from the values of the Red Blood Cells which can be measured using radioactive isotopes of chromium (^{51}Cr).

4. Measurement of intracellular fluid volume

The volume of ICF cannot be measured directly, since there is no substance which can be confined exclusively to this compartment after intravenous injection. Therefore, values of ICF volume are calculated from the values of TBW and ECF as

ICF volume = TBW volume – ECF volume.

5. Measurement of interstitial fluid volume

Like ICF volume, the volume of interstitial fluid also cannot be measured directly for the same reasons. Its values can be roughly calculated from the values of ECF volume and plasma volume as

Interstitial fluid volume = ECF volume – plasma volume.

Note. The ECF volume/intracellular fluid volume ratio is larger in infants and children as compared to adults, but absolute volume of ECF in children is smaller than in adults. Therefore, dehydration develops rapidly, more frequently and severe in children than in adults.

BODY ELECTROLYTES

The electrolytes constitute about 7% of the total body weight. The distribution of electrolytes in various compartments differs markedly. Table 1.1-2 shows the distribution of electrolytes in two major compartments of body fluid: the ECF and the ICF.

From Table 1.1-2, it may be noted that in the ICF the main cations are K^+ and Mg^{2+}, and the main anions are PO_4^{3-} and proteins. While in the ECF, the predominant cation is Na^+ and the principal anions are Cl^- and HCO_3^- Besides these, a small proportion of non-diffusible proteins,

Table 1.1-2 Distribution of ions in the ECF and ICF (Values are in mEq/L of H_2O)

Ion	Extracellular fluid	Intracellular fluid
Cations		
Na^+	142	14
K^+	5.5	150
Ca^{2+}	5	<1
Mg^{2+}	3	58
Anions		
Cl^-	103	4
HCO_3^-	28	10
PO_4^{3-}	4	75
Proteins	1 g/dL	5 g/dL

nutrients and metabolites such as glucose and urea are also present in ECF.

The essential difference between the two main subdivisions of ECF is the higher protein content in plasma than in the interstitial fluid which plays an important role in maintaining fluid balance.

It is important to note that:

- Essentially all of the body K^+ is in the exchangeable pool.
- Only 65–70% of the body Na^+ is exchangeable.
- Almost all of the body Ca^{2+} and Mg^{2+} are non-exchangeable.
- Only the exchangeable solutes are osmotically active.

Functions of electrolytes

1. Electrolytes are the main solutes in the body fluids for maintenance of acid–base balance.
2. Electrolytes maintain the proper osmolality and volume of body fluids.
3. The concentration of certain electrolytes determines their specific physiologic functions, e.g. the effect of calcium ions on neuromuscular excitability.

INTERNAL ENVIRONMENT AND HOMEOSTASIS

INTERNAL ENVIRONMENT

Claude Bernarde (1949), the great French physiologist, introduced the term *internal environment* of the body or the *milieu interieur* for the ECF of the body. He said so since all the body cells essentially depend upon the ECF for maintenance of cellular life. Cells are capable of living, growing and performing their special functions so long as the proper concentration of oxygen, glucose, different ions, amino acids, fatty substances and other constituents are available in the internal environment.

HOMEOSTASIS

Homeostasis, a term introduced by W. B. Cannon, refers to the mechanism by which the constancy of the internal environment is maintained and ensured. For this purpose, living membranes with varying permeabilities such as vascular endothelium and cell membrane play important role.

The factors involved in the maintenance of internal environment can be summarized as:

- Transport of ECF,
- Maintenance of pH of ECF (acid–base balance),
- Regulation of temperature,
- Maintenance of water and electrolyte balance,
- Supply of nutrients, oxygen, enzymes and hormones,
- Removal of metabolic and other waste products, and
- Reproduction.

MODE OF ACTION OF HOMEOSTATIC CONTROL SYSTEM

The homeostasis is a complex phenomenon. The mode of operation of all the systems, which are involved in the homeostasis is through 'feedback' mechanism and the adaptive control system. Feedback mechanism is of two types: the negative feedback mechanism and the positive feedback mechanism.

Negative feedback mechanism

Most control systems of the body act by the negative feedback. That is, in general if the activity of a particular system is increased or decreased, a control system initiates a negative feedback, which consists of a series of changes that return the activity toward normal.

Examples of a feedback mechanism:

1. When the blood pressure suddenly rises or lowers, it initiates a series of reactions that tries to bring the blood pressure to normal levels.
2. When thyroxine secretion is more, it inhibits the secretion of thyroid stimulating hormone from pituitary so that, thyroxine is not secreted from the thyroid gland.

Positive feedback mechanism

Positive feedback is better known as a vicious circle. Usually it is harmful and in some instances even death can occur due to positive feedback. For example, as shown in Fig. 1.1-2, when a person has suddenly bled 2 L of blood, a vicious circle of progressively weakening of the heart is set which ultimately causes death.

A mild degree of feedback can be overcome by the negative feedback control mechanisms of the body, and a vicious cycle fails to develop. For example, when a patient bleeds

1 L of blood instead of 2 L, the negative feedback mechanisms of controlling the blood pressure may overcome the positive feedback, and the blood pressure will return to normal, as shown in Fig. 1.1-2.

Further, sometimes positive feedback can serve useful purposes, e.g. under following circumstances:

- *Clot formation* followed by rupture of vessels is accelerated by the vicious cycle of thrombin formation (for details see page 153). This stops the bleeding.
- *Child birth* during labour is facilitated by progressively increasing uterine contractions due to positive feedback from stretching of the cervix by head of the baby (for details see page 671).
- *Generation of nerve signals* by the vicious cycle of progressive leakage of Na^+ ions from the channels set up following stimulation of membrane of nerve fibre is due to the positive feedback.

Adaptive control system

Adaptive control system refers to a delayed type of negative feedback mechanism. This is seen in the nervous system. For example, when some movements of the body occur very rapidly, there is not enough time for nerve signals to travel from the peripheral parts of the body all the way to the brain and then back to the periphery again in time to control the movements. Under such circumstances brain uses a principle called *feed-forward control* to cause the required muscle contraction which retrospectively is conveyed to the brain by the sensory nerve signals from the moving part. If the movement performed is found incorrect, then the brain corrects the *feed-forward signals* that it sends to the muscle the next time the movement is required. Such a correction made by successive retrospective feedback mechanism is called *adaptive control.*

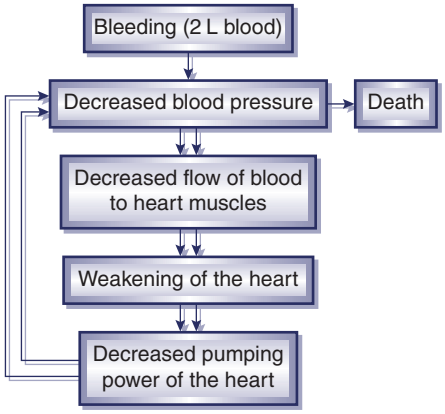


Fig. 1.1-2 Flowchart showing how a positive feedback mechanism can cause death.

Chapter 1.2

The Cell Physiology

CELL STRUCTURE
- **Cell membrane**
- **Cytoplasm**
- **Nucleus**

THE CELL MEMBRANE
- Fluid mosaic model of membrane structure
- Arrangement of different molecules in cell membrane

INTERCELLULAR JUNCTIONS
- Tight junction
- Adherens junction
- Gap junction

CELL STRUCTURE

The cell is the smallest structural and functional unit of the body. The human body contains about 100 trillion cells. Different types of cells of the body possess features which distinguish one type from the other and are specially adapted to perform particular functions, e.g. the red blood cells transport oxygen from lungs to the tissues, muscle cell is specialized for the function of contraction.

A typical cell, as seen by the light microscope, consists of three basic components:

- Cell membrane,
- Cytoplasm and
- Nucleus.

CELL MEMBRANE

Cell membrane or the plasma membrane is the protective sheath, enveloping the cell body. It separates the contents of cell from the external environment and controls exchange of materials between the fluid outside the cell (extracellular fluid) and the fluid inside the cell (intracellular fluid). A detailed knowledge of its structure (Fig. 1.2-1) is essential for the understanding of cell functions. Therefore, it will be discussed separately.

CYTOPLASM

Cytoplasm is an aqueous substance (cytosol) containing a variety of cell organelles and other structures. The structures dispersed in the cytoplasm can be broadly divided into three groups: organelles, inclusion bodies and cytoskeleton.

A. ORGANELLES

The organelles are the permanent components of the cells which are bounded by limiting membrane and contain enzymes hence participate in the cellular metabolic activity. These include:

1. Mitochondria

Mitochondria are the major sites for aerobic respiration. These are oval structures and more numerous in metabolically active cells.

Structure. The mitochondria consist of:
- *Membrane.* There are two layers of the membrane. The outer smooth and inner folded into incomplete septa called cristae (Fig. 1.2-1A).
- *Matrix* of the mitochondria contains enzymes required in Krebs' cycle by which products of carbohydrate, fat and protein metabolism are oxidised to produce energy which is stored in the form of ATP in the lollipop-like globular structures.

Functions. In addition to their role as power generating units, the mitochondria may have a role in synthesizing membrane bound proteins since they also possess deoxyribonucleic acid (DNA) and ribosomes.

2. Endoplasmic reticulum

Endoplasmic reticulum (ER) is a system of flattened membrane-bound vesicles and tubules called cisternae

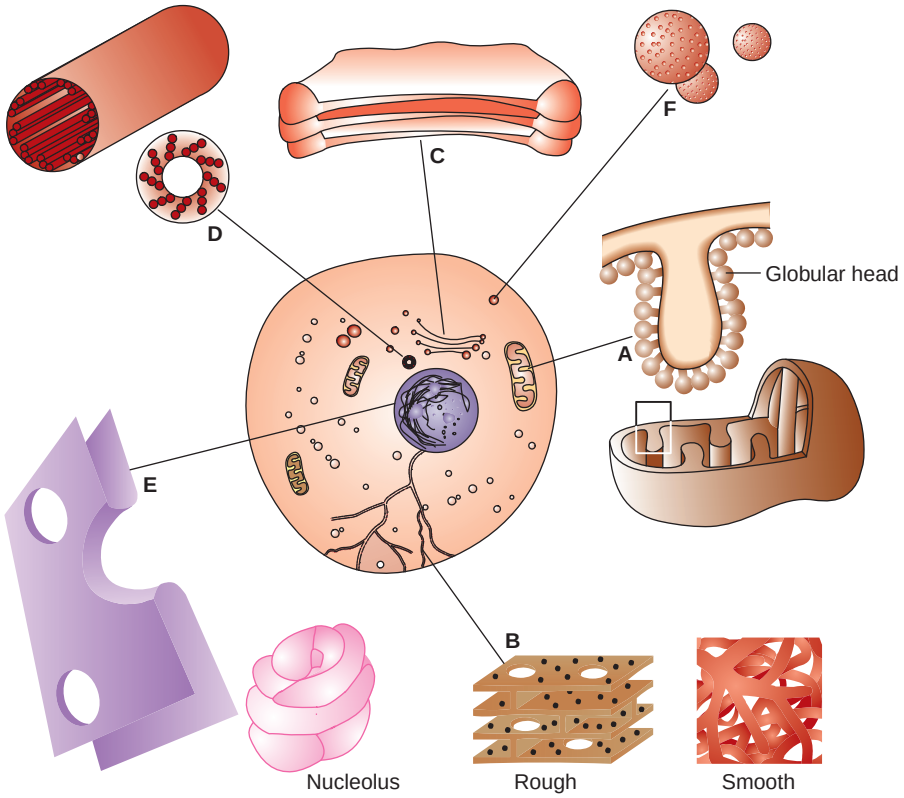


Fig. 1.2-1 Structure of a typical cell (in the centre) showing various organelles: A, mitochondrion; B, endoplasmic reticulum (rough and smooth); C, Golgi apparatus; D, centrosome; E, nucleus and F, secretory granules.

(Fig. 1.2-1B). It is continuous with the outer membrane of the nuclear envelop, Golgi apparatus and possibly with the cell membrane. Morphologically, two types of endoplasmic reticulum can be identified: rough or granular and smooth or agranular.

(i) *Rough endoplasmic reticulum.* The rough ER is characterized by the presence of a number of ribosomes on its surface and transports proteins made by the ribosomes through the cisternae. Thus, the rough ER is especially well developed in cells active in protein synthesis, e.g. Russell's bodies of plasma cells, Nissl granules of nerve cells and acinar cells of pancreas.

(ii) *Smooth endoplasmic reticulum.* Smooth ER is devoid of ribosomes on its surface. It is a site of lipid and steroid synthesis. Therefore, it is found in abundance with the Leydig cells and cells of the adrenal cortex. In the skeletal and cardiac muscles, smooth ER is modified to form sarcoplasmic reticulum which is involved in the release and sequestration of calcium ions during muscular contraction.

3. Golgi apparatus

The Golgi apparatus or complex is a collection of membranous vesicles, sacs or tubules which is generally located close to the nucleus. It is continuous with the endoplasmic reticulum. Golgi apparatus is particularly well developed in exocrine glandular cells (Fig. 1.2-1C).

Functions. Its main functions are:

- Synthesis of carbohydrates and complex proteins.
- Packaging of proteins synthesized in the rough ER into vesicles.
- Site of formation of lysosomal enzymes.
- Transport of the material to the other parts of cell or to the cell surface membrane and secretion.
- Glycosylation of proteins to form glycoproteins.

4. Ribosomes

Ribosomes are spherical particles which contain 80–85% of the cell's ribonucleic acid (RNA). They may be present in the cytosol as free (unattached) or in bound form (attached to the membrane of endoplasmic reticulum). Slightly smaller form of ribosomes is also found in mitochondria.

Functions. They are the site of protein synthesis. They synthesize all transmembrane proteins, secreted proteins and most proteins that are stored in the Golgi apparatus, lysosomes and endosomes.

5. Lysosomes

Lysosomes are rounded to oval membrane bound organelles containing powerful lysosomal digestive (hydrolytic) enzymes. They are formed by the Golgi apparatus. As many as 40 different lysosomal enzymes have been synthesized. Lysosomes are particularly abundant in cells involved in

phagocytic activity, e.g. neutrophils and macrophages. There are three forms of lysosomes:

- ***Primary lysosomes or storage vacuoles*** are formed from the various hydrolytic enzymes synthesized by rough ER and packaged in the Golgi apparatus.
- ***Secondary lysosomes or autophagic vacuoles*** are formed by fusion of primary lysosomes with parts of damaged or worn out cell components.
- ***Residual bodies*** are undigestible materials in the lysosomes.

6. Peroxisomes

Peroxisomes, also known as microbodies, are spherical structures enclosed by a single layer of unit membrane. These are predominantly present in hepatocytes and tubular epithelial cells.

Functions. They essentially contain two types of enzymes:
- *Oxidases* which are active in oxidation of lipid and
- *Catalases* which act on hydrogen peroxide to liberate oxygen.

7. Centrosome

The centrosome consists of two short cylindrical structures called centrioles (Fig. 1.2-1D). It is situated near the centre of the cell close to the nucleus. The centrioles are responsible for movement of chromosomes during cell division.

B. CYTOPLASMIC INCLUSIONS

The cytoplasmic inclusions are the temporary components of certain cells. These may or may not be enclosed in the membrane. A few examples of cytoplasmic inclusions are:

- *Lipid droplets.* These are seen in the cells of adipose tissue, liver and adrenal cortex.
- *Glycogen.* It is seen in the cells of liver and skeletal muscles.
- *Proteins as secretory granules* are seen in the secretory glandular cells (Fig. 1.2-1F).
- *Melanin pigment* is seen in the cells of epidermis, retina and basal ganglia.
- *Lipofuscin.* It is a yellow brown pigment believed to be derived from secondary lysosomes and is seen in the cardiac muscle and brain cells of elderly people.

C. CYTOSKELETON

The cytoskeleton is a complex network of fibres that maintains the structure of the cell and allows it to change shape and move. It primarily consists of (Fig. 1.2-2):

Microtubules

Microtubules are long hollow tubular structures without limiting membrane about 25 nm in diameter. These are made up of two globular protein subunits α- and β-tubulin. The bundles of tubulin give structural strength to the cells. Microtubules form the transport system of the cells. Some of the other organelles and protein molecules move to a different part of the cell through the microtubules. *Kinesin* and *dynein* known as molecular motors help in the movement of molecules through the microtubules.

The *cilia* and *flagella* which project from surface of certain cells (spermatozoa, respiratory mucosa and fallopian tubes) are also composed of microtubules enclosed in the plasma membrane and are active in the locomotion of the cells.

Intermediate filaments

Intermediate filaments are filamentous structures about 10 nm in diameter. Some of these filaments connect the nuclear membrane to the cell membrane. Their main function is to mechanically integrate the cell organelles within the cytoplasm. In their absence, cells rupture more easily; and when they are abnormal in human, blistering of the skin is common.

Microfilaments

Microfilaments are long solid filamentous structures having a diameter of 6–8 nm. These are made up of contractile proteins, actin and myosin. Actin is the most abundant protein in the mammalian cell. It attaches to various parts of

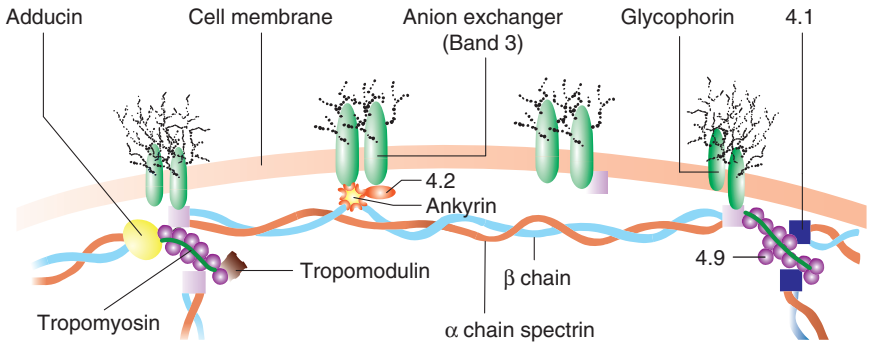


Fig. 1.2-2 Cytoskeleton showing various proteins.

cytoskeleton by other proteins (anchor proteins). These are identified by numbers as 4.1, 4.2 and 4.9. Extension of microfilaments along with the plasma membrane on the surface of the cells forms *microvilli* which increase the absorptive surface of the cells (e.g. intestinal epithelium). In the skeletal muscle, presence of actin and myosin filaments is responsible for their contractile property.

MOLECULAR MOTORS

Molecular motors help in the movement of different proteins, organelles and other cell parts (their cargo) to all parts of the cell. These can be divided into two types:

1. Microtubule-based molecular motors. This is a super-family of molecular motors that produce motion along microtubules. Two important molecular motors are:

(i) Conventional kinesin.
(ii) Dyneins.

2. Actin-based molecular motors. This is a super-family of molecular motors that produce motion along the actin. The important example of this group is myosin.

NUCLEUS

Nucleus is present in all the eukaryotic cells. It controls all the cellular activities including reproduction of the cell. Most of the cells are uninucleated except few types of cells like skeletal muscle cells which are multinucleated. The nucleus consists of (Fig. 1.2-1E):

1. Nuclear membrane

The nuclear membrane is double layered porous structure having a 40,270 nm wide space called perinuclear cistern which is continuous with the lumen of endoplasmic reticulum. The outer layer of the nuclear membrane is continuous with endoplasmic reticulum. The exchange of materials between the nucleoplasm and cytoplasm occurs through the nuclear membrane.

2. Nucleoplasm

The nucleoplasm or the nuclear matrix is a gel-like ground substance containing a large quantity of genetic material in the form of DNA. When a cell is not dividing, the nucleoplasm appears as dark staining thread-like material called nuclear *chromatin.* During cell division, the chromatin material is converted into rod-shaped structures, the chromosomes. There are 46 chromosomes (23 pairs) in all the dividing cells of the body except the gamete (sex cells) which contain only 23 chromosomes (haploid number). Each chromosome is composed of two chromatids connected at the centromere to form 'X' configuration having variation of the location of centromere.

The chromosomes are composed of three components: DNA, RNA and other nuclear proteins. The nuclear DNA carries the genetic information which is passed via RNA into the cytoplasm for synthesis of proteins of similar composition.

3. Nucleolus

The nucleus may contain one or more rounded bodies called nucleoli. The nucleoli are the site of synthesis of ribosomal RNA. The nucleoli are more common in growing cells or in cells that actively synthesize proteins.

THE CELL MEMBRANE

An understanding of the structure and properties of the cell membrane is most essential to understand the various physiological activities of the cell. Electron microscopy has shown that cell membrane/plasma membrane has a trilayer structure having a total thickness of 7–10 nm (70–100 Å) and is known as *unit membrane.* The three layers consist of two *electron dense layers* separated by an *electron lucent layer* (clear zone). Biochemically the cell membrane is composed of a complex mixture of lipids (40%), proteins (55%) and carbohydrates (5%).

A few hypotheses have been proposed to explain the distribution of various biochemical components in the cell membrane. The most important hypothesis is the fluid mosaic model of Singer and Nicholson.

FLUID MOSAIC MODEL OF MEMBRANE STRUCTURE

In 1972, Singer and Nicholson put forward the fluid mosaic model of membrane structure (Fig. 1.2-3), which is presently most accepted. According to this model:

- *Phospholipid bilayer* is the basic continuous structure forming the cell membrane. The phospholipids are present in fluid form. This fluidity makes the membrane quite flexible and thus allows the cells to undergo considerable changes in the shape without disruption of structural integrity.
- *The protein molecules* are present as a discontinuous mosaic of globular proteins which float about in the fluid phospholipid bilayer forming a *fluid mosaic pattern.*

ARRANGEMENT OF DIFFERENT MOLECULES IN CELL MEMBRANE

Arrangement of lipid bilayer of the cell membrane

Each lipid molecule in the lipid bilayer of the cell membrane primarily consists of phospholipid, cholesterol and glycolipids. The lipid molecule is clothes pin shape and consists of a *head end* and a *tail end.*

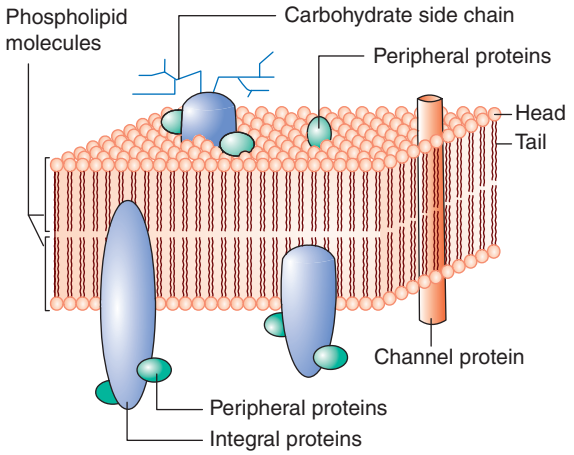


Fig. 1.2-3 Fluid mosaic model of the structure of cell membrane.

The head end or the globular end of the molecule. It is positively charged and quite soluble in water (i.e. *polar or hydrophilic*). The tail end consists of two chains of fatty acids or steroid radicle of cholesterol. It is quite insoluble in water (*nonpolar* or *hydrophobic*). These lipid molecules are arranged as bilayer in such a way that their nonpolar hydrophobic tail ends are directed towards the centre of the membrane whereas their polar hydrophilic head ends are directed outwards on either side of the membrane (Fig. 1.2-3). In this way head ends of molecules face the aqueous phase, i.e. extracellular fluid on outside and the intracellular fluid (cytoplasm) on inner side.

Functional significance of the lipid bilayer. The lipid bilayer of the cell membrane makes it a semipermeable membrane which constitutes the major barrier for the water soluble molecules like electrolytes, urea and glucose. On the other hand, fat soluble substances like oxygen, fatty acids and alcohol can pass through the membrane with ease.

Arrangement of proteins in the cell membrane

Most protein molecules float about in the phospholipid bilayer forming a fluid mosaic pattern. The two types of proteins recognized in the cell membrane are:

- *Lipoproteins,* i.e. the proteins containing lipids which function as enzymes and ion channels, and
- *Glycoproteins,* i.e. the proteins containing carbohydrates which function as receptors for hormones and neurotransmitters.

The proteins in the cell membrane are described to be arranged as:

1. Peripheral proteins. These are present peripheral to the lipid bilayer both inside and outside to it.

(i) *Intrinsic proteins.* These are located in the inner surface of the lipid bilayer and serve mainly as enzymes. Some of these are anchored to the cytoskeleton of the cell (Fig. 1.2-2).

(ii) *Extrinsic or surface proteins.* These are the proteins located on the outer surface of the lipid bilayer. These protein molecules are not associated tightly with the cell membrane and thus can dissociate readily from the cell membrane. Some of these proteins serve as cell adhesion molecules (CAMs) that anchor cells to neighbouring cells and to the basal lamina.

2. Integral proteins or transmembrane proteins. These are the proteins which extend into the lipid bilayer (Fig. 1.2-3). Some proteins penetrate only part of the way into the membrane while others penetrate all the way through. The integral proteins, on the basis of functions they serve are:

- *Channel proteins.* Some of the integral protein molecules serve as channels for water soluble substances like glucose and electrolytes. These are also called the channel proteins.
- *Carrier proteins.* The protein molecules which help in transport of substances across the cell membrane by means of active and passive (facilitated diffusion) transport are called carrier proteins.
- *Receptor proteins.* Some of the proteins function as receptors that bind neurotransmitters and hormones, initiating physiologic changes inside the cell.
- *Antigens.* Some proteins in the cell membrane also act as antigens. These are glycoproteins with branching carbohydrate side chains like antennae.
- *Pumps.* There are certain proteins in the cell membrane which act as pumps and form active transport system of the cell, e.g. Na^+–K^+ ATPase pump, K^+–H^+ ATPase pump and Ca^{2+} pump.

Arrangement of carbohydrates in the cell membrane

The carbohydrates are attached either to the proteins (glycoproteins) or the lipids (glycolipids). Throughout the surface of cell membrane, carbohydrate molecules form a thin loose covering called glycocalyx.

Functions of cell membrane carbohydrates

- Being negatively charged the carbohydrate molecules of the cell membrane do not allow the negatively charged particles to move out of the cell.
- The glycocalyx helps in tight fixation of the cells with one another.
- Some of the carbohydrate molecules also serve as *receptors.*

INTERCELLULAR JUNCTIONS

The cell membranes of the neighbouring cells are connected with one another through the intercellular junctions or the junctional complexes, which are of three types.

Types of intercellular junctions

1. Tight junction. This is also called *zona occludens* or the *occluding zone* (Fig. 1.2-4A). In this type of intercellular junction, the outer layer of the cell membrane of the neighbouring cells fuse with each other, thus obliterating the space between the cells. Such junctions form a barrier to the movement of ions and other solutes from one cell to another.

2. Adherens junction. This is also called zonula adherens. In this type of junction, cell membranes of the adjacent cells are separated by a 15–20 nm wide space which is at focal places obliterated by the dense accumulation of the proteins at the cell surface. Bundles of intermediate filaments project from the intercellular junctional areas and radiate into the cytoplasm. This holds the adjacent cells at these focal places. These are of two types:

- *Desmosomes* are the adherens junctions where thickened focal areas are formed on both the apposing cell membranes (Fig. 1.2-4B).
- *Hemidesmosomes* are the adherens junctions where focal thickening is seen only on the membrane of one of the two adjacent cells. So, this is also known as half desmosome. Adherens junctions are seen in the cells of epidermis.

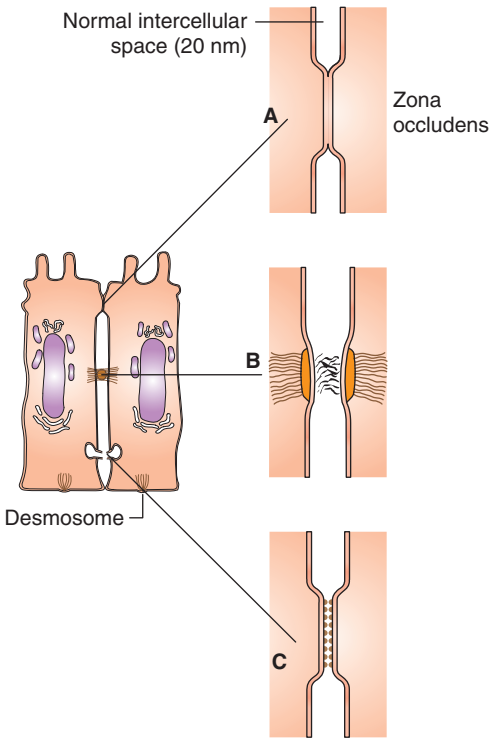


Fig. 1.2-4 Schematic diagram of a cell to show various intercellular junctions: A, tight junction; B, adherens junction and C, gap junction.

3. Gap junction. Gap junctions or the nexus are the channels on the lateral surfaces of the two adjacent cells through which the molecules are exchanged between the cells (Fig. 1.2-4C). Each half of the channel is surrounded by six subunits of proteins (the *connexins*). The intercellular space is reduced from the usual size of 15–20 nm to 2–3 nm at such junction. The gap junctions are seen in the heart and basal part of epithelial cells of intestinal mucous membrane. Gap junctions serve the following functions:

- These permit the intercellular passage of glucose, amino acids, ions and other substances which have a molecular weight of about 1000.
- These permit rapid propagation of electrical potential changes from one cell to another as seen in cardiac muscle and other smooth muscle cells.
- These help in the exchange of chemical messengers between the cells.

Cell adhesion molecules

Cell adhesion molecules (CAMs) are the prominent parts of the intercellular connections by which the cells are attached to the basal lamina and to each other.

Types of CAMs. CAMs have been variously classified. Most simply they can be divided in four broad families:

1. *Integrins.* These are molecules that bind to various receptors.
2. *Adhesion molecules of IgG subfamily.* Through these molecules the IgG immunoglobulins bind to various antigens.
3. *Cadherins.* These are Ca^{2+} dependent molecules that mediate cell-to-cell adhesions.
4. *Selectins.* These are lectin-like domains that bind carbohydrates.

Functions of CAMs. In addition to binding the neighbouring cells to each other the CAMs perform following other functions:

- They transmit signals into and out of the cells.
- They play a role in embryonic development and formation of the nervous system and other tissue.
- They hold tissue together in adults.
- They play an important role in inflammation and wound healing.
- They also play a role in metastasis of tumours.

Chapter 1.3

Transport Through Cell Membrane

PASSIVE TRANSPORT
- **Diffusion**
  - Simple diffusion
  - Facilitated diffusion
- **Osmosis**
  - Osmotic pressure
  - Osmole, osmolality and osmolarity
  - Tonicity of fluids

ACTIVE TRANSPORT
- **Primary active transport processes**
  - Sodium–potassium pump
  - Calcium pump
  - Potassium–hydrogen pump
- **Secondary active transport processes**
  - Sodium co-transport
  - Sodium counter-transport

VESICULAR TRANSPORT
- **Endocytosis**
- **Exocytosis**
- **Transcytosis**

OTHER TRANSPORT PROCESSES
- **Transport across epithelia**
- **Ultrafiltration**

The physiological activities of a cell depend upon the substances like nutrients, oxygen and water, which must be transported into the cell, and at the same time metabolic waste must be transported out of the cell. Various processes involved in the transport of substances across the cell membrane may be grouped as under:

- Passive transport,
- Active transport and
- Vesicular transport.

PASSIVE TRANSPORT

Passive transport refers to the mechanism of transport of substances along the gradient without expenditure of any energy. It depends upon the physical factors like concentration gradient, electrical gradient and pressure gradient. Since the transport of substances occurs along the gradient, this process is also called *down-hill movement.* The passive transport mechanisms operating at the cell membrane level are diffusion and osmosis.

DIFFUSION

Diffusion refers to passive transport of molecules from areas of higher concentration to areas of lower concentration. Diffusion through cell membrane is divided into two subtypes called: simple diffusion and facilitated diffusion.

SIMPLE DIFFUSION

In simple diffusion, transport of atoms or molecules occurs from one place to another due to their random movement. Due to constant random movement, the molecules collide with each other and also strike with the cell membrane. The frequency of collision and the probability of striking to the cell membrane will be higher on the side of the membrane having higher concentration of that particular molecule. In this process there occurs a net flux of the molecules from the areas of high concentration to areas of low concentration. The net movement of the molecules ceases when the concentration of molecules equals, and there occurs a condition of *diffusional equilibrium.* Quantitatively, the net movement of the molecules across a permeable membrane where only simple diffusion is occurring is expressed by *Fick's law of diffusion,* which states that rate of diffusion (J) is directly proportional to the difference in the concentration of the substance in two regions (concentration gradient, i.e. $C_1 - C_2$) and cross-sectional area (A) and inversely proportional to the distance to be travelled, i.e. thickness of the membrane (T).

Thus, $J = D \frac{A(C_1 - C_2)}{T}$, where D is the diffusion coefficient.

The diffusion of molecules across the biological membranes differs depending upon the lipid solubility, water solubility, type of electrical charge and size of the molecules. Further, selective permeability of the semipermeable cell membrane also affects the diffusion of different molecules. How the different molecules diffuse across a cell membrane are discussed below.

Simple diffusion of lipid soluble substances through the cell membrane

The rate of diffusion through the lipid bilayer of the cell membrane is directly proportional to the solubility of a substance in lipids. Therefore, molecules of substances like oxygen, nitrogen, carbon dioxide, alcohol, steroid hormones and weak organic acids and bases, being lipid soluble, diffuse very rapidly through the lipid bilayer of the cell membrane.

Simple diffusion of water and other lipid insoluble molecules through the cell membrane

Astonishingly, water and other lipid insoluble substances can also pass easily through the cell membrane. It has been shown that it is possible due to the presence of the so-called protein channels (made from transmembrane proteins) in the cell membrane.

Diffusion through protein channels

The protein channels are tube-shaped channels which extend in the cell membrane from the extracellular to the intracellular ends (Fig. 1.3-1). Therefore, even the highly lipid insoluble substances can diffuse by simple diffusion directly through these channels of the cell membrane.

The protein channels have been equipped with following characteristics:

- Selective permeability and
- Gating mechanism.

Selective permeability of protein channels

The protein channels are highly selective, i.e. each channel can permit only one type of ion to pass through it. This results from the characteristics of the channel itself, such as its diameter, its shape and nature of electrical charges along its inside surfaces. Examples of some selective channels are:

- *Sodium channels* are specifically selective for the passage of sodium ions. These are 0.3 by 0.5 nm in size and their inner surfaces are *strongly negatively charged* (Fig. 1.3-2).
- *Potassium channels* are specifically selective for the passage of potassium ions. These are 0.3 by 0.3 nm in size and are not negatively charged (Fig. 1.3-3).

Gating mechanism in protein channels

Some protein channels are continuously open, whereas most others are 'gated', i.e. they are equipped with actual gate-like extensions of the transport protein molecule which can open and close as per requirement. This gating mechanism is a means of controlling the permeability of the channels. The opening and closing of gates are controlled by three principal ways:

1. ***Voltage-gated channels.*** These respond to the electrical potential across the cell membrane. As shown in Fig. 1.3-2 in the case of sodium channels the gates are located at the outer end of the channels and these remain tightly closed, when there is a strong negative charge on the inside of cell membrane. When the inside of cell membrane loses its negative charge, these gates open and there occurs a tremendous inflow of sodium ions. This is the basis of occurrence of action potentials in nerves that are responsible for nerve signals.

   In the case of potassium channels, the gates are located at inner end of the channel (Fig. 1.3-3) and they too open when inside of the cell membrane loses its negative charge, but this response is much slower than that for sodium channel. The opening of potassium channel gates is partly responsible for terminating the action potential.
2. ***Ligand-gated channels.*** Gates of these channels open when some other chemical molecule binds with the gate proteins that is why this is also called chemical gating. One of the most important example of ligand channel gating is the effect of acetylcholine on the so-called

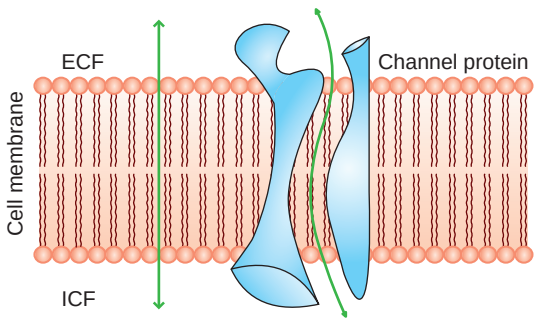


Fig. 1.3-1 Simple diffusion.

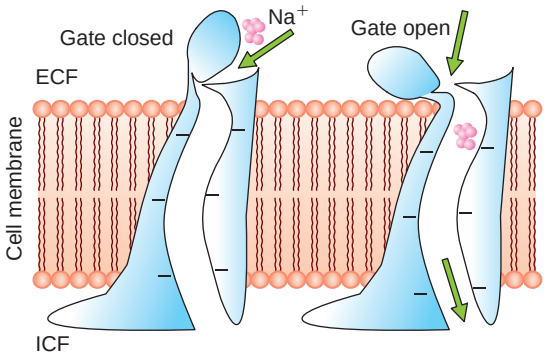


Fig. 1.3-2 Voltage-gated sodium channels.

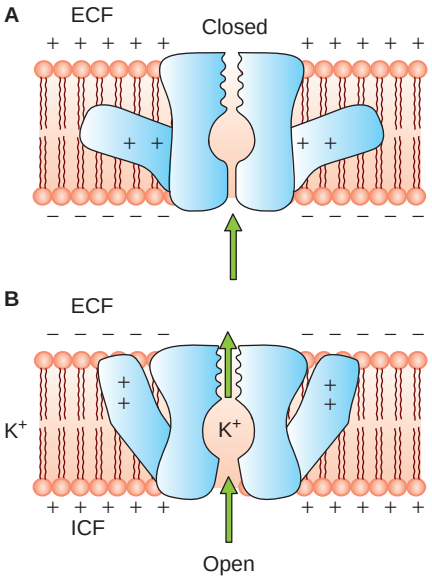


Fig. 1.3-3 Voltage-gated potassium channels.

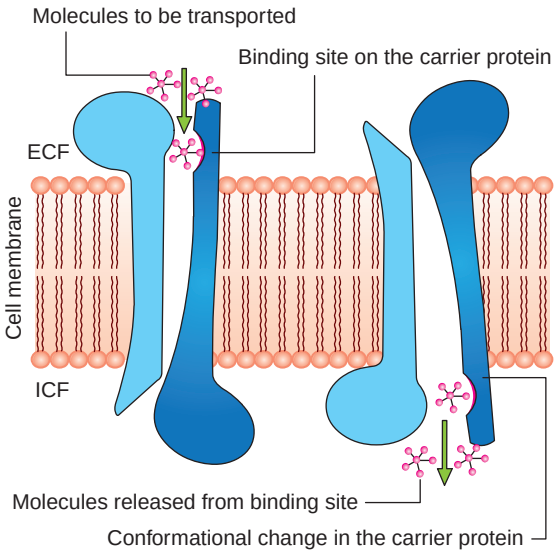


Fig. 1.3-4 Postulated mechanism of facilitated diffusion.

acetylcholine channels. This gate plays an important role in transmission of nerve signals from one nerve cell to another and from nerve cells to muscle cells.

3. ***Mechanical-gated channels.*** Some protein channels are opened by mechanical stretch. These mechano-sensitive channels play an important role in cell movements.

FACILITATED DIFFUSION

The water soluble substances having larger molecules such as glucose cannot diffuse through the protein channels by simple diffusion. Such substances diffuse through the cell membrane with the help of some carrier proteins. Therefore, this type of diffusion is called facilitated or the carrier-mediated diffusion. There are many types of carrier proteins in the cell membrane, each type having binding sites that are specific for a particular substance. Among the most important substances that cross cell membranes by facilitated diffusion are *glucose* and most of the *amino acids.*

Mechanism of facilitated diffusion

Postulated mechanism for facilitated diffusion is shown in Fig. 1.3-4. As shown in the figure, a conformational change occurs in the carrier protein after the molecule to be transported is bound at the receptor site. The repetitive spontaneous configurational changes allow the diffusion of the molecule.

Types of carrier protein systems

Three types of carrier protein systems are known: uniport, symport and antiport (Fig. 1.3-5). The symports and antiport are together known as co-transport.

1. ***Uniport.*** In this system the carrier proteins transport only one type of molecules (Fig. 1.3-5A).
2. ***Symport.*** In this system transport of one substance is linked with transfer of another substance. For example, facilitated diffusion of glucose in the renal tubular cells is linked with the transport of sodium (Fig. 1.3-5B).

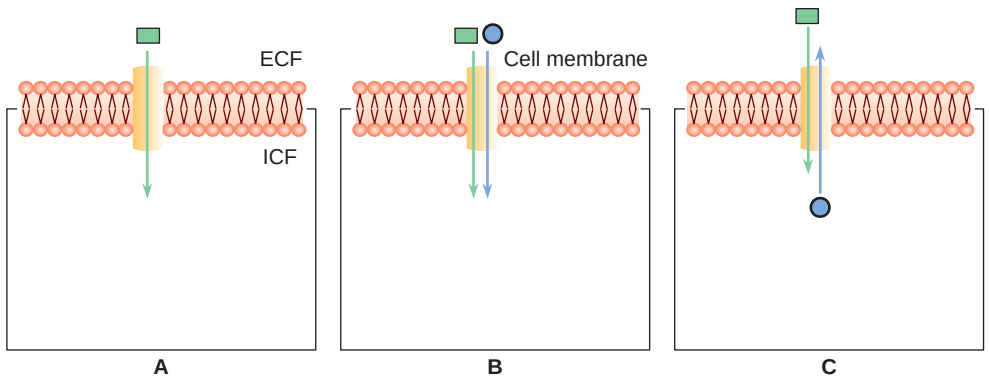


Fig. 1.3-5 Various types of carrier protein systems: A, uniport; B, symport and C, antiport.

3. ***Antiport.*** In this system the carrier proteins exchange one substance for another. For example, Na^+–K^+ exchange or Na^+–H^+ exchange in the renal tubules (Fig. 1.3-5C).

Differences between simple and facilitated diffusion

1. ***Specificity.*** The carrier proteins are highly specific for different molecules.
2. ***Saturation.*** As shown in Fig. 1.3-6, in simple diffusion the rate of diffusion increases proportionately with the increase in the concentration of the substance and there is no limit to it. However, in facilitated diffusion the rate of diffusion increases with increase in concentration gradient to reach a limit beyond which a further increase in the diffusion cannot occur. This is called *saturation* point and here all the binding sites on the carrier proteins are occupied and the system operates at its maximum capacity.
3. ***Competition.*** When two molecules, say A and B are carried by the same protein there occur a competition between the two molecules for the transport. Thus, an increase in the concentration of 'A' molecule will decrease the transport of molecule 'B' and vice versa. No such competition is known to occur in simple diffusion.

FACTORS AFFECTING NET RATE OF DIFFUSION

The diffusion of the substance can occur either way, i.e. extracellular fluid (ECF) to intracellular fluid (ICF) or vice versa depending upon the prevailing environment. The factors which affect the net rate of diffusion in the desired direction are:

1. ***Cell membrane permeability.*** Permeability of the cell membrane (P) is the major determining factor for the net diffusion, which in turn depends upon the following factors:
   - *Thickness of the membrane.* The diffusion is inversely proportional to the thickness of the cell membrane.
   - *Lipid solubility.* Diffusion is directly proportional to the lipid solubility of the substance.
   - *Distribution of protein channels in the cell membrane.* The rate of diffusion of lipid insoluble substance is directly proportional to the number of channels per unit area of the cell membrane.
   - *Temperature.* Rate of diffusion increases with increase in the temperature. This is because of the increased motion of the molecules and ions of the solution with increase in temperature.
   - *Size of the molecules.* Rate of simple diffusion is inversely proportional to the size of molecules.
   - *Area of the membrane.* The net diffusion of the substance is directly proportional to the total area of the membrane.
2. ***Concentration gradient.*** The simple diffusion is directly proportional to the concentration gradient (Fig. 1.3-7) but, the facilitated diffusion, however, has certain limitations beyond certain level of concentration gradient (Fig. 1.3-6B).

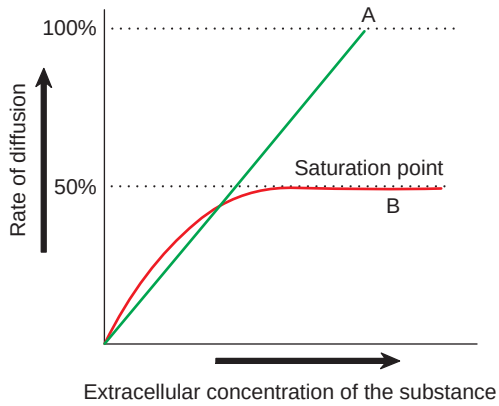


Fig. 1.3-6 Effect of concentration of substance on rate of diffusion in: A, simple diffusion and B, facilitated diffusion.

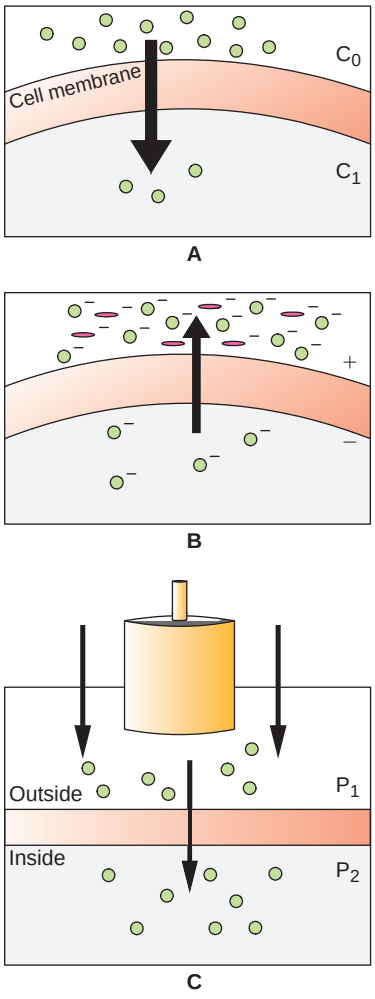


Fig. 1.3-7 Factors affecting net rate of diffusion: A, concentration gradient; B, electrical gradient and C, pressure gradient.

3. ***Electrical potential gradient.*** Electrical potential across the cell membrane is another important factor which affects the diffusion of ions across the cell membrane.

   As shown in Fig. 1.3-7B, the concentrations of negative ions are the same on both sides of the membrane, but there is an electrical gradient across the cell membrane because of positive charge outside and negative charge inside the membrane. The positive charge attracts the negative ions, whereas the negative charge repels them. Therefore, net diffusion occurs from inside to outside till the concentration gradient created balances the electrical gradient.
4. ***Pressure gradient.*** It has been observed that the increased amounts of energy are available to cause net movements of molecules from the high pressure side towards the low pressure side. The pressure gradient effect is demonstrated in Fig. 1.3-7C, which shows that high pressure developed by the piston on one side of the cell membrane causes greater number of molecules to strike the membrane resulting net diffusion to the other side.

OSMOSIS

Osmosis refers to diffusion of water or any other solvent molecules through a semipermeable membrane (i.e. membrane permeable to solvent but not to the solute) from a solution containing lower concentration of solutes towards the solution containing higher concentration of solutes; Fig. 1.3-8 shows osmosis across a selective permeable membrane. When a sodium chloride solution is placed on one side of the membrane and water on the other side (Fig. 1.3-8A) the net movement of water occurs from the pure water into the sodium chloride solution (Fig. 1.3-8B).

OSMOTIC PRESSURE

Osmotic pressure refers to the minimum pressure which when applied on the side of higher solute concentration prevents the osmosis. Figure 1.3-8C shows that when appropriate pressure is applied the net diffusion of water into the sodium chloride solution is prevented.

The osmotic pressure in the body fluids refers to the pressure exerted by the solutes dissolved in water or other solvents. The osmotic pressure exerted by the colloidal substances in the body is called *colloidal osmotic pressure.* The colloidal osmotic pressure due to plasma colloids (proteins) is called *oncotic pressure.*

The osmotic pressure depends upon the number of molecules or ions dissolved in a solution rather than their size, type or chemical composition.

OSMOLE, OSMOLALITY AND OSMOLARITY

Osmole is the unit used in place of grams to express the concentration in terms of number of osmotically active particles in a given solution. One osmole is equal to the molecular weight of a substance in grams divided by the number of freely moving particles liberated in solution by each molecule. Thus:

- A molar solution of glucose contains 1 mole and exerts osmotic pressure of 1 atm.
- A molar solution of NaCl contains 2 osmoles (1 mole of Na^+ and 1 mole of Cl^-) and exerts osmotic pressure of 2 atmospheres.
- A molar solution of $CaCl_2$ contains 3 osmoles (1 mole of Ca^{2+} and 2 moles of Cl^-) and thus exerts osmotic pressure of 3 atm.
- One milli osmole (mOsm) is 1/1000 of an osmole.

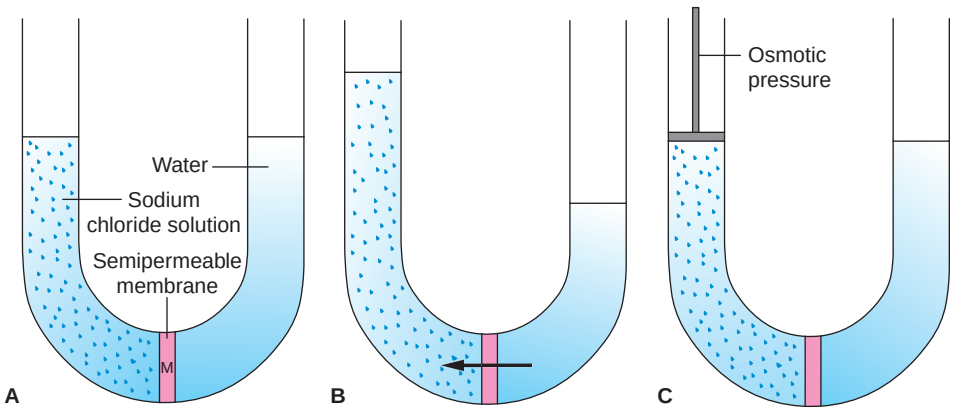


Fig. 1.3-8 Diagrammatic representation of phenomenon of osmosis: A, semipermeable membrane 'M' separates sodium chloride solution from pure water; B, net movement of water occurs through (M) from pure water side into the sodium chloride solution side and C, demonstration of osmotic pressure (net movement of water from pure water side to sodium chloride solution is prevented by applying appropriate pressure on the solution side).

Osmolality of a solution refers to the number of osmotically active particles (osmoles) per kilogram (kg) of a solution, whereas, ***osmolarity*** refers to the number of osmoles per litre (L) of a solution. Therefore, osmolarity is affected by the volume of the various solutes in the solution and the temperature, while the osmolality is not. The osmotic pressure is determined by the osmolality and not the osmolarity. However, the quantitative differences between the osmolarity and osmolality are less than 1%. In practice, osmolarity is more frequently used in physiological studies, since it is far more easier to measure osmolarity *vis-a-vis* osmolality.

Normal plasma osmolality. The normal osmolality of the extracellular and intracellular fluids is 290 milliosmoles per kilogram (mOsm/kg). In the plasma of the total osmolality 270 mOsm are contributed by Na^+, Cl^- and HCO_3^-. The remaining 20 mOsm are contributed by glucose and urea. Because of the large molecular weight and hence lesser number of particles, plasma proteins (70 g/L) contribute 2 mOsm to the total plasma osmolality.

TONICITY OF FLUIDS

In clinical practice, the word tonicity always refers to tonicity of a solution with respect to that of plasma (290 mOsm). In other words, it is the red blood cell (RBC) membrane across which the tonicity is tested. Thus:

- *Isotonic fluids* are those which have osmolality similar to plasma. RBCs neither shrink nor swell in such solution (Fig. 1.3-9A). A solution of 0.9% NaCl is isotonic with plasma.
- *Hypertonic fluids* have osmolality higher than the plasma. The RBCs shrink in such solutions by losing water by osmosis (Fig. 1.3-9B).
- *Hypotonic fluids* are those whose osmolality is lower than that of plasma. The RBCs swell up in hypotonic solutions by gaining water by osmosis (Fig. 1.3-9C).

APPLIED ASPECTS

- The total plasma osmolality may increase in patients having severe dehydration.
- Increased blood glucose levels in patients with severe diabetes also increase the plasma osmolality.
- Excessive intravenous administration of 5% glucose decreases plasma osmolality leading to swelling of the body tissues.
- Hyperosmolality can cause coma by causing water to flow out of the brain cells (hyperosmolar coma).
- Raised plasma levels of urea in patients with renal diseases also cause hyperosmolality.

ACTIVE TRANSPORT

Active transport refers to the mechanism of transport of substances against the chemical and/or electrical gradient. Active transport involves expenditure of energy which is liberated by breakdown of high energy compounds like adenosine triphosphate (ATP). Since the transport of substances occur against the chemico-electrical gradient, this process is also called *up-hill movement.* Substances transported actively across the cell membrane include:

- ***Ionic substances*** such as Na^+, K^+, Ca^{2+}, Cl^- and I^-, and
- ***Non-ionic substances*** like glucose, amino acids and urea.

TYPES OF ACTIVE TRANSPORT

The active transport is of two types:

- Primary active transport and
- Secondary active transport.

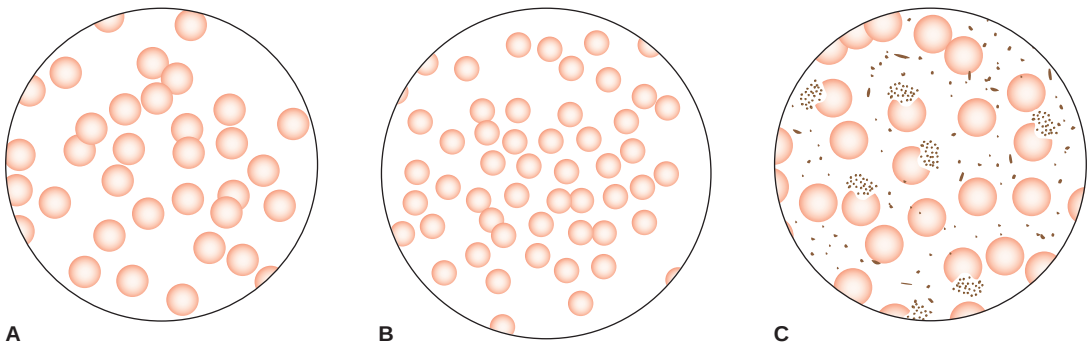


Fig. 1.3-9 Tonicity of fluids: A, isotonic fluid (0.9% NaCl) has osmolarity similar to plasma, RBCs neither shrink nor swell in it; B, hypertonic fluid (2% NaCl) has osmolarity higher than plasma, RBCs shrink in it and C, hypotonic fluid (0.3% NaCl) has osmolarity lower than plasma, RBCs swell in it.

A. PRIMARY ACTIVE TRANSPORT PROCESSES

In primary active transport process, the energy is derived directly from the breakdown of ATP or some other high-energy phosphate compound. Some of the important pumps involved in the primary active transport processes are:

- Sodium–potassium pump,
- Calcium pump and
- Potassium–hydrogen pump.

1. Sodium–potassium pump

Sodium–potassium (Na^+–K^+) pump is present in all the cells of the body. It is involved with the active transport of sodium ions outwards through the cell membrane and potassium ions inwards simultaneously. Thus, this pump is responsible for maintaining the Na^+ and K^+ concentration differences across the cell membrane and for establishing a negative electrical potential inside the cells.

Structure of Na^+–K^+ pump (Fig. 1.3-10). The carrier protein involved in Na^+–K^+ pump is a complex consisting of two separate protein units, a larger α subunit (molecular weight approximately 100,000) and a smaller β subunit (molecular weight approximately 55,000). The α subunit is mainly concerned with Na^+–K^+ transport. It has got following binding sites:

- Three intracellular sites, one each for binding sodium ions ($3Na^+$) and ATP, and one phosphorylation site.
- Two extracellular sites, one each for binding potassium ions ($2K^+$) and ouabain.

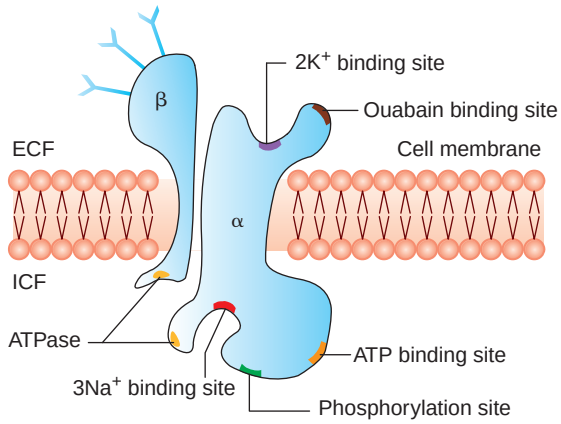


Fig. 1.3-10 Structure of sodium–potassium ATPase pump.

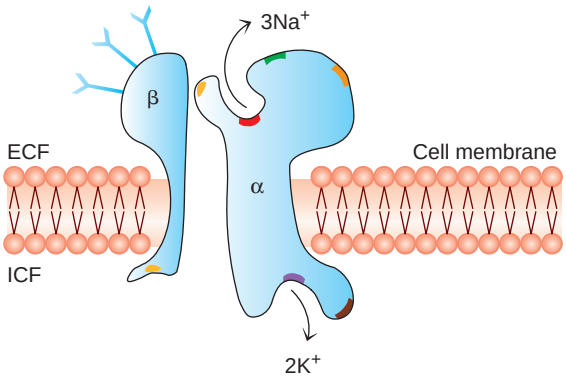


Fig. 1.3-11 Mechanism of operation of sodium–potassium ATPase pump.

Mechanism of operation of Na^+–K^+ pump (Fig. 1.3-11). The functioning of Na^+–K^+ pump involves the use of enzyme ATPase. The enzyme ATPase is activated when three sodium ions and one ATP molecule bind to their respective binding sites. The activated ATPase catalyzes the hydrolysis of ATP to ADP and liberates a high-energy phosphate bond of energy (phosphorylation). The energy so liberated is believed to cause a conformational change in the carrier protein molecule extruding sodium into the extracellular fluid. This is followed by binding of two potassium ions to the receptor site on extracellular surface of the carrier protein and dephosphorylation of a subunit which returns to its previous conformation, releasing potassium into the cytoplasm.

Functions of Na^+–K^+ pump. The Na^+–K^+ pump subserves two main functions:

1. *Controlling the cell volume.* It is the most important function of the Na^+–K^+ pump, without which most of cells of the body will swell up until they burst. When the Na^+–K^+ pump fails the cells swell up and burst.
2. *Electrogenic activity.* Na^+–K^+ pump acts as an electrogenic pump since it produces a net movement of positive charge out of the cell ($3Na^+$ out and $2K^+$ in); thus creating electrical potential across the cell membrane. This is basic requirement in nerves and muscles to transmit the signals.

2. Calcium pump

The calcium pump forms another important active transport mechanism. Like Na^+–K^+ pump, it also operates through a carrier protein which has ATPase activity. But the difference from Na^+–K^+ pump is that the carrier protein binds calcium ions rather than sodium and potassium ions. The calcium pump helps in maintaining extremely low concentration of calcium in the intracellular fluid (10,000 times less than the ECF).

3. Potassium–hydrogen pump

The primary active transport system of hydrogen ion also operates through ATPase (K^+–H^+ ATPase) activity. These are present at following two places in the human body:

- Parietal cells of gastric glands (see page 465) and
- Renal tubules (see page 395).

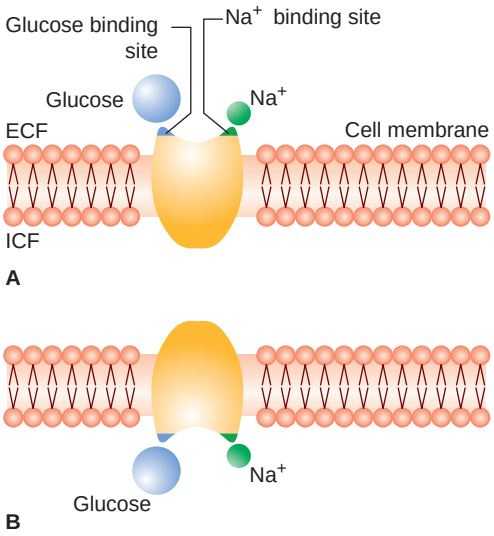


Fig. 1.3-12 Postulated mechanism of sodium co-transport of glucose (secondary active transport): A, carrier protein has two receptor sites, one for sodium and one for glucose and B, conformational change in carrier proteins causes transport of both glucose and sodium inside the cell simultaneously.

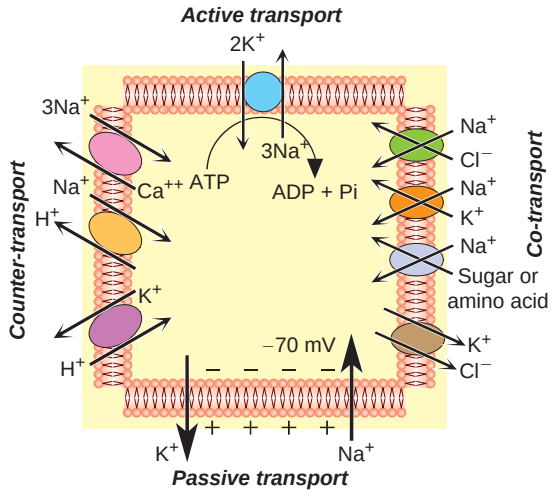


Fig. 1.3-13 Composite diagram of the cell showing the various co-transport and counter-transport mechanisms and maintenance of membrane potential as an effect of primary active transport of Na^+ and K^+.

B. SECONDARY ACTIVE TRANSPORT PROCESSES

In secondary active transport processes, the energy is derived secondarily from the energy which has been stored in the form of ionic concentration differences between the two sides of a membrane, created in the first place by primary active transport. At many areas in the body, transport of some other substance is coupled with the active transport of Na^+, i.e. the same carrier protein which is involved in the active transport of Na^+ also secondarily transports some other substance. The secondary active transport of substance may occur in the form of sodium co-transport or sodium counter-transport.

Sodium co-transport

The carrier protein here acts as a symport, i.e. transports some other substance along with the sodium. Substances carried by sodium co-transport include glucose, amino acids, chloride and iodine.

Sodium co-transport of glucose. The glucose is transported into most cells against large concentration gradient. As shown in Fig. 1.3-12A, the carrier protein has two receptor sites on the outer surface, one for sodium and other for glucose. The special feature of the carrier protein is that the conformational change in it occurs only when both the sodium and glucose molecules are attached to it. Due to conformational change in the carrier protein both the sodium and the glucose are transported simultaneously inside the cell (Fig. 1.3-12B). The co-transport of glucose occurs during its absorption from the intestine into the blood and during the reabsorption of glucose from renal tubule in the blood.

Sodium co-transport of amino acids. Occurs especially in the epithelial cells of intestinal tract and renal tubules during absorption of the amino acids into the blood. The mechanism of sodium co-transport of amino acids is similar to that of glucose, except that the carrier proteins involved are different.

Sodium counter-transport

The carrier protein involved here acts as an *antiport,* i.e. sodium ion is exchanged for some other substance. Some of the sodium counter-transport mechanism occurring in the body are:

1. ***Sodium–calcium counter-transport*** is known to occur in almost all cell membranes with sodium ions moving inside and calcium outside the cell (Fig. 1.3-13).
2. ***Sodium–hydrogen counter-transport*** is especially known in the proximal tubules of kidney. Here the Na^+ ions move inside the cell and the H^+ ions move out of the cell by the same carrier protein.
3. ***Other counter-transport systems*** which exist somewhere in the body are sodium–potassium counter-transport system, sodium–magnesium counter-transport, calcium–magnesium counter-transport system and chloride–bicarbonate counter-transport system.

VESICULAR TRANSPORT

Vesicular transport mechanisms are involved in the transport of macromolecules such as large protein molecules

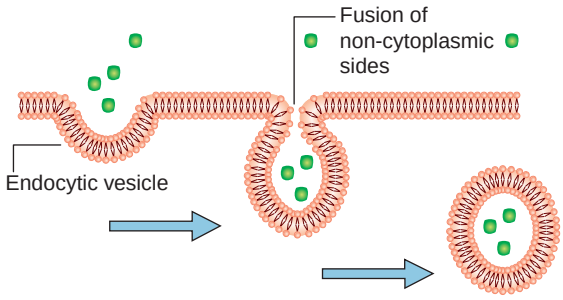


Fig. 1.3-14 Endocytosis.

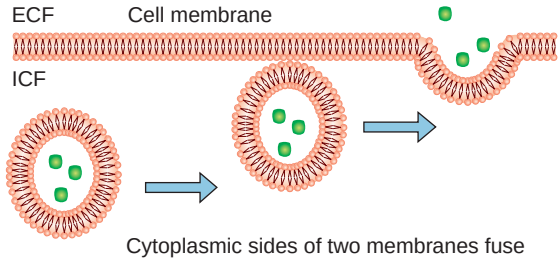


Fig. 1.3-15 Exocytosis.

which can neither pass through the membrane by diffusion nor by active transport mechanisms. The vesicular transport mechanisms include endocytosis, exocytosis and transcytosis.

ENDOCYTOSIS

Endocytosis is the process in which the substance is transported into the cell by infolding of the cell membrane around the substance and internalising it (Fig. 1.3-14). It is further categorized into three types:

1. ***Pinocytosis,*** i.e. cell drinking refers to the process of engulfing liquid substances by the enfolding of cell membrane, e.g. reabsorption by renal tubular epithelial cells.
2. ***Phagocytosis,*** i.e. cell eating is the process of engulfing of solid particles, such as bacteria, dead tissue and foreign particles by the cells. The process of phagocytosis involves three steps: (i) the attachment stage, (ii) the engulfment stage and (iii) killing or degradation stage (see page 126).
3. ***Receptor-mediated endocytosis.*** In this process the substance to be transported binds with the special receptor protein present on the cell surface. The receptor protein–substance complex is then engulfed by the cell membrane by the process of endocytosis. Transport of iron and cholesterol into the cells occurs by receptor-mediated endocytosis.

EXOCYTOSIS

Exocytosis (Fig. 1.3-15) is reverse of endocytosis, i.e. by this process the substances are expelled from the cell without passing through the cell membrane. In this process, the substances which are to be extruded are collected in the form of granules or vesicles which move towards the cell membrane. Their membrane then fuses to the cell membrane. The area of fusion breaks down releasing the contents to the exterior and leaving the cell membrane intact. Release of hormones and enzymes by secretory cells of the body occurs by exocytosis. The process of exocytosis requires Ca^{2+} and energy along with docking proteins.

TRANSCYTOSIS

Vesicular transport within the cell is called transcytosis or cytopempsis. It is quite similar to exocytosis and endocytosis. Three basic steps involved in this process are: (i) vesicle formation, (ii) vesicle transportation and (iii) docking in the cell.

OTHER TRANSPORT PROCESSES

So far we have considered transport across the cell membrane, i.e. movement of substances between the ICF and the ECF through the cell membrane. In addition to this, there are many situations in the body where transport of substances occurs through the epithelia and the capillary endothelial cell membrane. Some of these processes discussed briefly are:

- Transport across epithelia and
- Ultrafiltration.

TRANSPORT ACROSS EPITHELIA

Transport across epithelia involves movement of the substances from one side of the epithelium to the other. The transepithelial transport occurs in body cavities lined by continuous sheet of cells, such as in gastrointestinal tract, renal tubules, pulmonary airways and other structures. For transepithelial transport to occur, the cells need to be bound by tight junctions and have different ion channels and transport protein in different parts of their membrane.

ULTRAFILTRATION

When a solution of protein and salt is separated from plain water or a less concentrated salt solution by a membrane permeable to salt and water and not to the protein, there will be a net movement of water on the protein side by

diffusion and a movement of salt away from the protein side. This process is called dialysis.

Ultrafiltration refers to occurrence of dialysis under hydrostatic pressure (see page 441). Ultrafiltration is occurring at the capillary level in the body. The capillary blood is under hydrostatic pressure. The pressure is 35 mm Hg near the arteriolar end and gradually declines to 12 mm Hg near the venous end of the capillary. Through the capillaries there occurs ultrafiltration of all the constituents of the plasma except the proteins into the interstitial spaces. Ultrafiltration plays important role in the formation of body fluids (see page 238) (Fig. 4.4-19).

Chapter 1.4

Membrane Potential

INTRODUCTION

There exists a potential difference across the membrane of all living cells with the inside being negative in relation to the outside. This potential difference is named *membrane potential* because the cations and anions arrange themselves along the outer and inner surfaces of the cell membrane. The magnitude of membrane potential varies from cell to cell and in a particular cell varies according to its functional status. For example, a nerve cell has a membrane potential of –70 mV (inside negative) at rest, but when it gets excited the membrane potential becomes about +30 mV (inside positive). The membrane potential at rest is called *resting membrane potential* or resting transmembrane potential or simply resting potential. The term rest does not imply that cell is metabolically quiescent but that it is not undergoing any electrical change. The membrane potential measured during excited state of the cell is called *action potential.*

GENESIS OF MEMBRANE POTENTIAL

Membrane potential is basically due to unequal distribution of ions across the cell membrane, which in turn results due to the combined effect of various forces acting on the ions. The factors involved in genesis of membrane potential are:

- Selective permeability of the cell membrane,
- Gibbs'–Donnan membrane equilibrium,
- Nernst equation,
- Constant field Goldmann equation and
- Sodium–potassium ATPase pump.

SELECTIVE PERMEABILITY OF THE CELL MEMBRANE

The cell membrane is selectively permeable, that is, to some ions it is freely permeable, to others impermeable and to some others it has variable permeability as:

- Ions like Na^+, K^+, Cl^- and HCO_3^- are diffusible ions. The cell membrane is freely permeable to K^+ and Cl^- and moderately permeable to Na^+.
- The cell membrane is practically impermeable to intracellular proteins and organic phosphate which are negatively charged ions.
- Presence of gated channels in the cell membrane is responsible for the variable permeability of certain ions in different circumstances.

GIBBS'–DONNAN MEMBRANE EQUILIBRIUM

According to Gibbs'–Donnan membrane equilibrium, when two ionized solutions are separated by a semipermeable membrane at equilibrium:

- Each solution shall be electrically neutral, i.e. total charges on cations will be equal to total charges on anions.
- The product of diffusible ions on one side of the membrane will be equal to product of diffusible ions on the other side of the membrane.

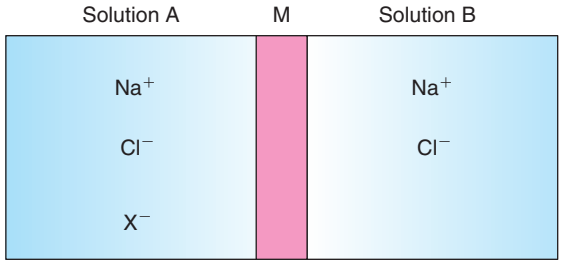


Fig. 1.4-1 Semipermeable membrane 'M' separates ionic solutions of sodium chloride A and B.

To understand let us consider 'M' is a semipermeable membrane separating two ionized solutions of sodium chloride A and B (Fig. 1.4-1). Then according to the *Gibbs'–Donnan equilibrium:*

1. Each solution is electrically neutral, i.e. $(cations)_A = (anions)_A$ and $(cations)_B = (anions)_B$

   OR

   $(Na^+)_A = (Cl^-)_A$ and $(Na^+)_B = (Cl^-)_B$.

2. The product of diffusible ions on both sides will be equal, i.e. $(diffusible\ cations)_A \times (diffusible\ anions)_A = (diffusible\ cations)_B \times (diffusible\ anions)_B$

   OR

   $[(Na^+)_A \times (Cl^-)_A] = [(Na^+)_B \times (Cl^-)_B]$

From the above, the concentration ratio of diffusible ions at equilibrium will be as below:

$$\frac{(diffusible\ cations)_A}{(diffusible\ cations)_B} = \frac{(diffusible\ anions)_B}{(diffusible\ anions)_A}$$

OR

$$\frac{[Na^+]_A}{[Na^+]_B} = \frac{[Cl^-]_B}{[Cl^-]_A}$$

Thus there will be symmetrical distribution of ions at equilibrium. But if one or more non-diffusible ions 'X^-' are present on one side (A side) of the membrane, then according to Gibbs'–Donnan equilibrium the distribution of diffusible ions will be as under:

1. Both solutions will be electrically neutral, i.e. $(Na^+)_A = (Cl^-)_A + (X^-)_A$ and $(Na^+)_B = (Cl^-)_B$
   So, $(Na^+)_A + (Cl^-)_A + (X^-) > (Na^+)_B + (Cl^-)_B$ ..... (1)
2. The product of diffusible ions on two sides will be equal, i.e. $(Na^+)_A \times (Cl^-)_A = (Na^+)_B \times (Cl^-)_B$ ..... (2)
   From the relationship of (1) and (2), it is found that:
   - $(Na^+)_A > (Na^+)_B$ and
   - $(Cl^-)_A < (Cl^-)_B$.

Hence there is unequal distribution of diffusible ions (asymmetrical). At equilibrium, Na^+ being greater on the side which contains non-diffusible anions 'X^-' (side A) and anion Cl^- is greater on the other side (side B). However, their concentration ratio are equal.

Since the intracellular fluid (ICF) contains non-diffusible anions like proteins and organic phosphate, so, according to Gibbs'–Donnan equilibrium there should be an asymmetrical distribution of diffusible ions across the cell membrane with cations being more inside than the outside. However, in reality interior of the cell is negatively charged which will be explained in the ensuing discussion.

NERNST EQUATION

The asymmetrical distribution of diffusible ions across the cell membrane in the form of excess diffusible cation inside due to the Gibbs'–Donnan equilibrium (as explained above) results in concentration gradient. As a result of which diffusible cations (K^+) will try to diffuse back into the ECF from ICF, but it is counteracted by the electrical gradient, which will be created due to the presence of non-diffusible anions inside the cell. Thus equilibrium will be reached between the concentration gradient and the electrical gradient resulting in diffusion potential (equilibrium potential) across the cell membrane. The magnitude of this equilibrium potential can be determined by the Nernst equation as:

$$E_{(m)} = \pm 61 \log \frac{(conc)o}{(conc)i}$$

where,

$E_{(m)}$ = Equilibrium potential (in millivolts) of the ions at which efflux and influx of the ions are equal

R = The natural gas constant and its value is 8.316 joules/degree.

T = The absolute temperature

F = The Faraday constant and its value is represented as Number of coulomb/mole of charge = 96,500 coulomb/mole

Z = The valency of the ion

ln = Symbol for natural logarithm

(conc)i = The concentration of the ions in the intracellular fluid (inside).

(conc)o = The concentration of the ions in the extracellular fluid (outside).

The equilibrium potential, $E_{(m)}$, for some of the important ions in the mammalian spinal motor neuron calculated from the simplified Nernst equation is shown in Table 1.4-1.

GOLDMANN–HODGKIN–KATZ EQUATION

The Nernst equation helps in calculating the equilibrium potential for each ion individually. However, the magnitude of the membrane potential at any given time depends on the distribution of Na^+, K^+ and Cl^- and the permeability of each of these ions. The integrated role of different ions in

Table 1.4-1 Equilibrium potential, $E_{(m)}$, for important ions in a mammalian spinal motor neuron

Ion	Concentration (mmol/L of H_2O)		Equilibrium potential (mV)
	Outside the cell	Inside the cell	
Na^+	150	15	+60
K^+	5.5	150	–90
Cl^-	125	9	–70
Ca^{2+}	5	<1	+130

the generation of membrane potential can be described accurately by the Goldmann's constant field equation or the so-called Goldmann–Hodgkin–Katz (GHK) equation:

$$V = \frac{RT}{F} \ln \frac{P_k [K^+]_i + P_{Na^+} [Na^+]_i + P_{Cl^-} [Cl^-]_o}{P_k [K^+]_o + P_{Na^+} [Na^+]_o + P_{Cl^-} [Cl^-]_i}$$

V = The membrane potential
R = The gas constant
T = The absolute temperature
F = The Faraday constant and
P_{k+}, P_{Na+} and P_{Cl-} = The permeabilities of the membrane to K^+, Na^+ and Cl^-, and brackets signify concentration and i and o refer to inside and outside of the cell, respectively.

Inferences of the Goldmann constant field equation

Following important inferences can be drawn from the Goldmann constant field equation:

1. ***Most important ions for development*** of membrane potentials in nerve and muscle fibres are sodium, potassium and chloride. The voltage of membrane potential is determined by the concentration gradient of each of these ions.
2. ***Degree of importance*** of each of the ions in determining the voltage depends upon the membrane permeability of the individual ion. For example, if the membrane is impermeable to K^+ and Cl^- then the membrane potential will be determined by the Na^+ gradient alone and the resulting potential will be equal to the Nernst potential for sodium.
3. ***Positive ion concentration from inside the membrane to outside*** is responsible for electronegativity inside the membrane. This is because of the fact that due to concentration gradient, the positive ions diffuse outside leaving the non-diffusible negative ions inside the cell.
4. ***Signal transmission*** in the nerves is primarily due to change in the sodium and potassium permeability because their channels undergo rapid change during conduction of the nerve impulse and not much change is seen in the chloride channels.

ROLE OF NA^+–K^+ ATPASE PUMP

The role of Na^+–K^+ ATPase lies in building the concentration gradient. It serves to pump back the Na^+ that diffuses into the cell and K^+ that diffuses out of the cell. In the resting membrane, these diffusions are negligible, so the Na^+–K^+ pump works very feebly in this stage. Further, although the Na^+–K^+ pump potentially electrogenic (since it pumps out $3Na^+$ ions for $2K^+$ ions), at no stage the pump is able to build up a significant membrane potential. This is because of the fact that, as soon as the pump creates a negative potential inside the cell, chloride ions rush out of the cell and restore electroneutrality. Thus, in other words, the pump, pumps out $3Na^+$ ions and one Cl^- ion for every $2K^+$ ions it pumps in.

RECORDING OF MEMBRANE POTENTIAL

INSTRUMENTS USED FOR RECORDING

The essential instruments used in recording the activity of an excitable tissue are:

- Microelectrodes,
- Electronic amplifiers and
- Cathode ray oscilloscope.

Basic principles of the functioning of these instruments are described on page 55.

TECHNIQUE OF RECORDING

Technique of recording of membrane potential is described on page 56.

Chapter 1.5

Genetics: An Overview

STRUCTURAL AND FUNCTIONAL CHARACTERISTICS OF SUBSTRATE FOR GENETICS

- **Chromosomes**
- **Structure and function of DNA and RNA**
- **Genes**
  - General considerations
  - Gene expression: central dogma
  - Regulation of gene expression

APPLIED GENETICS

- **Molecular genetics and biotechnology**
  - Genetic engineering/recombinant DNA technology
  - Polymerase chain reaction
  - Blotting techniques
  - Cloning
  - Apoptosis
- **Molecular genetics and medicine**
  - Mutations and genetic human diseases
  - Genetic screening
  - Genetics and cancer
  - Gene therapy

Genetics may rightly be claimed to be one of the most important branches of biology. Foundation for the present day genetics was laid by the Mendel's work published in 1866. He demonstrated that characteristics do not blend but pass from parents to offsprings as discrete (separate) units. These units, which appear in the offspring in pairs, remain discrete and are passed on to subsequent generations by the male and female gametes each of which contain a single unit. The Danish botanist Johannied called these units genes in 1909 and the American geneticist Morgan, in 1912, demonstrated that they are carried on chromosomes. Since early 1900, the study of genetics has made great advances. An overview of which is given here in brief.

STRUCTURAL AND FUNCTIONAL CHARACTERISTICS OF SUBSTRATE FOR GENETICS

CHROMOSOMES

Waldeyer in 1888 coined the term chromosomes to denote the thread-like structures present in the nucleus of eukaryotic cells during division. It is now established that the chromosomes are responsible for the transmission of the hereditary information from one generation to next. There are 46 chromosomes (23 pairs in all the dividing cells of the body except the gametes (sex cells which contain only 23 chromosomes) haploid number.

Morphology of chromosomes

Each chromosome is composed of two chromatids connected at the *centromere.* Each chromatid consists of two *chromonemes. Telomeres* are the terminal ends of chromosomes DNA molecule.

Morphological types of chromosomes. Depending upon the location of centromere four morphological types of chromosomes are recognized (Fig. 1.5-1):

- *Metacentric chromosomes* in which the centromere divides the chromosomes into two equal arms (Fig. 1.5-1A).
- *Submetacentric chromosomes.* The centromere divides the chromosomes into two unequal arms (Fig. 1.5-1B).
- *Acrocentric chromosomes.* The centromere is located in such a way that a very short arm of chromosomes is visible (Fig. 1.5-1C).
- *Telocentric chromosomes.* The centromere is located at one end (Fig. 1.5-1D).

Functional types of chromosomes. There are three types of eukaryotic chromosomes:

- *Autosomes* are the chromosomes present in somatic cells. The number of autosomes in a cell is fixed and is expressed as 2n or *diploid number.*

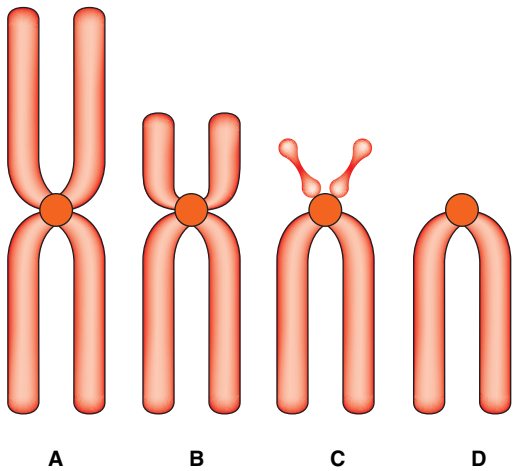


Fig. 1.5-1 Morphological types of chromosomes: A, metacentric; B, submetacentric; C, acrocentric and D, telocentric.

- *Sex chromosomes* are present in the sex cells and are responsible for determining the sex of individual.
- *Supernumerary or redundant chromosomes* are also found in eukaryotic cells but their occurrence is quite uncommon.

Chemical structure of chromosome

The chromosomes are mainly composed of deoxyribonucleic acid (DNA). The chromosome also contains ribonucleic acid (RNA), basic proteins called histones, complex proteins including enzymes, some organic phosphorus compounds and inorganic salts. The amount of DNA in a haploid cell is half the amount present in a diploid cell of the same species. Further, the concentration of DNA in any cell remains constant in every circumstances. An important feature of DNA is that it is metabolically stable.

Organization of DNA in a chromosome

See page 30.

STRUCTURE AND FUNCTION OF DNA AND RNA

DNA

DNA, i.e. deoxyribonucleic acid is a molecule of inheritance and thus may be regarded as the reserve bank of genetic information. DNA is exclusively responsible for maintaining the identity of different species of organisms for millions of years.

Structure of DNA

DNA is a polymer of four monomeric deoxyribonucleotides, namely deoxyadenylate (dAMP), deoxyguanylate (dGMP), deoxycytidylate (dCMP) and deoxythymidylate (dTMP). Each deoxyribonucleotide in turn is composed of

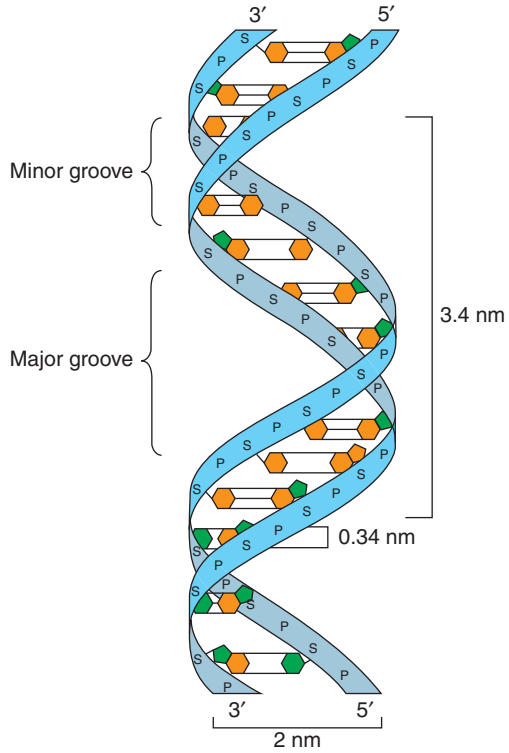


Fig. 1.5-2 Watson–Crick model of DNA structure.

a nitrogenous base purines or pyrimidines (A, G, C or T), a pentose sugar, i.e. a deoxyribose and a phosphate. Each molecule of DNA has equal number of adenine and thymine residues (A=T) and equal number of guanine and cytosine residues (G=C). This is known as *Chargaff's rule.*

Watson–Crick model of DNA structure. The salient features of Watson–Crick model of DNA (now known as B-DNA) are (Fig. 1.5-2):

- *Double helix structure.* Each DNA molecule is right handed double helix composed of two polydeoxyribonucleotides chains (strands) twisted around each other on a common axis.
- *Antiparallel chains.* The two chains of each DNA molecule are antiparallel, i.e. one chain runs in the 5′ to 3′ direction while the other in 3′ to 5′ direction.
- *Dimensions.* The width of a double helix is 20 Å (2 nm). Each turn (pitch) of the helix contains 10 pairs of nucleotides, each placed at distance of about 3.4 Å (0.34 nm), thus each turn is 34 Å (3.4 nm) in dimension.
- *Arrangement of base, sugar and phosphate molecule.* Each chain has a sugar phosphate backbone with bases which project at right angles and hydrogen bond with the bases of the opposite chain across the double helix (Fig. 1.5–3).
- *Complementary chains.* The two polynucleotide chains are not identical but complementary due to base pairing.
- *Genetic information.* The genetic information resides in one of two strands known as template strand or sense strand. The opposite strand is antisense strand.

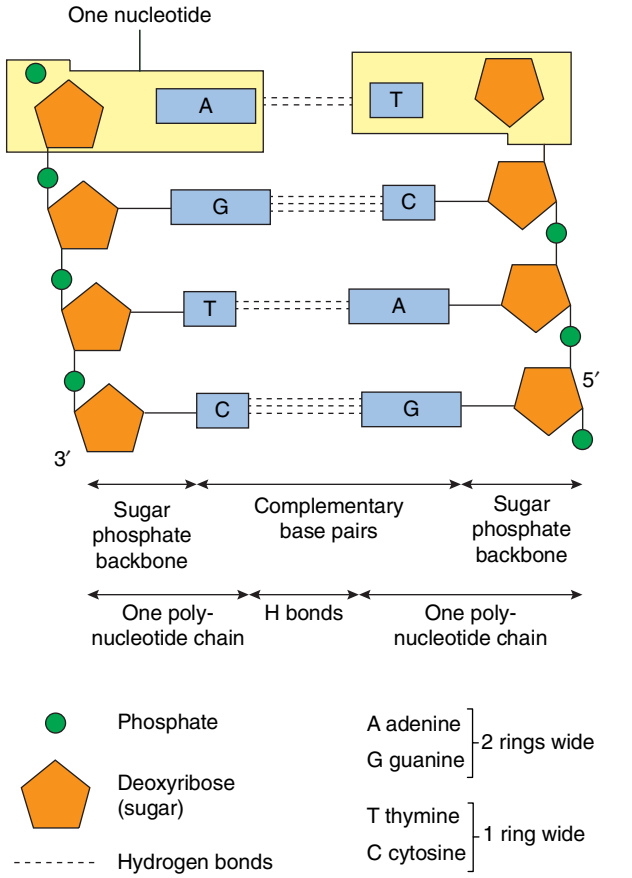


Fig. 1.5-3 Diagrammatic structure of straightened chains of DNA.

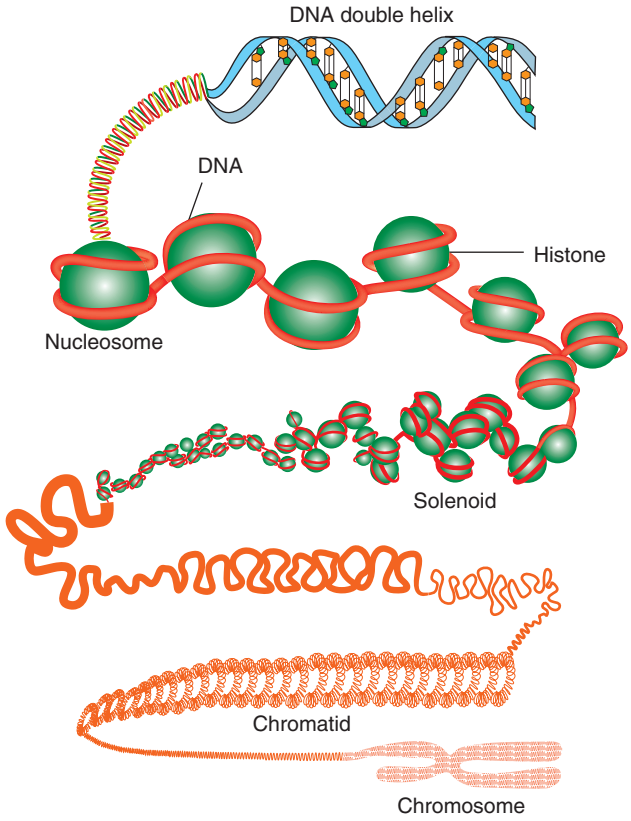


Fig. 1.5-4 Diagrammatic organisation of DNA in a chromosome of the human cell.

Size of DNA

DNA molecules are huge in size. The term kilobase pair (Kb = 1000 base pairs) is commonly used in DNA structure. In humans 23 haploid chromosomes have 2,900,000 Kb with a total contour length of 990 nm. Thus a human cell contains about 2 m of DNA distributed among 46 chromosomes. Each chromosome, therefore, contains about 4.8 cm (48.000 µm) of DNA. Human chromosomes are on average about 6 µm long, a packing ratio of 8000:1. In order to maintain a high degree of organisation when the DNA is folded, the histone proteins form precise architectural scaffolding for the DNA.

Organization of DNA in the cell

In human cells the DNA is found in association with positively charged protein molecules called *histones.* Each DNA helix combines with group of eight histone molecules to form structures known as *nucleosomes* which have an appearance of 'beads and string'. These nucleosomes, and the DNA strands linking them, are packed closely together to produce a 30 nm diameter helix with about six nucleosome per turn. This is known as 30 nm fibre or the *solenoid fibre.* The solenoid fibres in turn coil to form *chromatin fibres* which are further coiled and packed in the form of *chromatin* in which form DNA is present in the chromosome (Fig. 1.5-4).

RNA

Structure of RNA

RNA is a polymer of ribonucleotides held together by 3′,5′-phosphodiester bridges. Though RNA molecule like that of DNA is composed of nucleotides consisting of a base sugar and phosphate but has following structural differences:

- *Single strand.* RNA is commonly single stranded structure unlike DNA. However in certain forms of RNA this strand may fold at certain places to give a double stranded structure if complementary base pairs are in close proximity.
- *Ribose sugar.* The sugar molecule in a RNA molecule is ribose in contrast to deoxyribose.
- *Base.* The pyrimidine base in a RNA molecule is uracil in place of thymine of a DNA molecule.
- *Chargaff's rule.* Due to the single stranded structure Chargaff's rule is not obeyed, i.e. there is no specific relation between purine and pyrimidine contents.

Types of RNA

Following types of RNAs have been recognized:

- ***Messenger RNA (mRNA).*** In the human cell it is synthesized in the nucleus and enters the cytoplasm to participate in protein synthesis.
- ***Transfer RNA (tRNA).*** There are about 20 species of tRNA corresponding to 20 amino acids present in protein structure. The structure of tRNA resembles that of clover leaf with four arms. tRNA delivers amino acids for protein synthesis.
- ***Ribosomal RNA (rRNA).*** rRNAs are present in ribosomes (factories of protein synthesis). It is believed that rRNAs play a significant role in binding of mRNA to ribosomes in protein synthesis.

DNA REPLICATION

DNA replication is a process by which each original DNA molecule gives rise to two copies with identical structure. The method by which the DNA replicates is called *semiconservation replication* since each new double helix retains (conserves) one of the two strands of the original DNA double helices. Steps involved in the DNA replication are (Fig. 1.5-5):

1. Initiation of replication. The site from where the replication of DNA is initiated is called *origin of replication.* In prokaryotes, DNA replication initiates from only one site hence called monorepliconic replication and in eukaryotes it starts from multiple sites *(multirepliconic replication).* The origin of replication mostly consists of A–T base pairs. When a specific binding protein (DNA protein) binds to the site of replication then there occurs separation of double stranded DNA, and separated strands of DNA form a bubble at the site of origin.

2. Formation of replication fork and replication eye. The next step in the DNA replication is unwinding of double helix leading to formation of either Y-shaped *replication fork,* (when DNA replication initiates from the terminal end of the double helix), or θ-shaped replication eye, (when DNA replication starts from the intercalary position). This step is controlled by an enzyme called *helicase* and a protein called *single strand binding* (SSB) protein.

- *Role of DNA helicases.* These enzymes bind to both the strands of DNA at replication fork and move along the DNA helix and separate the strands of the DNA double helix. The function of helicases can be compared to a zip opener.
- *Role of single strand DNA binding (SSB) proteins.* As the name indicates SSB protein binds only to single stranded DNA (separated by helicase). Main function of this protein is to keep the two DNA strands separate hence also called helix destabilizing protein. SSB protein also provides template for new DNA synthesis and prevent degradation of single stranded DNA.

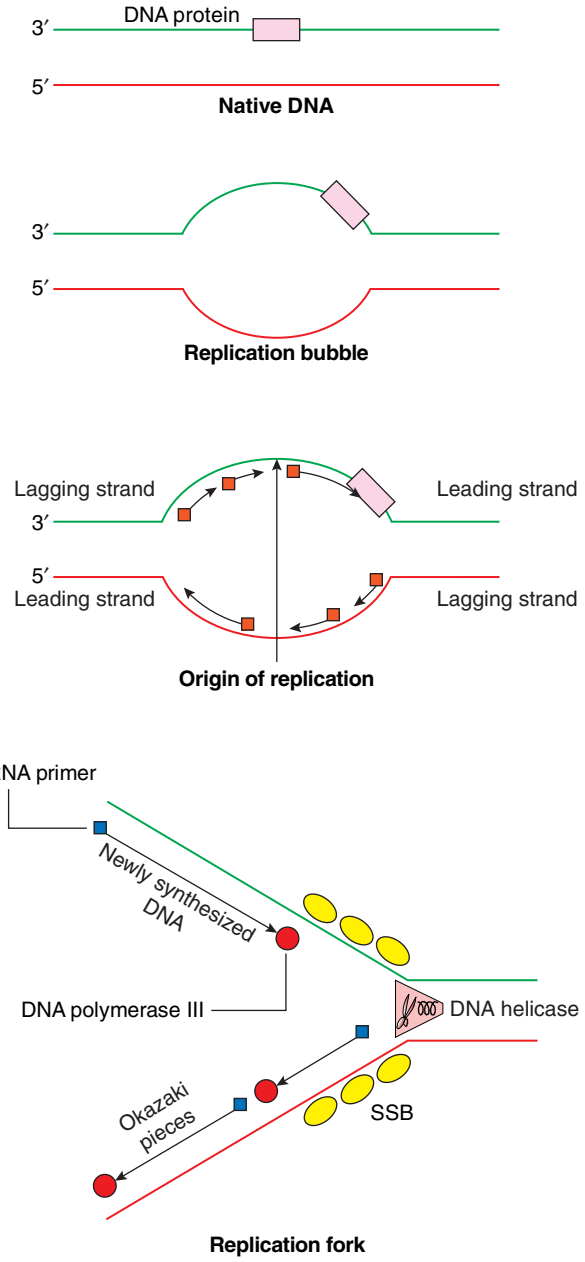


Fig. 1.5-5 Simplified diagram showing main steps of DNA replication.

3. Formation of RNA primer. RNA primer consists of a short fragment of RNA (about 5–50 nucleotides). It is required for synthesis of new DNA. The RNA primer is synthesized on DNA template by specific *RNA polymerase* (primase).

4. DNA synthesis along the replication fork. DNA replication occurs simultaneously in both the leading as well as lagging strands of Y-shaped replication fork and is of two types:

- *Continuous DNA replication.* In the leading strand, DNA polymerase III binds to the single stranded DNA

and starts to move along the strand. Each time it meets the next base on DNA, free nucleotides approach the DNA strand, and one with the correct complementary base hydrogen bonds to the base in the DNA. The free nucleotide is then in place by the enzyme until it binds to the preceding nucleotide thus extending the new strand of DNA. The enzyme continues to move along one base at a time with new DNA strand growing as it does so.

- *Discontinuous DNA replication.* Occurs in the lagging strand.

GENES

GENERAL CONSIDERATIONS

The gene is the functional unit of DNA. A gene could therefore be defined as a piece of DNA which codes for a protein. In strictest sense the gene can be defined as the DNA code for a polypeptide. Since some proteins are made up of more than one polypeptide chains and are therefore coded for by more than one genes.

Genome. The term genome refers to total genetic information contained in a cell.

Human genome. For human, the genome is essentially equivalent to all of the genetic information which is present in a single set of 23 chromosomes.

Human genome project (1990–2003). The human genome project which completed on April 14, 2003 has accomplished the following goals:

- Identified all the *approximate 30,000 genes in human DNA.*
- Determined the sequences of 3 billion chemical base pairs that make the human DNA.

Functional genomics. Understanding the functions of genes and other parts of genome is known as functional genomics.

Comparative genomics. Comparative genomics is the analysis and comparison of genome from different species.

Constitutive and inducible genes. The genes are generally considered under two categories:

- *Constitutive.* The products (proteins) of these genes are required all the time in a cell. Therefore, the constitutive genes (or housekeeping genes) are expressed more or less at constant rate in almost all the cell and, further, they may not be subjected to regulation, e.g. enzymes of citric acid cycle.
- *Inducible genes.* The concentration of the proteins synthesized by inducible genes is regulated by various molecular signals. An inducer increases the expression of these genes while a repressor decreases, e.g. tryptophan pyrrolase of liver is induced by tryptophan.

GENE EXPRESSION: CENTRAL DOGMA

As mentioned above, each cell of human body contains entire genome, yet the genetic expression is very selective and different patterns of protein synthesis occur in different tissues. Not only this, even in the same tissue there is wide variation in the proteins produced during the course of development.

The expression of genetic material occurs through the production of proteins. This involves two consecutive steps—transcription and translation. In transcription, the genetic information, stored in DNA, is transferred to an RNA intermediate, which in turn uses this information to direct the synthesis of proteins during translation. This unidirectional flow of information was described by F. H. C. Crick in 1958 as the *central dogma* of molecular biology (Fig. 1.5-6). However, an important modification of this information flow was given by David Baltimore and H. Temin, who described reversible sequence through reverse transcription or teminism in the presence of transcripts (revised central dogma).

Transcription

Transcription is a process in which RNA is synthesized from DNA. All the three types of RNAs (mRNA, tRNA and rRNA) are produced through transcription. The transcription process is selective, i.e. the entire molecule of DNA is not expressed in transcription, but the RNAs are synthesized only for selected regions of DNA. The strand of DNA that directs the synthesis of mRNA via complementary base pairing is called the *template strand* or coding or sense

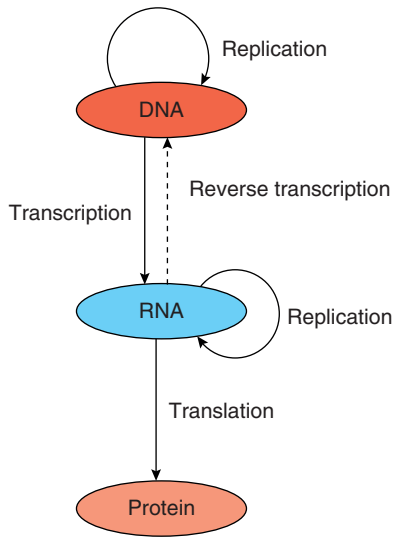


Fig. 1.5-6 Central dogma: the flow of genetic information.

strand, and the other strand is known as noncoding strand or antisense strand. Transcription is accomplished by an enzyme *RNA polymerase* that gets physically associated with DNA. Only one type of such an enzyme is found in prokaryotes in contrast to eukaryotes (where three different forms of RNA polymerase are found). RNA I, II and III catalyze the synthesis of rRNA, mRNA and tRNA, respectively.

Promotor sites. RNA polymerase binds to a region of DNA called promotor site. In eukaryotes, a sequence of DNA bases has been identified. This sequence, known as *Hogness box* or TATA box (Fig. 1.5-7), is located on left about 25 nucleotides away (upstream) from the starting site of mRNA synthesis. There also exists another site of recognition between 70 and 80 nucleotides upstream from the start of transcription. This second site is referred to as *CAAT box*. One of these two sites (or sometimes both) helps RNA polymerase II to recognize requisite sequence of DNA for transcription.

Salient features of transcription in eukaryotes vis-a-vis prokaryotes are:

- Transcription in eukaryotes unlike prokaryotes occurs within the nucleus and mRNA moves out of the nucleus into the cytoplasm for translation.
- The initiation and regulation of transcription in eukaryotes is more extensive than prokaryotes.

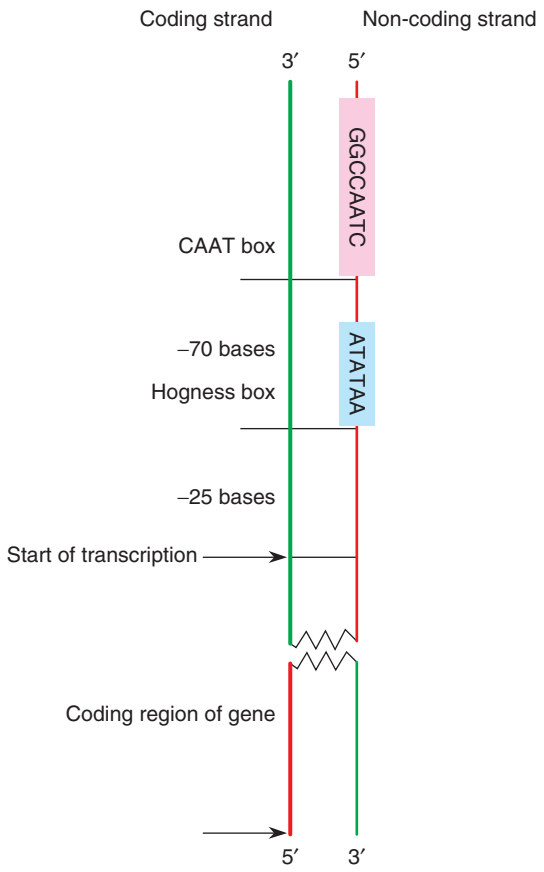


Fig. 1.5-7 Promoter sites of DNA in eukaryotes.

Post-transcriptional modifications

- The mRNA in eukaryotes is processed from the primary RNA transcript; a process called maturation which includes:
- Releases of the introns and joining with two adjacent exons to produce mature mRNA.
- *RNA editing.* Besides, these two post-transcriptional modifications, RNA editing may also take place before translation begins.

Reverse transcription refers to formation of DNA from RNA. The enzyme reverse transcriptase is responsible for this process. The DNA so formed is complementary (cDNA) to viral RNA can be transmitted to host DNA. Reverse transcription is known to occur in retroviruses which include human immunodeficiency virus that cause AIDS.

Translation: Biosynthesis of proteins

Translation is the process by which genetic message carried by mRNA from the DNA is converted in the form of a polypeptide chain having specific sequence of amino acids. Before discussing the process of translation, it will be worthwhile to know something about genetic code.

Genetic code. The process by which the information coded in the mRNA is decoded into polypeptide is referred to as *deciphering the genetic code.* Dr. Hargobind Khorana shared 'Nobel Prize' in 1968 with Nirenberg and Holly for the discovery of genetic code. The genetic code (codons) is formed by three nucleotides (triplet) base sequences in mRNA. The codons are formed of four nucleotide base (A, G, C and U). These four bases produce 64 different combinations of three base codons. Of the 64 codons, the 61 codons code for 20 amino acids found in proteins and the three codons (UAA, UAG and UGA) are *termination codons* which act as stop signals in protein synthesis. The codons AUG and sometimes GUG act as *initiating codons.*

Characteristics of genetic code are:

- *Universality,* i.e. same codons are used to code for the same amino acids in all the living organisms with a few exceptions.
- *Specificity,* i.e. a particular codon always codes for the same amino acid, e.g. AUG is the codon for methionine.
- *Non-overlapping,* i.e. the genetic code is read from a fixed point as a continuous base sequence.
- *Degenerate,* i.e. one amino acid is coded by more than one codons. The codons that designate the same amino acid are called synonyms.

Process of protein biosynthesis. The process of protein synthesis in addition to mRNA requires amino acids,

tRNA, energy sources (ATP and GTP) and protein factors. Protein synthesis occurs over ribosomes which are also called *protein factories.* The protein biosynthesis involves three processes:

I. ***Activation of amino acids.*** Amino acids are activated and attached to tRNA in a two-step reaction. A group of enzymes, namely aminoacyl tRNA synthetases are required for this process. In the first step an amino acid reacts with ATP in the presence of specific amino acid tRNA to form *enzyme–AMP–amino acid complex*. This complex then reacts with a specific tRNA and the amino acid is transferred to 3′ end of the tRNA to form aminoacyl tRNA.

II. ***Translation proper*** involves three steps—initiation, elongation and termination.
   1. *Initiation.* The translation of mRNA begins with the formation of initiation complex.
   2. *Elongation.* Ribosomes elongate the polypeptide chain by a sequential addition of amino acids to the growing carboxyl end.
      The elongation process is repeated again and again with addition of one amino acid each time till signal for termination is reached.
   3. *Termination* of polypeptide synthesis is evoked by a nonsense or termination codon (UAA, UAG or UGA).

III. ***Post-translational modifications.*** The proteins synthesized in translation are, as such not functional. Many changes take place in the polypeptides after the initiation of their synthesis or, most frequently, after the protein synthesis is completed. Post-translational modification include:
   - Proteolytic degradation, and
   - Covalent modifications (Phosphorylation, hydroxylation and glycosylation).

REGULATION OF GENE EXPRESSION

As discussed earlier, each nucleated somatic cell in the body contains full genetic message, yet there is great differentiation and specialization in the functions of various types of adult cells. It is because of the fact that there exists a full proof system for regulation of gene expression that maintains orderly growth in cells and prevents uncontrolled growth. The genes are controlled both spatially and temporally. The regulation of gene expression is thus absolutely essential for growth, development and differentiation of an organism. A positive regulator increases the gene expression whereas a negative regulator decreases.

Regulation of gene expression in prokaryotes

In prokaryotes, gene expression is regulated by operon system. Operon are segments of genetic material which function as regulated units that can be switched on and switched off. The operon systems are of two types:

1. ***Inducible operon system*** (Lac operon system). An inducible operon system is that regulated genetic material which remains switched off normally but becomes operational in the presence of an inducer. It occurs in catabolic pathway.
2. ***Repressible operon system*** (Tryptophan operon system). A repressible operon system is that regulated genetic material which normally remains active/operational. It usually occurs in anabolic pathways.

Regulation of gene expression in eukaryotes

The regulation of gene expression in eukaryotes is very complex and involves various mechanisms. Some of the mechanisms are:

1. ***Gene amplification.*** In this mechanism, the expression of gene is increased several folds. An example of gene amplification in humans includes development of drug resistance by the malignant cells to long-term administration of methotrexate. This occurs by amplifying the gene coding for dihydrofolate reductase.
2. ***Gene rearrangement.*** The process of gene rearrangement is responsible for the generation of 10 billion antigen specific immunoglobulins.
3. ***Regulation of gene expression through transcription factors.*** Transcription factors are products of other genes and hence mediate transregulation by binding to specific DNA segments. This specific interaction of protein to DNA in over 80% of the non-transcription factors is brought about by one of the four DNA-binding motifs: zinc finger motif, Lucine zipper-motif, helix-turn-helix motif and helix-loop–helix motif.
4. ***Regulation of gene through mRNA.*** Gene expression is also regulated by regulation of synthesis transport, processing and stability of mRNA.

APPLIED GENETICS

Applications of genetics are many more and beyond the scope of a book on physiology, only a few of interest described here include some aspects of:

- *Molecular genetics and biotechnology, and*
- *Molecular genetics and medicine.*

MOLECULAR GENETICS AND BIOTECHNOLOGY

Biotechnology involves use of living organisms or products of living organisms for the welfare of human. Presently, the

molecular biology has combined with genetics to give us more powerful biotechnology. The tools of molecular genetics have provided improved way to make use of living organisms for the benefit of humans. The subject of molecular genetics and biotechnology is expanding fast and has already become very vast. The present discussion includes only a few important aspects viz.

- Genetic engineering,
  - Stages of recombinant DNA technology,
  - Application of recombinant DNA technology,
- Polymerase chain reaction,
- Blotting techniques,
- Cloning,
- Apoptosis.

GENETIC ENGINEERING/RECOMBINANT DNA TECHNOLOGY

The terms genetic engineering/recombinant DNA technology/DNA cloning/molecular cloning/gene cloning all refer to the process of transfer of a DNA fragment of interest from one organism to a self-replicating genetic element, such as a bacterial plasmid. In other words, this technology involves cutting, modifying and joining DNA molecules using enzymes, such as restriction enzymes and DNA ligase.

Stages of recombinant technology

Stages of recombinant technology are:

Stage 1: Generation of a copy of gene required. This is the most difficult part of the process. Three methods are used to get a copy of a gene:

- Making a copy of the gene from its mRNA using reverse transcriptase,
- Synthesising the gene artificially and
- Using a *shotgun approach* which involves chopping up the DNA with *restriction enzymes* and searching for the piece with the required gene (Fig. 1.5-8).

Stage 2: Joining the gene to a vector or carrier molecule. A vector is a carrier of a DNA molecule to which the generated gene is attached for cloning. Three commonly used vectors in recombinant DNA technology are (Fig. 1.5-8):

- Plasmids,
- Bacteriophages and
- Cosmids.

Stage 3: Introduction of vector DNA into the host cell to produce chimeric DNA. The main aim of genetic engineering is to insert a DNA of interest (generated gene) into a vector DNA so that the DNA fragment replicates along with the vector after annealing. This hybrid combination of two fragments of DNA is referred to as *chimeric DNA* or *hybrid DNA* or *recombinant DNA*.

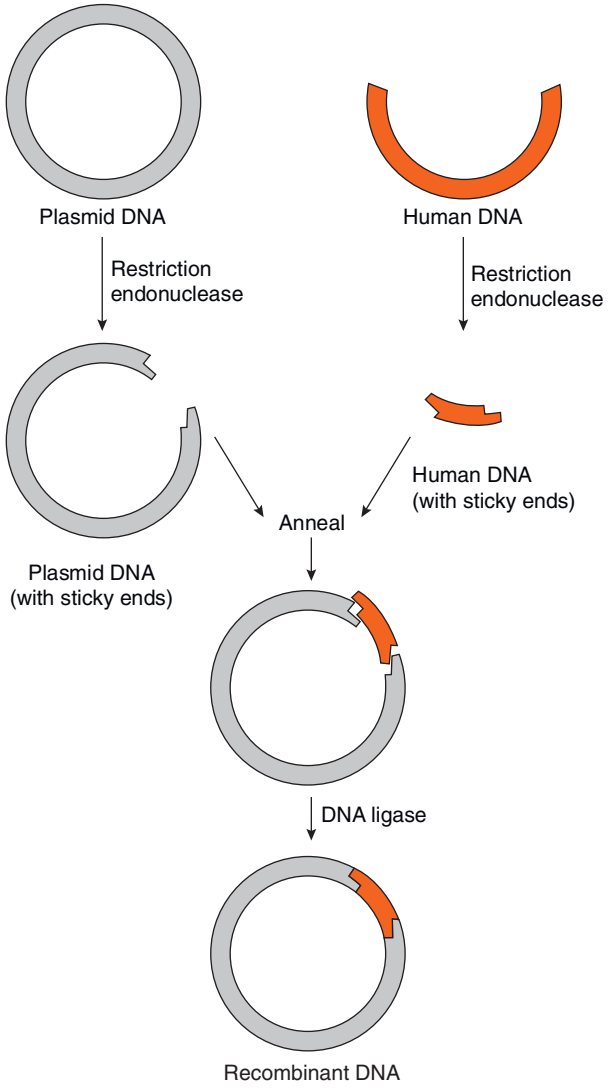


Fig. 1.5-8 Stages of recombinant DNA technology.

Stage 4: Cloning of chimeric DNA (Fig. 1.5-9). A clone is a large population of identical molecules, bacteria or cells that arise from a common ancestor. The chimeric DNA contained in a plasma (vector) can be introduced into bacterial cells by a process called *transfection.* The replicating bacterial cell (host cell) permits the amplification of the chimeric DNA of the vector. In this way, cloning results in the production of large number of identical target DNA molecules. The cloned target DNA is released from the vector by cleavage (using appropriate restriction endonucleases), isolated characterized and used for various purposes.

Applications of genetic engineering

The genetic engineering has revolutionized the application of molecular biology to medical/agricultural sciences that

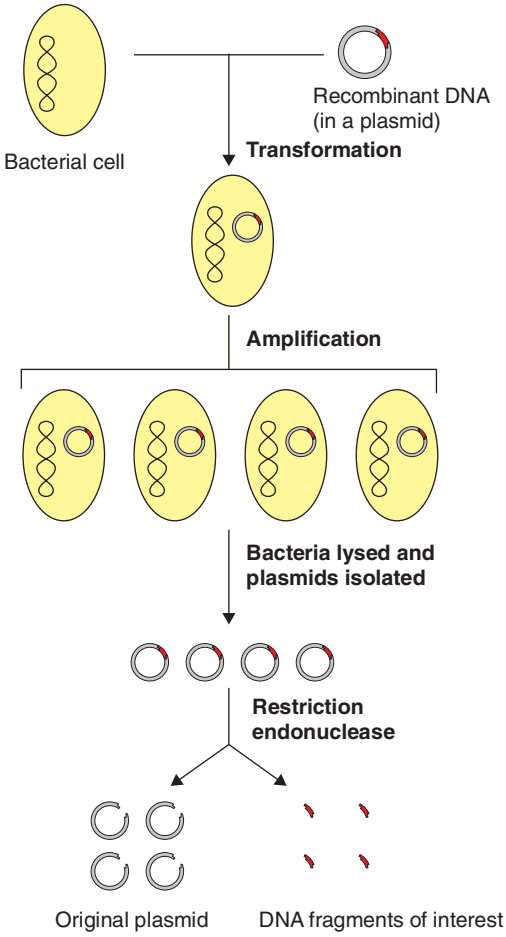


Fig. 1.5-9 Cloning of a recombinant DNA.

has immensely benefitted the mankind. A few important applications are:

1. *Production of proteins/hormones.* It is possible to produce proteins/hormones in large amount for therapeutic purposes. These include insulin, growth hormone, erythropoietin, interferons, vaccines and blood clotting factors.
2. *Molecular analysis of diseases* such as sickle cell anaemia, thalassaemia, cystic fibrosis using recombinant DNA technology has led to better understanding of these diseases.
3. *Laboratory diagnostic applications.* Using this technique the diagnosis of diseases like AIDS has become simple and rapid.
4. *Gene therapy* for correcting a genetic defect is very useful application of RBT (see page 41).
5. *Prenatal diagnosis of genetic diseases* such as sickle cell anaemia is possible from the DNA collected from the amniotic fluid by using DNA probes.
6. *Transgenesis* refers to transfer of genes into the fertilized ovum which will be found in somatic as well as germ cells and passed on to successive generation.
7. *Application in forensic medicine: DNA fingerprinting* applying southern blot technique is useful in identifying criminals and settling the disputes of parenthood of children.
8. *Industrial applications.* Enzymes synthesized by this technology are used to produce sugar, cheese and detergents.
9. *Agricultural applications* include development of genetically engineered plants to increase the yield of crops, to resist draught and to resist diseases.
10. *Evolution.* This technique helps in bridging several missing links in the evolution by amplifying the DNA by polymerase chain reactions (PCRs) from the archeological sample of extinct animals.

POLYMERASE CHAIN REACTION

Polymerase chain reaction (PCR) is a sensitive, selective and extremely rapid method of amplifying a target sequence of DNA.

Technique of PCR involves following steps:

- *Denaturation of DNA* refers to separation of double stranded target DNA into single strand by heating.
- *Annealing with two primers* (one for each strand) is allowed to occur after cooling the single DNA strand.
- *DNA amplification* occurs by synthesis of new DNA strand in the presence of enzyme DNA polymerase and the substrates deoxyribonucleotide triphosphates. These strands are compliments to the target DNA. The cycle of DNA amplification is repeated again and again. The Taq DNA polymerase is *heat resistant;* this special feature makes it suitable for automation of PCR. PCR results in the amplification of target DNA by about a million (10^6) fold with high specificity.

Application of PCR. PCR is highly sensitive, it can detect even the presence of single molecule of DNA. PCR amplification of sequences of human genomic DNA produces an amount of target DNA (up to 1 mg) which is sufficient for direct application in any one of a wide range of molecular biological procedures including *direct DNA sequencing.*

PCR has several applications:

- Rapid diagnosis of AIDS.
- DNA fingerprinting is useful in kinship analysis and in the identification of crime suspects.
- Prenatal diagnosis of genetic diseases.
- Study of evolution from DNA of archeological samples.
- Sex identification.

BLOTTING TECHNIQUES

Blotting techniques refer to the analytical techniques used for the identification of a special DNA, or RNA or a protein. These include:

- Southern blotting (for DNA),
- Northern blotting (for RNA) and
- Western blotting (for protein).

Southern blotting

Southern blotting technique is named after the scientist who identified it while the northern and Western blot technique are the laboratory jargons which are now accepted.

Applications. Important applications of Southern blotting are:

- DNA fingerprinting and
- Detection of mutant gene causing cystic fibrosis.

Northern blotting

Northern blotting is similar to Southern blotting except that it is for RNA instead of DNA.

Applications. Important applications of northern blotting are analysis of the expression of a gene in a particular tissue.

Western blotting

Western blotting is the technique for identification of a specific protein.

Applications. Western blot test is widely used as confirmatory test of HIV. In combination with a positive enzyme-linked immunosorbent assay (ELISA), a positive western blot is 99.9% accurate in detecting HIV infection.

CLONING

Cloning refers to making many identical copies of a molecule.

Certain terms which need to be defined before discussing the different types of cloning include:

- *Transformation* refers to introduction of any DNA molecule into any living cell.
- *Transfection* refers to introduction of purified DNA molecules into cultured cells.
- *Transduction* is the transfer of genetic material or DNA from one cell to another with the help of a virus.
- *Microinjection* is a method of introducing new DNA into a cell by injecting it directly into the nucleus.
- *Biolistics* refers to the means of introducing DNA into cells that involves bombardment with high-velocity microprojectiles located with DNA.

Types of cloning. Some important types of cloning are:

- Gene cloning/DNA cloning,
- Reproductive cloning,
- Therapeutic cloning and
- Tissue culture.

Gene cloning

Gene cloning refers to the process of transfer of a DNA fragment of interest from one organism to a self-replicating genetic element (cloning vector) such as bacterial plasmid and subsequent propagation of the recombinant DNA molecule in the host organism (Fig. 1.5-9). Other terms used to denote gene cloning are: DNA cloning and recombinant DNA technology.

Details of this process has already been described on page 35.

Applications of gene cloning, i.e. recombinant DNA technology have already been highlighted (see page 35). Important genetic applications include:

- Gene therapy,
- Gene engineering of organisms and
- Sequencing genomes.

Reproductive cloning

Reproductive cloning is a technology used to generate an animal that has same nuclear DNA as another currently or previously existing animal.

Technique. Reproductive cloning uses the technique of somatic cell nuclear transfer. In this technique genetic material is transferred from the nucleus of a donor adult cell to an egg whose nucleus, and thus its genetic material, has been removed. A sheep (Dolly) was created by somatic cell nuclear transfer process by the research team at the Roslin Institute in Edinburgh (Scotland) in Feb 1997.

Applications. Reproductive cloning can be used to repopulate endangered animals or animals that are difficult to breed. Examples are:

- In 2001, the first clone of an endangered animal, a wild ox called a *gaur* was born. The young gaur died from an infection about 48 h after its birth.
- In Italy (2001) a successful cloning of a healthy baby *mouflon,* an endangered wild sheep was reported. The cloned mouflon is living at a wild centre in Sardinia.

Embryo cloning (Therapeutic cloning)

- Therapeutic cloning refers to the production of human embryos for use in research.

- This process is aimed at harvesting stem cells that can be used to study human development and to treat diseases rather than to create a cloned human being.
- Stem cells are extracted from the egg at the blastocysts stage of development and can be used to generate virtually any type of specialized cells in the human body.
- First human embryo for the purpose of therapeutic research was cloned in Nov, 2001 by the scientists from 'Advanced Cell Technologies (ACT), a biotech company in Massachusetts.

Applications. Stem cells can be used to serve as replacement cells to treat heart disease, Alzheimer's diseases, cancer and other diseases.

Tissue culture

Cells may also be cloned for special purposes. This technique is called *tissue culture.* Certain cells when placed in a suitable medium can be cultured indefinitely.

Applications. The use of cloned cells allows a study of the action of such chemicals as hormones, drugs, antibiotics, cosmetics and pharmaceutical products to be made on cells. Such a technique is useful substitute for laboratory animals, such as rats, cats and dogs.

APOPTOSIS

Apoptosis, also called as programmed cell death, occurs under genetic control. Since the cells's own genes play an active role in its demise, so apoptosis can also be called *cell suicide.*

Need and examples of apoptosis. Apoptosis is a very common process during development as well as adulthood, the need of which is highlighted with some examples:

1. Apoptosis for proper development of tissues
- *In central nervous system* large number of neurons are produced and then die during the remodelling that occurs during development and synapse formation.
- *During formation of the fingers and toes* of the fetus, the apoptosis plays an important role of removing the web tissue between the finger and toes.
- *During sexual development* in fetal life the apoptosis is responsible for regression of duct systems.

2. Apoptosis for normal functioning of adult tissues. Examples are:
- *Cyclic breakdown of endometrium* that leads to start of menstruation is caused by apoptosis.
- *Epithelial cells* that lose their connection to the basal lamina and surrounding cells undergo apoptosis.
- *Enterocytes sloughed off* the tips of intestinal villi undergo apoptosis.

3. Apoptosis to destroy the cells that represent a threat to the integrity of the organism. Examples are:
- Apoptosis of the cells infected with virus.
- Apoptosis of the cells of the immune system to prevent them from attacking body constituents.
- Apoptosis of the cells with DNA damage.
- Apoptosis of the cancer cells.

Mechanism of apoptosis

The final common pathway leading to apoptosis is the activation of a group of cysteine proteases called *caspases* which exist in cells as inactive proenzymes (Fig. 1.5-10).

Triggering stimuli. The apoptosis can be triggered by external and internal stimuli (Fig. 1.5-10).

- *Internal stimuli.* One of the important pathways goes through the mitochondria, which release *cytochrome* and a protein called smac/DIABL0, causing activation of the *caspase 9.*
- *Apoptosis's inducing factor* located in the intermembrane space of mitochondria, migrates to nucleus and destroys DNA after binding leading to cell death.
- *External stimuli.* One ligand that activates receptors triggering apoptosis is Fas (a transmembrane protein that project from natural killer cells and T lymphocytes but also exists in circulating form). Another ligand is tumour necrosis factor (TNF). These activate the enzyme caspase 8 followed by cascade of caspase activation.

Net result of caspases activation is DNA fragmentation, cytoplasmic and chromatin condensation and eventually membrane bleb formation with cell break-up and removal of the debris by phagocytosis (Fig. 1.5-10).

MOLECULAR GENETICS AND MEDICINE

Clinical applications of molecular genetics in medicine are rapidly increasing owing to research and advances in the molecular aspect of genetics, regulation of gene expression and protein synthesis. Some of the aspects in relation to molecular genetics and medicine described here are:

- Mutations and genetic human diseases.
- Detecting human genetic variations (genetic screening).
- Genetics and cancer.
- Gene therapy.

MUTATIONS AND GENETIC HUMAN DISEASES

Mutations

Mutation refers to a change in the DNA structure of a gene. The substances or factors which are responsible for the

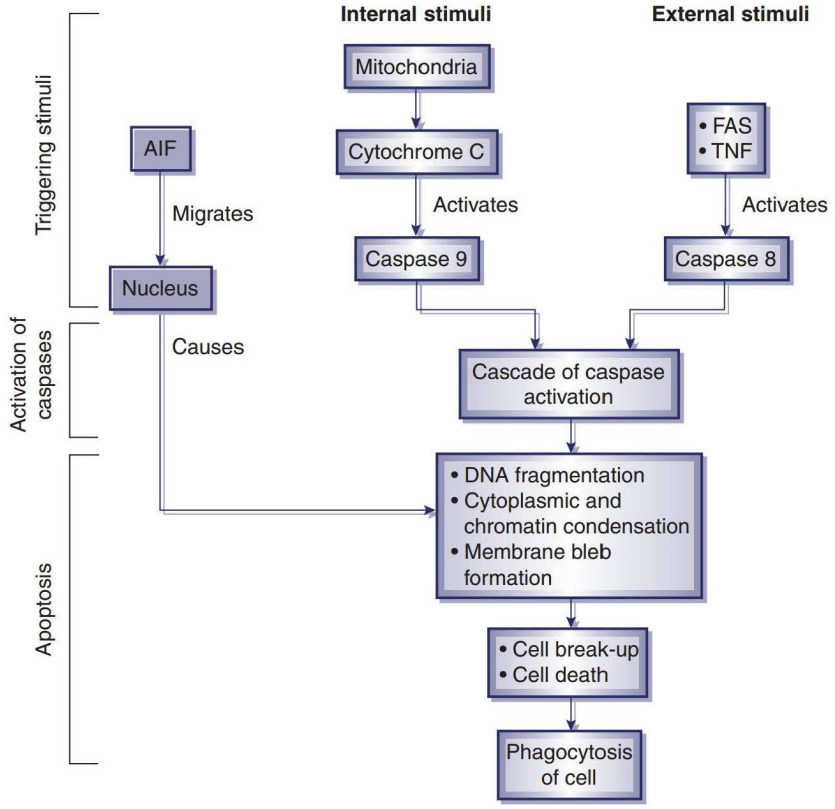


Fig. 1.5-10 Steps involved in apoptosis.

mutations are called *mutagens,* e.g. X-rays, ultraviolet light, certain chemicals etc.

There are two major types of mutations:

1. ***Point mutations.*** In this type one base pair of DNA is replaced by another.
2. ***Frame shift mutations.*** When one or more than one base pairs are either deleted or inserted into the DNA of gene. Therefore frame shift mutations are also called deletion or insertional mutations.

Genetic human diseases

Salient aspects of some common genetic human diseases associated with mutations are summarized in Table 1.5-1.

GENETIC SCREENING

Genetic screening refers to detection of mutant genes in an individual (detecting human genetic variations). Modern genetics is making this much earlier than it was in the past.

Field of genetic screening

There are three situations where genetic screening is of particular relevance:

1. Prenatal diagnosis. Prenatal diagnosis aims at identifying the health problems of unborn babies. Parents can be provided counseling and option for abortion as per situation.

Techniques of prenatal diagnosis include:

- Chorionic villus sampling,
- Amniocentesis and
- Pre-implantation diagnosis.

2. Carrier diagnosis. This is the identification of people who carry a particular genetic disease, usually with no visible symptom or harm to themselves. Examples include sickle cell anaemia, cystic fibrosis and phenylketonuria.

3. Predictive diagnosis. This is the prediction of a future disease from which one is likely to suffer. The classic example of this 'genetic time bomb' is Huntington's chorea where the onset of disease occurs in middle age.

GENETICS AND CANCER

Cancer is a disease characterized by uncontrolled cell growth.

Points favouring genetic basis for cancer

- *Hereditary predisposition* is noted in some cancers like colon cancer and retinoblastoma.
- *Chromosomal abnormalities* are noted in many forms of cancers like Burkitt's lymphoma and acute myeloid leukaemia.

Table 1.5-1 Some common genetic diseases

Genetic disease/ disorder	Chromosome affected	Type of mutation	Expression of gene	Main symptoms	Defect	Frequency at the birth
Gene mutations						
Sickle cell anaemia	11	Substitution	Codominant (sometimes described as recessive) autosomal	Anaemia and interference with circulation	Abnormal haemoglobin molecule	1 in 1600 among black people
Cystic fibrosis	7	In 70% of cases is a deletion of three bases	Recessive autosomal	Unusually thick mucus clogs lungs, liver and pancreas	Failure of chloride ion transport mechanism in cell surface membranes of epithelial cells	1 in 1800 among white people
PKU (Phenylketonuria)	12	Substitution	Recessive autosomal	Brain fails to develop normally	Enzyme Phenylalanine hydroxylase defective	1 in 18,000
Huntington's chorea (disease)	4	A newly discovered type of mutation—the normal gene has 10–34 repeats of CAG at one end, the HC gene has 42–100 repeats of CAG	Dominant autosomal	Gradual deterioration of brain tissue starting on average in middle age	Brain cell metabolism is inhibited	1 in 10,000 to 1 in 20,000 worldwide
Haemophilia	X	Substitution	Recessive sex-linked	Blood does not clot	Factor VIII or IX protein defective	1 in 7000
Chromosome mutations						
Down's syndrome	21	Extra chromosome (trisomy 21)		Reduced intelligence, Characteristic facial features		1 in 750
Klinefelter's syndrome (XXY)	Sex	Extra X chromosome in male (trisomy)		Feminised male		1 in 500
Turner's syndrome (XO)	Sex	Missing X chromosome in female (monosomy)		Sterile female		1 in 2500

Autosomal – affecting non-sex chromosome (autosome).
Monosomy – one chromosome missing (2n – 1).
Trisomy – one extra chromosome (2n + 1).
Monosomy and trisomy are examples of **aneuploidy,** where the total number of chromosomes is not an exact multiple of the haploid number.

- *Defective DNA repair mechanisms* have been associated with occurrence of cancers.
- *Genetic damages* (mutagenesis) by the action of various agents like ionizing radiations, UV rays are associated with occurrence of cancer.

Genes and molecular factor involved in pathogenesis of cancer

1. Oncogenes. Over 100 oncogenes (cancer causing genes) have been described. These genes are derived by somatic mutations from the closely related *proto-oncogenes* (normal genes that encode proteins having a role in cell's normal activities).

Genetic changes converting proto-oncogenes into oncogenes are:

- *Missense mutation,* i.e. change in the amino acid sequence of proto-oncogene protein converts it into oncogene.
- *Gene amplification,* e.g. Myc genes have been amplified in human leukaemia, breast, stomach, lung and colon cancer.
- *Chromosomal translocations,* e.g. in Burkitt's lymphoma a region of chromosome 8 is translocated to either chromosome 2, 14 or 22. The breakpoint in chromosome 8 causes the overexpression of c-myc gene.
- *Retroviral integration,* e.g. in avian lymphomas the integration of the avian leucosis virus can enhance the transcription of the c-myc gene.

2. Tumour suppressor gene (antioncogenes). Over 10 tumour suppressor genes have been described which produce proteins that suppress tumour. When a tumour suppressor gene becomes inactivated by mutation, it becomes more likely that cancer will occur. It is important to note that for causing cancer, both alleles of tumour suppressor genes should be mutated. The most studied of these genes are RB gene (retinoblastoma gene) and P 53 gene.

3. Mutator genes. Normal cells have caretaker genes that regulate DNA repair, i.e. they prevent faulty DNA transcription and regulation. The mutated version or mutator gene is characterized by loss of normal surveillance function that renders DNA susceptible to accumulation of mutation and therefore progression of cancer.

4. Telomeres in cancer. Telomeres care the terminal tips of chromosome which progressively shorten due to repetitive cell division. Telomerase is the enzyme required for continued recognition of telomere in successive cell divisions. Cancer cell expresses telomerase with consequent telomerase lengthening and this helps the transformed cell to maintain their cancerous state (Fig. 1.5-11).

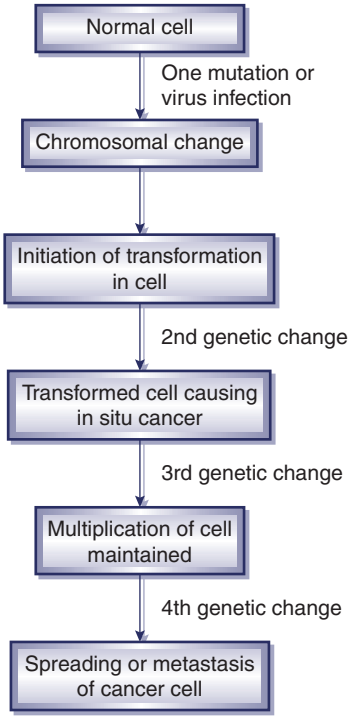


Fig. 1.5-11 Multistep phenomena of series of genetic changes leading to development of cancer.

Conclusion. To conclude, cancer is a multistep phenomenon where series of genetic changes (Fig. 1.5-10) can be produced by:

- Ionizing radiations,
- Chemical carcinogens,
- Radioactive material,
- Spontaneous mutations and
- Inherited mutations.

GENE THERAPY

Gene therapy is the name given to methods that aim to cure an inherited disease by providing the patient with a correct copy of the defective gene. Gene therapy has now been extended to include the attempts to cure any disease by introduction of a cloned gene into the patient.

Basic principles

Basic principles involved in correcting a genetic defect are:

Gene replacement, i.e. replacement of a mutant gene with a normal gene.

Gene correction, i.e. correction of the mutated area (specific bases) of DNA leaving the rest of DNA unchanged.

Gene augmentation, i.e. insertion of a foreign DNA into the genome of a cell to rectify the genetic defect.

Basic approaches to gene therapy are:

Germ line therapy can be carried out by microinjection of DNA into the isolated egg cell that is reimplanted into the mother. If successful the gene is present and expressed in all cells of the resulting individual. Thus, theoretically the germ line therapy can be used to treat any inherited disease. At the moment such treatment is regarded as unethical in humans because the gene would be passed on to future generations.

Somatic cell therapy. In humans, at the moment the focus is on somatic cell therapy. This involves changing some, though not all, of the somatic cells which are non-sex cells of the body. Changes in these cells cannot be inherited. The patients treated will therefore be cured but they will still be able to pass faulty gene on their offspring. Steps involved in this therapy are:

- Isolation of the cells with the gene defect from a patient,
- Growing the isolated cells in culture,
- Transfecting the isolated cell with a remedial gene construct,
- Selecting, growing and testing the transfecting cells and
- Either transplanting or transfusing the transfecting cells back into the patients.

Examples of successful trials of gene therapy

Somatic cell therapy in cystic fibrosis of the lung. The gene (cDNA) cloned in adenovirus vectors or contained in liposomes, when introduced into the respiratory tract via an inhaler, is taken up by the epithelial cells. These epithelial cells then make normal protein called CFTR (cystic fibrosis transmembrane conductance regulator). Probably 10% of the cells need to be corrected to eliminate the problem.

Somatic cell therapy in severe combined immunodeficiency disease (SCID). In this disease, the mutated gene is unable to make the enzyme adenosine deaminase (ADA). The ADA is needed by the white blood cells (lymphocytes) responsible for immunity to infection. Without ADA the child develops SCID and dies of infection in the early childhood.

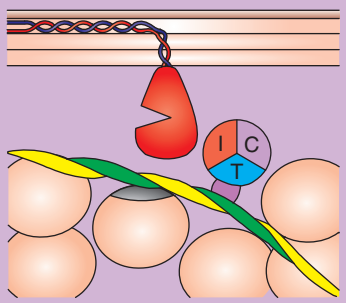
Two children, aged four and nine suffering from SCID were selected for gene therapy in USA in 1990. Lymphocytes were isolated from the children, a normal gene introduced by means of a retrovirus vector, and the cells replaced. These children had shown significant improvement after 1 year of therapy repeated every 1–2 month.

Gene therapy in cancer may have great potentiality

- *Inactivation of oncogenic gene* by introduction of a gene or an antisense RNA of an oncogenic gene.
- Introduction of an active version of tumour suppressor gene.
- Introduction of a gene that will selectively kill the cancer cells.
- Gene therapy to improve the natural killing of cancer cells by the patient's immune system.

SECTION 2

Nerve Muscle Physiology



The nerve and muscle cells are excitable, that is, capable of generation of electrical impulses at their membranes. For this very reason, the physiological aspects of these excitable tissues are discussed together in this section.

The electrical impulses generated in the excitable tissues, in most instances, can be used to transmit signals along the membranes. A neuron is the basic unit of nervous tissue. It is specialized for the function of reception, integration and transmission of information in the body. Muscles, like neurons, are excitable tissues but are characterized by the fact that a mechanical contraction follows an action potential.

To understand the physiological aspects, it is imperative to have knowledge about the functional anatomy and physiological properties of the nerve, the muscle and the neuromuscular junction.

“This page intentionally left blank"

Chapter 2.1

The Nerve

FUNCTIONAL ANATOMY
- Neuron
- Neuroglia
- Peripheral nerve

BIOLOGICAL ACTIVITIES
- Protein synthesis
- Metabolism and heat production in the nerve fibres

ELECTRICAL PROPERTIES OF NERVE FIBRE
- Excitability
- Conductivity
- Recording of membrane potential and action potentials
- Compound action potential

NERVE FIBRE TYPES
- Classification of nerve fibres

DEGENERATION AND REGENERATION OF NEURONS
- Causes and grading of nerve injury
- Stage of degeneration
- Stage of regeneration

FACTORS PROMOTING NEURONAL GROWTH
- Neurotrophins
- Other growth promoting factors

FUNCTIONAL ANATOMY

NEURON

Neuron, or the nerve cell, is the structural and functional unit of the nervous system. The nervous system of human is made up of innumerable neurons. The total number of estimated neurons in the human brain is more than 10^{12}. The neurons are linked together in a highly intricate manner. It is through these connections that the body is made aware of changes in the environment, or of those inside the body itself; and appropriate responses to such changes are produced.

STRUCTURE

Neurons vary considerably in size, shape and other features. However, most of them have some major features in common. The basic structure of a neuron is best studied in a spinal motor neuron. A neuron primarily consists of the cell body and processes called neurites, which are of two kinds, the dendrites and the axon (Fig. 2.1-1).

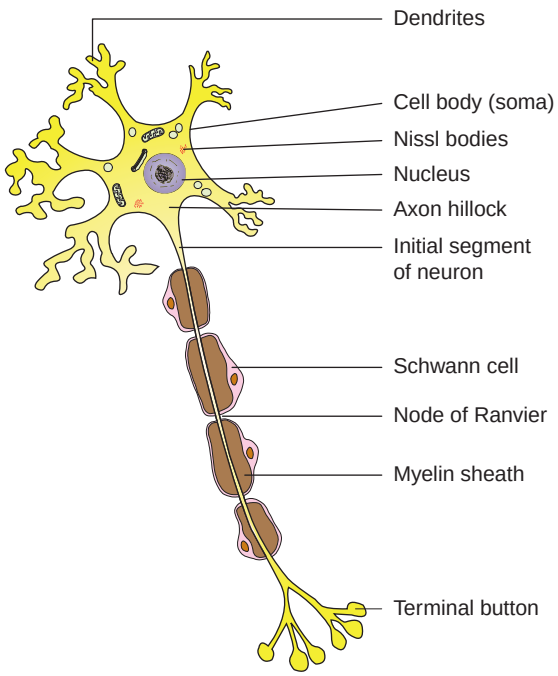


Fig. 2.1-1 Structure of a typical neuron.

Cell body

The cell body of a neuron is also called the *soma* or *perikaryon* and may be round, stellate, pyramidal or fusiform in shape. Like any other cell it consists of a mass of cytoplasm with all its principal constituents surrounded by a cell membrane. The cell body contains a large nucleus with one or two nucleoli but there is no centrosome.

Note. The absence of centrosome indicates that the neuron has lost ability for division. Thus, neurons once destroyed are replaced by neuroglia only.

In addition to the general features of a typical cell (page 9), the cytoplasm of a neuron has following distinctive characteristics (Fig. 2.1-1):

Nissl granules/bodies. These are basophilic granules, when seen under electron microscope, these bodies are seem to be composed of rough surfaced endoplasmic reticulum. The presence of abundant granular endoplasmic reticulum is an indication of the high level of protein synthesis in neurons. The proteins are needed for maintenance and repair, and for production of neurotransmitters and enzymes. The Nissl bodies are present in the dendrites as well but are usually absent from the axon hillock and the axon. These bodies disintegrate into fine dust and which finally disappears (*chromatolysis*) on fatigue, due to the effect of certain poisons and on sectioning of the axon.

Neurofibrillae. These consist of microfilaments and microtubules. In certain degenerative disease like Alzheimer's disease, the neurofilament protein gets altered, resulting in the formation of *neurofibrillary* tangles.

Pigment granules are seen in some neurons. For example, neuromelanin is present in the neurons of substantia nigra. Aging neurons contain a pigment *lipofuscin*.

Dendrites

The dendrites are multiple small branched processes which contain Nissl bodies and neurofibres. Dendrites are the *receptive processes* of the neuron receiving signals from other neurons via their synapses with axon terminals.

Axon

The axon is the single longer process of the nerve cell. It varies in length from a few microns to one metre. It arises from the conical extension of the cell body called *axon hillock,* which is devoid of the Nissl bodies. The part of the axon between the axon hillock and the beginning of myelin sheath is called the *initial segment.* In the axon, the cell membrane continues as *axolemma* and the cytoplasm as *axoplasm*. The axon terminates by dividing into a number of branches, each ending in a number of *synaptic knobs* also known as *terminal buttons* or axon telodendria. Synaptic knobs contain microvesicles in which chemical neurotransmitters are stored. *Myelin sheath* is present around the axon in the so-called myelinated nerve fibres (Fig. 2.1-2). Myelin sheath which consists of protein–lipid complex is produced by glial cells called Schwann cells which encircle the axon forming around it a thin sleeve. Each *Schwann cell* provides the myelin sheath for a short segment of the axon. At the junction of any two such segments, there is a short gap, i.e. periodic 1 µm constrictions at about 1 mm distance. These gaps are the *nodes of Ranvier*. There are some axons which are devoid of myelin sheath.

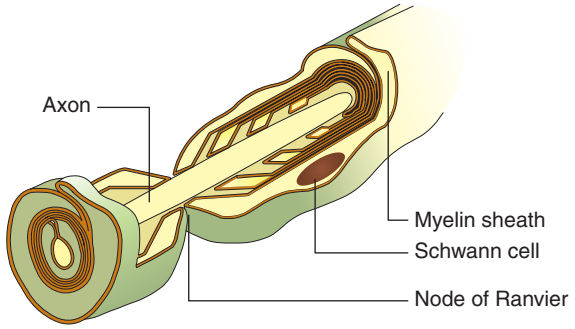


Fig. 2.1-2 Structure of a myelinated neuron. Note, Schwann cell encircles the axon to form myelin sheath.

Myelination of axons increases the speed of conduction, but greatly increases their diameter.

Axons perform the specialized function of conducting impulses away from the cell body.

TYPES OF NEURONS

Neurons have been variously classified as:

I. Depending upon the number of poles

Depending upon the number of poles from which processes arise, neurons are divided into (Fig. 2.1-3):

1. ***Unipolar neurons*** have a single pole, from which both the processes—axon and dendrite arise (Fig. 2.1-3A). True unipolar cells are present only in embryonic stage in human being. However, the primary sensory neurons (neurons conveying impulses from a sensory receptor to spinal cord) are pseudounipolar (Fig. 2.1-3B).
2. ***Bipolar neurons*** have two poles, one for axon and other for dendrite (Fig. 2.1-3C). Bipolar neurons are found in the vestibular and cochlear ganglia, in the nasal olfactory epithelium and as bipolar cells in the retina.
3. ***Multipolar neurons*** have many poles. One of the poles gives rise to axon and all others to dendrites (Fig. 2.1-3D). Most vertebrate neurons, especially in the central nervous system (CNS) are multipolar. The dendrites branch profusely to form the dendritic tree.

II. Depending upon the function

Depending upon the functions the neurons are of two types—motor and sensory.

1. ***Motor neurons,*** also known as efferent nerve cells, carry the motor impulses from the CNS to the peripheral effector organs like muscles, glands and blood

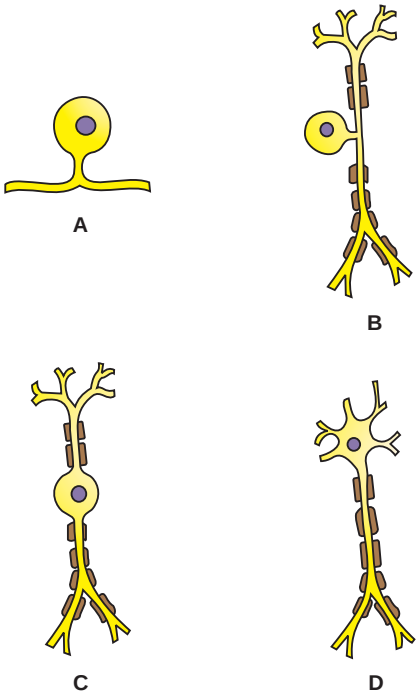


Fig. 2.1.3 Different types of neurons: A, unipolar; B, pseudo-unipolar; C, bipolar and D, multipolar.

vessels. These neurons have very long axon and short dendrites.

2. ***Sensory neurons,*** also known as afferent nerve cells, carry the sensory impulses from the periphery to the CNS. These neurons have short axon and long dendrites.

ZONES OF THE NEURON

From the functional point each neuron is divided into four zones (Fig. 2.1-4):

1. ***Receptor zone*** (dendritic zone) is the region where local potential changes are generated by integration of the synaptic connections.
2. ***Site of origin of conducted impulse*** is the site, where propagated action potentials are generated. In case of spinal motor neuron, *initial segment* and in cutaneous sensory neurons *first node* of Ranvier is the site of origin of conducted impulses.
3. ***Zone of all or none transmission*** in the neuron is the axon.
4. ***Zone of secretion of transmitter*** (nerve endings). The propagated impulses (action potential) to nerve endings cause the release of neurotransmitter.

NEUROGLIA

Neuroglia or the glial cells are the supporting cells present within the brain and spinal cord. They are numerous, about

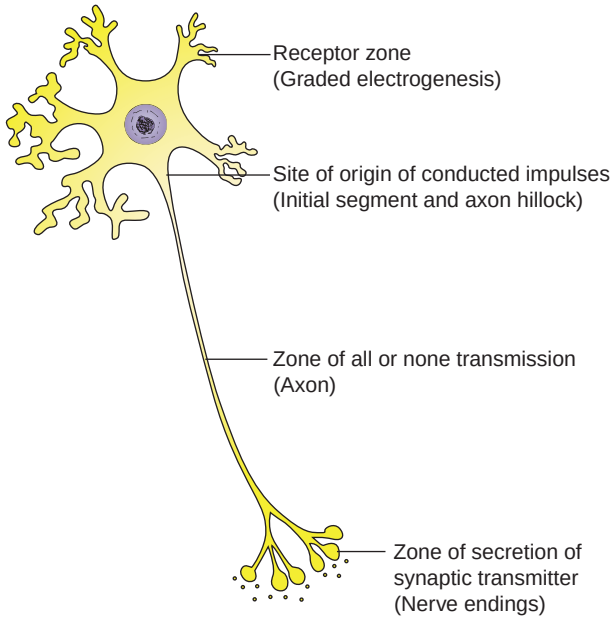


Fig. 2.1-4 Functional zones of the neuron.

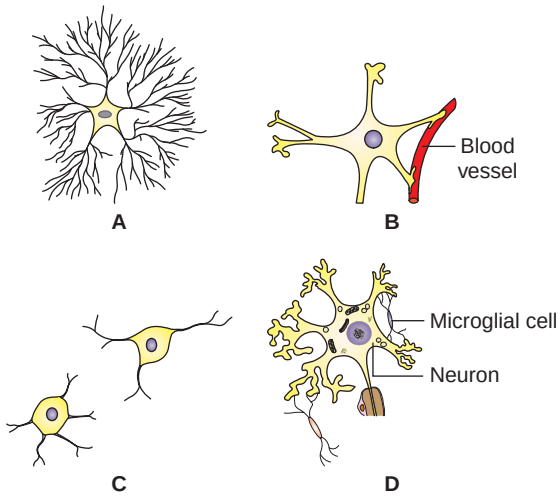


Fig. 2.1-5 Different types of glial cells: A, fibrous astrocyte; B, protoplasmic astrocyte; C, oligodendrocyte and D, microglial cells.

10 times more than the neurons. Glial cells may be divided into two major categories (Fig. 2.1-5):

1. Macroglia

Macroglia or large glial cells are ectodermal in origin. These are of two types:

- *Astrocytes* which may be subdivided into fibrous and protoplasmic astrocytes and
- *Oligodendrocytes*.

2. Microglia

Microglia or the small glial cells are mesodermal in origin. These are the smallest neuroglial cells having flattened cell

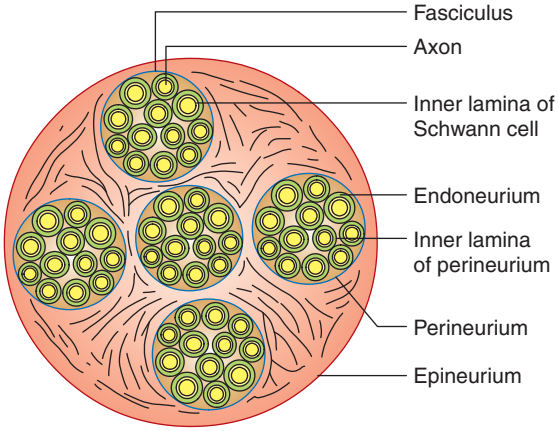


Fig. 2.1-6 Cross-section of a peripheral nerve.

body and short processes. They are more numerous in grey matter than in white matter. These act as phagocytes and become active after damage to nervous tissue by trauma or disease.

PERIPHERAL NERVE

A compact bundle of axons located outside the CNS is called a nerve. In a nerve the axons are arranged in different bundles called fasciculi (Fig. 2.1-6).

- Each axon or the nerve fibre is covered by *endoneurium* which is bounded internally by basal lamina around the Schwann cells and externally by the relatively impermeable inner basal lamina of the perineurium.
- Each fasciculus is covered by *perineurium.* The cells of the perineurium are tightly adherent and act as a barrier to the passage of particulate traces, dye molecules or toxins into the endoneurium.
- The whole nerve is covered by *epineurium* which is a tubular sheath formed by an areolar membrane. It limits the extent to which the nerve can be stretched by body movements or external pressure, thereby protecting the fragile axons inside the nerve.

BIOLOGICAL ACTIVITIES

PROTEIN SYNTHESIS

The cell bodies (soma) of all the neurons contain cellular apparatus required for the protein synthesis, i.e. the ribosomes and the Golgi apparatus. Since the axons do not have these organelles, all the proteins including the neurotransmitters are synthesized in the cell body and then transported along the axon to the synaptic knobs by the process of axoplasmic flow.

AXOPLASMIC TRANSPORT

Axoplasm, the cytoplasm of the neurons is in constant motion. The axoplasmic transport is vital to nerve cell functions, since movements of various materials occur through it. The axoplasmic transport is of two types: rapid and slow.

1. Rapid transport

Some materials travel 100–400 mm a day along the axoplasm and constitute the rapid transport. Microtubules play an important role in this form of transport. Rapid transport is bidirectional, i.e. both away from (anterograde) and towards the cell body (retrograde).

Note. Retrograde axoplasmic flow may also carry tetanus toxin and neurotropic viruses (e.g. polio, herpes simplex and rabies) along the axon into the neuronal cell bodies in the CNS. It has also been employed by the neuroanatomists for charting out neural pathways.

2. Slow transport

The materials travelling slowly (0.1–2 mm in a day) in the axoplasm constitute the slow transport. Slow transport is only unidirectional, away from the cell body (anterograde). It is responsible for flow of axoplasm containing protein subunits of neurofilaments, tubulins of the microtubules and soluble enzymes.

METABOLISM AND HEAT PRODUCTION IN THE NERVE FIBRES

Like other cells of the body, the metabolic activities occur in the nerve fibres as well; but the metabolism in nerve fibres occurs at a very low level. About 70% of the total energy required is used to maintain polarization of the membrane by the action of Na^+–K^+ ATPase pump. The energy is supplied mainly by combustion of sugars and phospholipids.

During nerve activity, the ATP and creatine phosphate breakdown, i.e. undergo hydrolysis and supply energy for the propagation of the nerve impulse.

Heat production in the nerve fibres

In a nerve fibre, heat is produced in three phases:

1. ***Resting heat*** is the amount of heat produced during the inactive stage.
2. ***Initial heat*** is the amount of heat produced during action potential (stage of activity). It is about 10% of the total heat produced. It results from anaerobic metabolic activity due to breakdown of ATP and creatine phosphate.

3. ***Delayed or recovery heat*** is produced during the recovery phase which follows the phase of activity. The energy is produced by aerobic metabolic activities and is about 30 times the initial heat. The energy produced during the recovery stage is used for resynthesis of ATP and creatine phosphate and as such for restoring the normal excitability of the nerve fibre.

ELECTRICAL PROPERTIES OF NERVE FIBRE

The main electrical properties of the nerve fibres are:

- *Excitability,* i.e. the capability of generating electrical impulses (action potential) and
- *Conductivity,* i.e. the ability of propagating the electrical impulses generated along the entire length of nerve fibres.

EXCITABILITY

Excitability is that property of the nerve fibre by virtue of which it responds by generating a nerve signal (electrical impulses or the so-called action potentials) when it is stimulated by a suitable stimulus which may be mechanical, thermal, chemical or electrical.

RESTING MEMBRANE POTENTIAL

The study of electrical activity of a tissue has been made possible due to advances in the method of the recording electrical potentials, especially the development of microelectrodes and 'cathode ray oscilloscope (CRO)'.

As shown in Fig. 2.1-7, when two electrodes are placed on the surface of a nerve fibre and connected to a CRO, no potential difference is observed. However, if one of the microelectrodes is inserted inside the nerve fibre (Fig. 2.1-7), a steady potential difference of –70 mV (inside negative) is observed on the CRO. This is resting membrane potential (RMP) and indicates the resting state of cell also called state of polarization. For details, see page 25.

ACTION POTENTIAL

The action potential may be defined as the brief sequence of changes which occur in the resting membrane potential when stimulated by a threshold stimulus. When the stimulus is subminimal or subthreshold, it does not produce action potential, but do produce some changes in the RMP. There is slight depolarization for about 7 mV which cannot be propagated, since propagation occurs only if the depolarization reaches a firing level of 15 mV (–55 mV). Once the firing level is reached, there occurs action potential, i.e. there occurs abrupt depolarization with propagation (action potential).

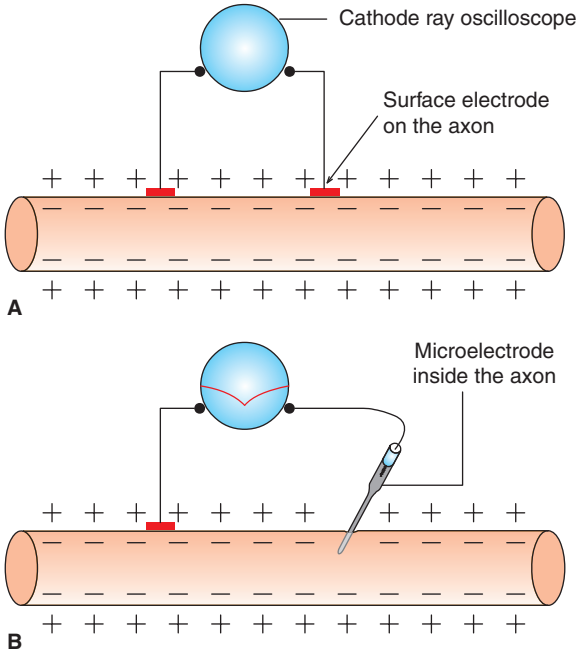


Fig. 2.1-7 Recording of resting membrane potential: A, both electrodes are on the surface of axon, no potential difference is recorded and B, one electrode on the surface and other inserted inside the axon, potential difference (–70 mV) is recorded.

PHASES OF ACTION POTENTIAL

The action potential basically occurs in two phases: depolarization and repolarization. When the nerve is stimulated, the polarized state (–70 mV) is altered, i.e. the RMP is abolished and the interior of the nerve becomes positive (+35 mV) as compared to the exterior. This is called depolarization phase. Within no time there occurs reverse to the nearly original potential and this second phase of action potential is called repolarization phase.

Action potential curve obtained when resting membrane potential is being recorded on a CRO and the nerve fibre is stimulated at a short distance away from the recording electrode (Fig. 2.1-8A) has following components (Fig. 2.1-8B):

1. *Resting membrane potential* is recorded as a straight baseline at –70 mV.
2. *Stimulus artefact* is recorded as a mild deflection of the baseline as soon as the stimulus is applied. The stimulus artefact occurs due to leakage of current from the stimulating electrode to the recording electrode.
3. *Latent period* is recorded as a short isoelectric period (0.5–1 ms) following the stimulus artefact. It represents

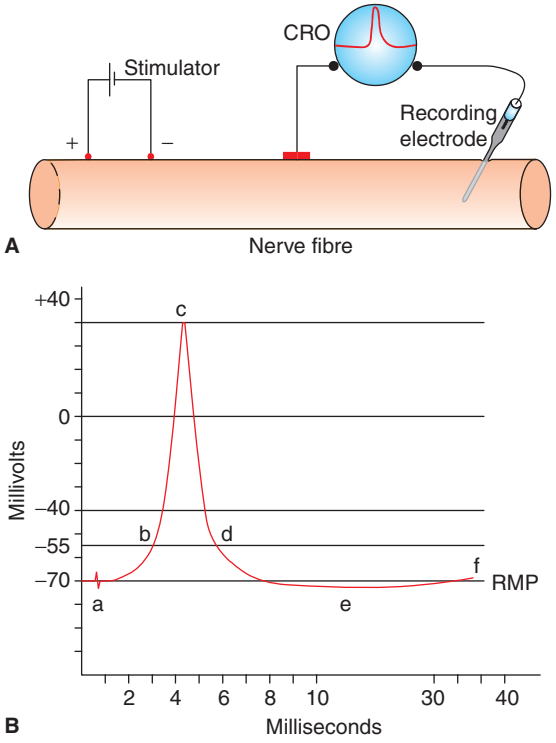


Fig. 2.1-8 Recording of action potential of a large mammalian myelinated nerve fibre: A, arrangement for recording action potential and B, various phases (components) of action potential. (a = Stimulus artefact, b = firing level, b–c = depolarization, c–d = repolarization, d–e = after depolarization; e–f = after hyperpolarization.)

the interval between the application of stimulus and the onset of action potential. It depends upon:

(i) The distance between the site of stimulation, and the point of recording, and

(ii) The velocity of conduction of action potential along the particular axon.

4. *Firing level.* After the latent period, phase of depolarization starts. To begin with, depolarization proceeds relatively slow up to a level called the *firing level* (–55 mV), at which depolarization occurs very rapidly.
5. *Overshoot.* From the firing level, the curve reaches the zero potential rapidly and then overshoots the zero line up to +35 mV.
6. *Spike potential.* After reaching the peak (+35 mV), the phase of depolarization is completed and the phase of repolarization starts, and the potential descends quickly near the firing level. The phase of rapid rise of potential in depolarization and a rapid fall in repolarization phase, combinedly constitute the so-called *spike potential.* Its duration is approximately 1 ms in an axon.
7. *After depolarization* is the slow repolarization phase which follows a rapid fall in spike potential and extends up to attainment of the RMP level. It is called phase of *negative after potential* and lasts for about 4 ms.

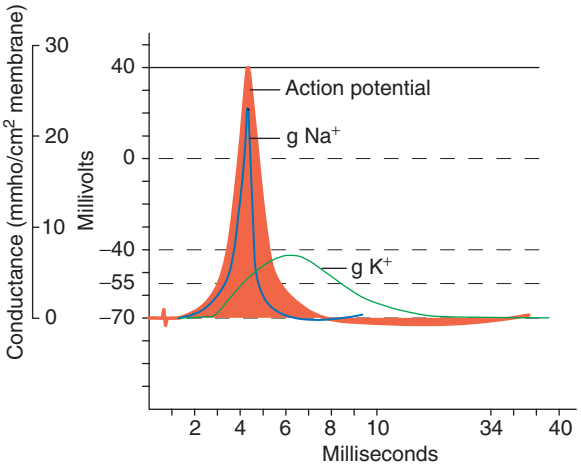


Fig. 2.1-9 Changes in sodium (Na^+) and potassium (K^+) conductance during action potential.

8. *After hyperpolarization.* After reaching the resting level (–70 mV) the potential further falls and becomes more negative (–72 mV). This phase is called *after hyperpolarization* or phase of *positive after potential.* It lasts for a prolonged period (35–40 ms). Finally, the RMP is restored.

IONIC BASIS OF ACTION POTENTIAL

According to the Hodgkin–Huxley theory, the sequence of events is:

1. Polarization phase. Resting membrane potential (–70 mV) is due to distribution of more cations outside the cell membrane and more anions inside the cell membrane. At this point though Na^+ are more in the extracellular fluid (ECF) but they cannot enter the cell due to the impermeability of the membrane. For details see page 25.

2. Depolarization phase. When threshold stimulus is applied to the cell membrane, at the point of stimulation (Fig. 2.1-9) the permeability of the membrane for Na^+ ions increases. At first the rise of permeability for Na^+ is slow till it reaches the *firing level.* When the membrane depolarizes, the Na^+ channels start opening up. The opening of the Na^+ channels (m gates) depolarizes the membrane further, leading to the opening of greater number of Na^+ channels. As the concentration gradient and electrical gradient of this ion are directed inward, there occurs a rapid influx of Na^+ ions into the cell. This rapid entry of Na^+ is sufficient to overwhelm the repolarizing forces.

Factors which limit further depolarization are:

- Inactivation of Na^+ channels due to activation of h-gates.
- During overshoot the direction of electrical gradient is reversed.
- Opening of voltage-gated K^+ channels. K^+ efflux starts along concentration gradient.

3. Repolarization phase. Repolarization occurs due to decrease in further Na^+ influx and K^+ efflux through the voltage-gated K^+ channels which open later than Na^+ channels but remain activated for prolonged period.

4. After depolarization. After the rapid initial repolarization (spike potential), the further repolarization occurs slowly. This is due to the fact that the rate of K^+ efflux slows down as the electrical gradient responsible for initial rapid diffusion declines. This last phase of slow repolarization due to slow efflux of K^+ is called *after depolarization.*

5. After hyperpolarization. The slow efflux of K^+ continues even after the resting membrane potential is reached, resulting in a prolonged phase of *hyperpolarization* during which the membrane potential falls up to –72 mV. However, little after, the voltage-gated K^+ channels also shut down. The final ionic distribution is brought to the resting state by the action of Na^+–K^+ pump and the leak channels (K^+, Cl^-).

Role of calcium ions

In addition to the Na^+ and K^+ ions, the Ca^{2+} ions also play some role in the development of action potential. The concentration of Ca^{2+} in the intracellular fluid (ICF) is very low as compared to ECF. Therefore, when Na^+ channels are open, some Ca^{2+} ions also move inside the cell through these opened up Na^+ channels.

CHARACTERISTICS OF NERVE EXCITABILITY VIS-A-VIS CHARACTERISTICS OF THE STIMULUS

1. STRENGTH–DURATION CURVE

The relationship between the strength and duration of a stimulus has been studied by varying the duration of a stimulus and finding out the threshold strength for each duration. The record of results plotted on a semilog graph paper gives the *strength–duration* curve (Fig. 2.1-10). Following inferences can be drawn from the strength–duration curve:

- *Rheobase (R)* refers to the minimum intensity of stimulus which if applied for adequate time (utilization time) produces a response.
- *Chronaxie (C)* refers to the minimum duration for which the stimulus of double the rheobase intensity must be applied to produce a response. Within limits, chronaxie of a given excitable tissue is constant. In other words, the chronaxie is an index of the excitability of a tissue and can be used to compare the excitability of various tissues. For example, a nerve fibre has far shorter chronaxie value than a muscle fibre indicating greater excitability of the former.
- When a stimulus of weaker intensity than the rheobase is applied, it will not produce a response, no matter how long the stimulus is applied.

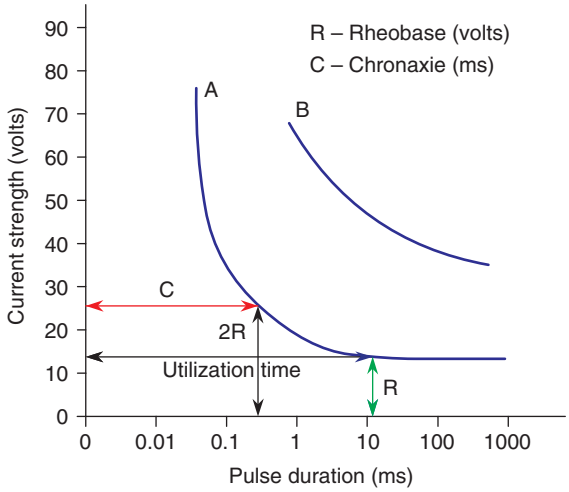


Fig. 2.1-10 Strength–duration curves: A, for nerve and B, for muscle.

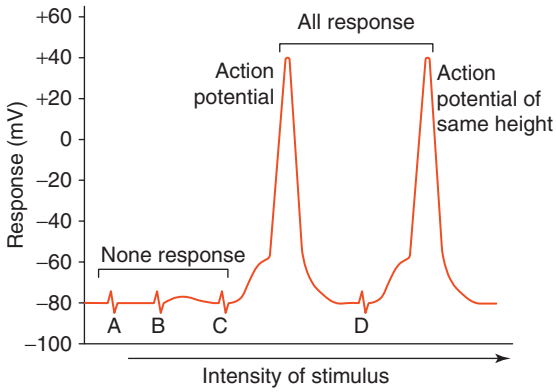


Fig. 2.1-11 All or none response in a single nerve fibre: A and B, represent subthreshold; C, threshold and D, suprathreshold stimuli.

- Stimulus of extremely short duration will not produce any response, no matter how intense that may be.

2. ALL OR NONE RESPONSE

A single nerve fibre always obeys 'all or none law', that is (Fig. 2.1-11):

- When a stimulus of subthreshold intensity is applied to the axon, then no action potential is produced (*none response*).
- A response in the form of spike of action potential is observed when the stimulus is of threshold intensity
- There occurs no increase in the magnitude of action potential when the strength of stimulus is more than the threshold level (all response).

This all or none relationship observed between strength of stimulus and response achieved is known as *All or None* law.

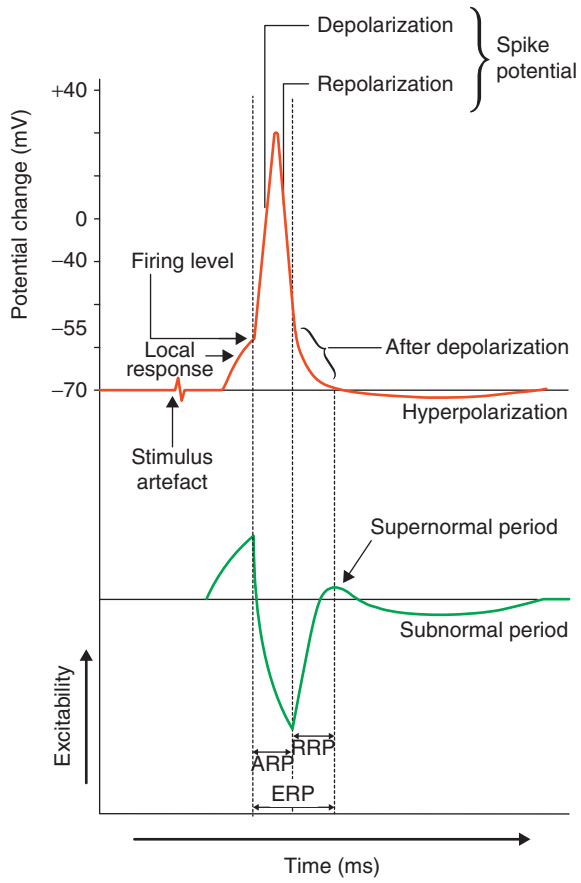


Fig. 2.1-12 Membrane excitability during different phases of action potential: ARP (absolute refractory period), RRP (relative refractory period) and ERP (effective refractory period).

MEMBRANE EXCITABILITY DURING ACTION POTENTIAL

Depending upon the response elicited to the stimulus, the period of action potential can be divided into: refractory period, supernormal period and subnormal period (Fig. 2.1-12).

(i) Refractory period

Refractory period refers to the period following action potential (produced by a threshold stimulus) during which a nerve fibre either does not respond or responds subnormally to a stimulus of threshold intensity or greater than threshold intensity. It is of two types:

(a) Absolute refractory period. It is a short period following action potential during which second stimulus, no matter how strong it may be, cannot evoke any response (another action potential). In other words, during absolute refractory period the nerve fibre completely loses its excitability. The absolute refractory period corresponds to the period of action potential from the firing level until repolarization is almost one third complete (spike potential).

Ionic basis of absolute refractory period. During up stroke of action potential (depolarization), the *m gates* of sodium channels in the membrane of nerve are opened rapidly. During down stroke (early repolarization), the channels are closed by closure of inactivation *(h) gates* of the sodium channels and slow potassium channels are not yet opened. These sodium channel gates do not open unless potential comes back to the resting level. Therefore, during this period (absolute refractory period) the nerve fibre is not stimulated at all.

(b) Relative refractory period. It is a short period during which the nerve fibre shows response, if the strength of stimulus is more than normal. It extends from the end of absolute refractory period to the start of *after depolarization* of the action potential.

Ionic basis of relative refractory period. During this stage the Na^+ channels are coming out of inactivated stage and voltage gated potassium channels are still opened. The stronger stimulus (suprathreshold) at this stage is able to open more Na^+ channels through 'm' gates and thus excite a response. The action potential elicited during this period, however, has a lower upstroke velocity and lower overshoot potential than the normal action potential.

Effective refractory period (ERP) includes ARP and early part of RRP. At the end of ERP, the nerve membrane is able to produce and conduct the impulse.

(ii) Supernormal period

During supernormal period, the membrane is hyperexcitable, i.e. the threshold of stimulus is decreased. This period corresponds with the after depolarization phase of the action potential.

Ionic basis of supernormal period. During 'after depolarization' phase the Na^+ channels have come out of inactivated state but the K^+ channels (voltage gated) are mostly closed and the membrane potential is nearer to the firing level, so a stimulus of low intensity will be able to excite action potential. In other words, the threshold level of stimulus is decreased during this stage.

(iii) Subnormal period

During this period the membrane excitability is low, i.e. the threshold of stimulus is increased. This period corresponds with the *after hyperpolarization* stage of the action potential.

3. ACCOMMODATION

We have studied that when a stimulus of sufficient strength (threshold level) is applied quickly, then the action potential is produced. However, when the stimulus strength is increased slowly to the firing level (during constant application),

no *action potential is produced.* This phenomenon of adaptation to the stimuli is called *accommodation.* Therefore, a *square pulse stimulus* (which rises sharply to its peak level) effectively triggers an action potential, whereas a *sawtooth pulse* (which rises to its peak slowly) often fails to trigger an action potential.

Ionic basis of accommodation. If depolarization occurs rapidly, the opening of the Na^+ channels overwhelm the repolarizing forces and the typical action potential is produced. However, when the induced depolarization is produced slowly, more and more Na^+ channels open up, only to get inactivated after 1 ms, while the K^+ channels remain open which tend to restore the membrane potential. Thus the repolarizing forces overwhelm the depolarizing forces and so the action potential is not produced.

4. INFATIGUABILITY

A nerve fibre cannot be fatigued, even if it is stimulated for a long time. This property of infatiguability is due to absolute refractory period (see page 52).

ELECTROTONIC POTENTIAL AND LOCAL RESPONSE

When a nerve fibre is stimulated by a subminimal or subthreshold stimuli, the action potential is not produced but there do occur some changes in the resting membrane potential. These local non-propagated changes are called *electrotonic potentials or acute subthreshold potentials.*

Types of electrotonic potentials

Depending upon the nature (negative or positive) of subthreshold level current used to stimulate (Fig. 2.1-13), the electrotonic potentials are of two types:

1. Catelectrotonic potential. Catelectrotonic potentials are localized depolarizing potential changes in the membrane potential produced when the stimulus of subthreshold strength is applied with cathode. This potential change rises sharply and then decays exponentially with time.

2. Anelectrotonic potential. Anelectrotonic potentials are the localized hyperpolarizing potential changes in the membrane potential when anodal subthreshold current is applied.

Graded potentials

As shown in Fig. 2.1-13, the electrotonic potentials produced by varying intensities of stimuli are proportionate to the magnitude of stimulation. Therefore, these are also known as *graded potentials.* The graded response is achieved both at cathode and anode up to 7 mV of depolarization or hyperpolarization.

Differences between graded potentials and action potential are summarized in Table 2.1-1.

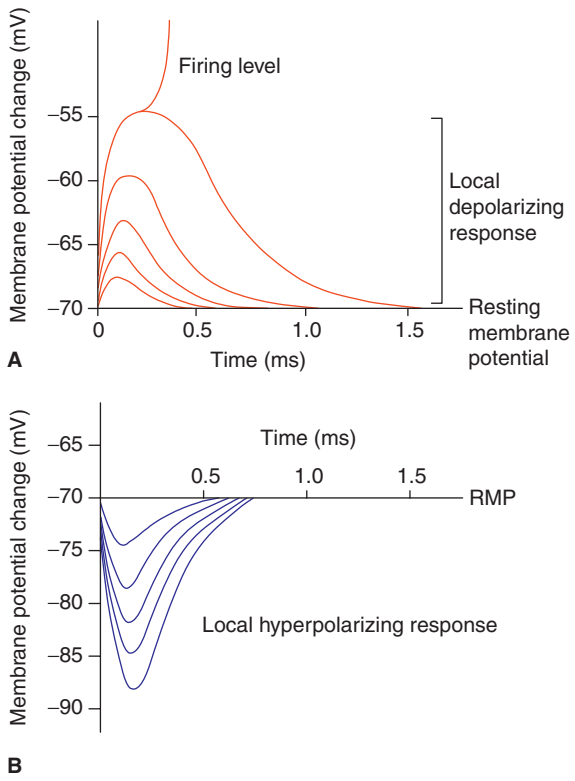


Fig. 2.1-13 Electrotonic potential: A, catelectrotonic potential and B, anelectrotonic potential.

Table 2.1-1 Differences between graded potential and action potential

Graded potential	Action potential
1. The amplitude of graded potential is proportionate to the intensity of the stimulus and can get summated.	The amplitude of action potential remains constant with increasing intensity of stimulus, therefore it cannot be summated.
2. Graded potentials can be either depolarizing or hyperpolarizing.	Action potential is always depolarizing.
3. Graded potential can be generated either spontaneously or in response to either physical or chemical stimuli.	Action potential is generated only in response to membrane depolarization.
4. Graded potentials cannot conduct impulse.	Action potential can conduct impulses.
5. Examples are: • Receptor potential at sensory nerve endings. • Motor end plate potential.	Examples are action potential of a nerve fibre, skeletal muscle and cardiac muscle.

APPLIED ASPECTS

Factors Affecting Excitability

1. ***High extracellular calcium concentration.*** A high extracellular Ca^{2+} concentration decreases membrane permeability to Na^+ ions thereby decreasing the membrane excitability.
2. ***Local anaesthetics.*** Local anaesthetic agents like procaine, tetracaine and lidocaine block the Na^+ channels thus reduce the membrane excitability.
3. **Hypocalcaemia** (low concentration of extracellular Ca^{2+}) increases excitability of nerve fibres by decreasing amount of depolarization required for opening of voltage-gated Na^+ channels.

CONDUCTIVITY

Conductivity refers to the propagation of nerve impulse (action potential) in the form of a wave of depolarization through the nerve fibre. Mechanism of conduction of action potential along an unmyelinated nerve fibre and a myelinated nerve fibre is described below.

PROPAGATION OF ACTION POTENTIAL IN AN UNMYELINATED AXON

The steps of propagation of action potential along an unmyelinated axon are summarized:

- In the resting phase (polarized state), the axonal membrane is outside positive and inside negative (Fig. 2.1-14A).
- When an unmyelinated axon is stimulated at the middle of the nerve fibre the impulse is conducted in both the direction (Fig. 2.1-14B), but when stimulated at one site by a threshold stimulus, there occurs action potential at that site, i.e. that site is depolarized. In other words, at that site outside becomes negative and inside positive (reversal of polarity) but the neighbouring areas until now remain in polarized state (Fig. 2.1-14C).
- As ECF and ICF are both conductive to electricity, a current will flow from positive polarized area to negative activated area through the ECF and in the reverse direction in the ICF (Fig. 2.1-14C). Thus, *local circuit* current flows between the resting polarized site to the depolarized site of the membrane (current sink).
- This circular current flow depolarizes the neighbouring area of the membrane up to the firing level and a new action potential is produced which in turn depolarizes the neighbouring area ahead. Thus, due to successive depolarization of the neighbouring area, the action potential is propagated along the entire length of the axon (Fig. 2.1-14D). This type of conduction is known as *electrotonic conduction.* Once initiated, the moving impulse cannot spread to reverse direction because the proximal site is in *refractory state* and thus the distal sites being in polarized state keep on getting depolarized. Thus, the direction of propagation of impulse is that of current flow inside the nerve fibre (Fig. 2.1-14E).
- The depolarization remains at any site for some length of time, therefore, portion which depolarizes first also repolarizes first. Thus, the repolarization is also propagated following the depolarization (Fig. 2.1-14F).

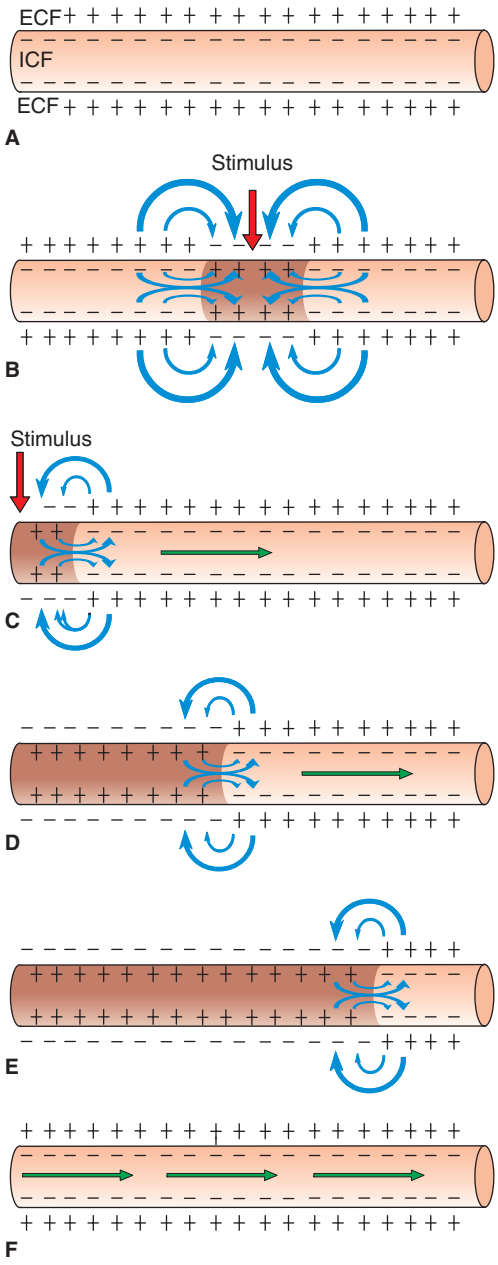


Fig. 2.1-14 Electrotonic conduction of impulse in an unmyelinated nerve fibre: A, resting phase (polarized state); B, conduction of impulse in both directions when stimulus applied at the middle of nerve fibre; C, when stimulus applied at one end of the nerve fibre; D and E, propagation of impulse in one direction along the nerve fibre and F, repolarization (occurs in same direction).

PROPAGATION OF ACTION POTENTIAL IN A MYELINATED AXON

The myelinated nerve fibres have a wrapping of myelin sheath with gaps at regular intervals which are devoid of myelin sheath (nodes of Ranvier). The myelin sheath acts as an insulator and does not allow the current flow. Therefore, in myelinated nerve fibres the local circuit of current flow only occurs from one node of Ranvier to the adjacent node (Fig. 2.1-15). That is, the impulse (action potential) jumps from one node of Ranvier to next. This is known as *saltatory conduction.* Since the impulse jumps from one node to other, the speed of conduction in the myelinated fibres is much rapid (50–100 times faster) than the unmyelinated fibre.

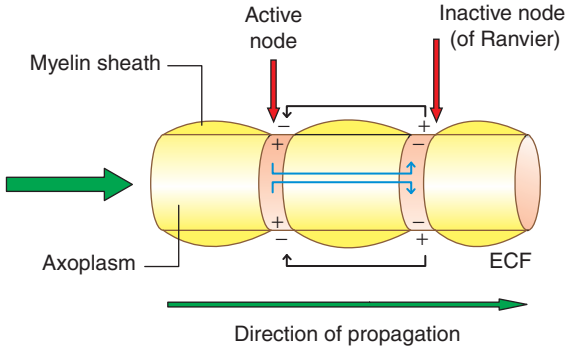


Fig. 2.1-15 Saltatory conduction along a myelinated nerve fibre.

ORTHODROMIC VERSUS ANTIDROMIC CONDUCTION

Normally, the action potential is propagated in one direction. That is, usually the nerve impulse from the receptors or synaptic junctions travels along the entire length of axon to their termination. This type of conduction is called *orthodromic conduction.* The conduction of nerve impulse in the opposite direction, as seen in the sensory nerve supplying the blood vessels, is called *antidromic conduction.*

CONDUCTION VELOCITY

FACTORS AFFECTING CONDUCTION VELOCITY

The velocity of conduction in the nerve fibres varies from as little as 0.25 m/s in very small unmyelinated fibres to as high as 100 m/s in very large myelinated fibres. In general, the factors affecting conduction velocity are:

1. ***Temperature.*** A decrease in temperature delays conduction, i.e. slows down the conduction velocity.
2. ***Axon diameter*** affects the conduction velocity through the *resistance offered by the axoplasm* (Ri) to the flow of axoplasmic current. If the diameter of the axon is greater, the Ri is lesser and hence the velocity of conduction is higher.
3. ***Myelination*** increases conduction velocity by its following effects:
   - By increasing the axon diameter and
   - By the saltatory conduction produced due to its insulating effect (as discussed above).

RECORDING OF MEMBRANE POTENTIAL AND ACTION POTENTIALS

The recording of membrane excitation and action potential is made possible by the use of highly sophisticated equipment. The mammalian axons are about 20 µm or less in diameter, so being relatively small, it is very difficult to separate them out from other axons. In certain invertebrate species like in crab (carcinus), cuttlefish (sepia) and squid (Loligo) have giant cells. The largest axon found in the neck region of Loligo is about 1 mm in diameter. The fundamental properties are quite similar to the human axons. Therefore, these animals can be used for recording various events.

INSTRUMENTS USED FOR RECORDING

The essential instruments used in recording the activity of excitable tissue are:

- Microelectrodes,
- Electronic amplifiers and
- Cathode ray oscilloscope.

Basic principles of the functioning of these instruments are discussed briefly.

1. Microelectrodes

The microelectrodes are usually very small pipette (micropipette) of tip size less than 1 µm diameter. The tip of micropipette electrode is impaled through the cell membrane of the nerve fibre. Another electrode called *indifferent electrode* is placed in the extracellular fluid (Fig. 2.1-16).

The microelectrode can be connected to the cathode ray oscilloscope through a suitable amplifier for recording rapid changes in membrane potential during nerve impulse transmission.

2. Electronic amplifier

This device can magnify the potential changes of the tissue to more than thousand times so that these can be recorded on the oscilloscope screen.

3. Cathode ray oscilloscope

Cathode ray oscilloscope (CRO) is almost an inertialess instrument which can record and measure the electrical

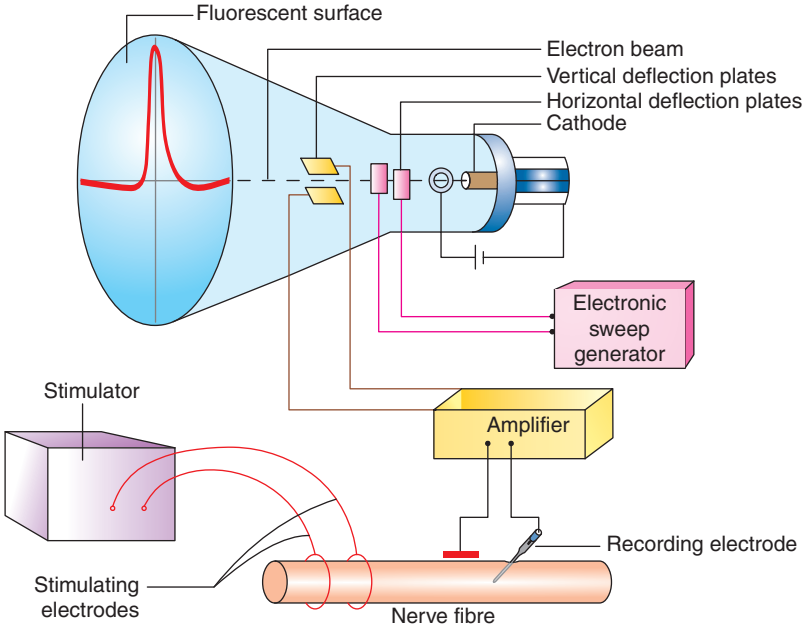


Fig. 2.1-16 Cathode ray oscilloscope and simplified diagram to record action potential from a nerve fibre.

events of living tissues instantaneously. The CRO primarily consists of a glass tube with a cathode, fluorescent surface (screen) and two sets of electrically charged plates (Fig. 2.1-16).

Cathode is in the form of an electronic gun. When connected to a suitable anode and electric current is passed, the electronic gun emits electrons.

Fluorescent surface (the face of glass tube coated with fluorescent material) acts as a screen. The electrons emitted from the cathode are directed into a beam which hits this screen.

Electrically charged plates are arranged in two sets:

- *Horizontal deflection plates* are placed on either side of the electron beam. These are connected to sweep generator (*electronic sweep circuit*). When a voltage is applied across these plates, the beam of electrons (being negatively charged) is attracted towards the positively charged plate and repelled away by the negatively charged plate. If saw-tooth voltage (i.e. voltage increased slowly, then suddenly reduced, and again slowly increased) is applied the electron beam will steadily move towards the positively charged plate (with slowly increasing voltage), then back to its original position (with sudden reduction in the voltage) and again move towards the positively charged plate. In this way the electron beam is made to sweep across the fluorescent screen horizontally, which will give a continuous marking of line of glowing light.
- *Vertical deflection plates* are placed above and below the electron beam. These are connected to recording electrodes placed on the nerve through an electronic amplifier. The potential change occurring in the nerve will charge these plates which will cause vertical (upward and downward) deflection of the electron beam. The magnitude of deflection will be proportional to the potential difference between the two plates. Thus action potential is recorded as vertical deflection of the beam as it moves across the fluorescent screen of the cathode ray oscilloscope.

RECORDING OF RESTING MEMBRANE POTENTIAL

The arrangement of electrodes for recording membrane potential is shown in Fig. 2.1-7B. The electrical potential is recorded at the surface of the membrane (exterior) and inside the membrane at each point in a nerve fibre. Starting from one end of the nerve fibre (left side of the figure) and passing to the other end (right side) and when both the electrodes (indifferent and micro electrodes) are on the surface (outside) of the nerve fibre membrane, then no potential difference was recorded (zero potential), which is potential of extracellular fluid. However, when microelectrode is made to penetrate into the interior of cell, a constant potential difference (–70 mV) is observed. This is known as resting membrane potential (Fig. 2.1-7).

RECORDING OF ACTION POTENTIAL

Monophasic recording of action potential

For monophasic recording of action potential, one microelectrode is placed inside the nerve fibre and the other electrode on the outside surface. These electrodes are then connected to the cathode ray oscilloscope (Fig. 2.1-16). When the nerve is stimulated, as shown in Fig. 2.1-17,

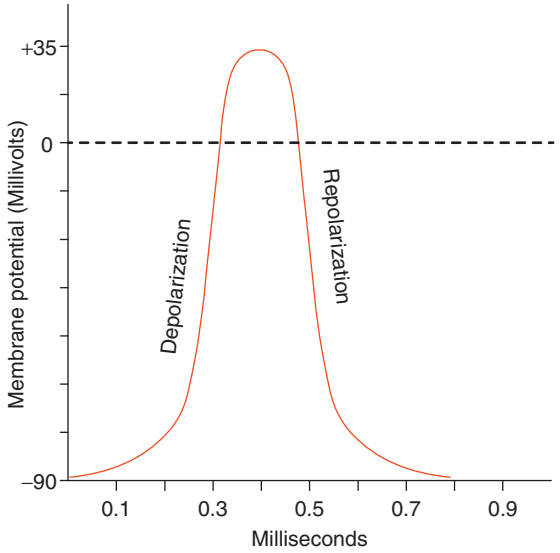


Fig. 2.1-17 A typical monophasic record of action potential.

a typical monophasic record of action potential obtained has the following components:

- Depolarization and
- Repolarization.

It has been described in detail on page 56.

Biphasic recording of action potential

For biphasic recording of action potential, both the exploring electrodes (x and y) are placed on the outside surface of nerve fibre, and they are connected to a CRO (Fig. 2.1-18). When nerve is stimulated, the excitation of one electrode will cause deflection opposite to that of another during the passage of impulse. Record of this alternate deflection one negative below the baseline and one positive above the baseline (called *biphasic action potential*) is depicted (Fig. 2.1-19).

COMPOUND ACTION POTENTIAL

Compound action potential is the monophasic recording of action potential from a *mixed* nerve which contains different types of nerve fibres with varying diameter. Therefore, the compound action potential represents an algebraic summation of the all or none action potentials of many axons.

Response of a mixed nerve to stimulus

Response of a mixed nerve to a stimulus will depend upon the thresholds of the individual axon in the nerve and their distance from the stimulating electrode (Fig. 2.1-20).

- Subthreshold stimulus stimulates none of the axons and thus no response is observed.

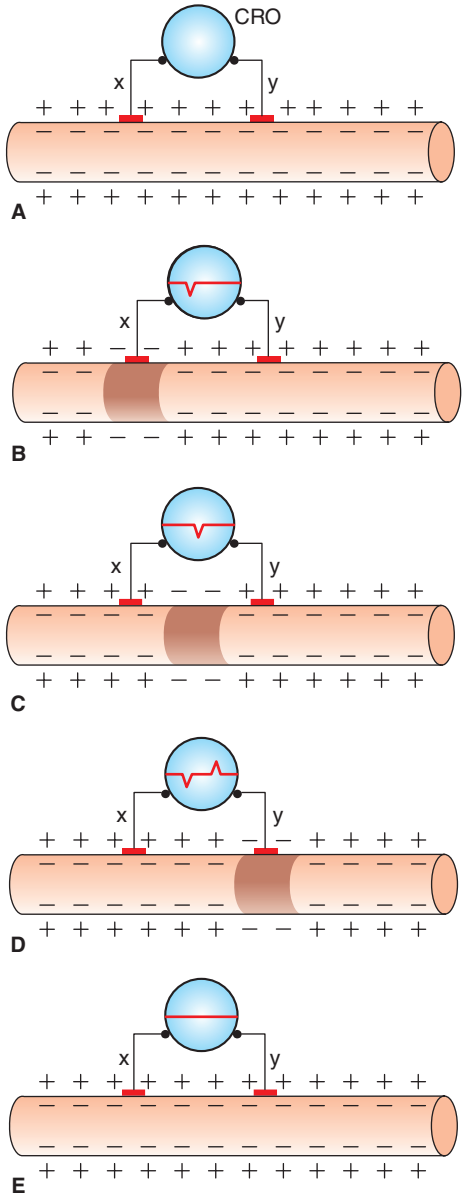


Fig. 2.1-18 Arrangement for recording biphasic action potential (recording electrodes x and y are on the surface of nerve fibre connected to cathode ray oscilloscope): A, resting state; B, on stimulation when wave of excitation reaches at electrode 'x'; C, when impulse passes beyond electrode 'x'; D, when impulse reaches at electrode 'y' and E, when impulse passes away from electrode 'y'.

- *Threshold intensity stimulus* makes the axons with low threshold to fire and thus a small potential change is observed.
- *With further increase in the intensity of stimulus,* the axons with higher thresholds also fire. The electrical response increases proportionately until the stimulus is strong enough to excite all of the axons in the nerve. The stimulus which excites all the axons is called *maximal stimulus.*
- *Supramaximal stimulus* does not produce any further increase in the potential.

It has a unique shape (multiple peaks) because a mixed nerve is made up of different fibres with various speeds of conduction.

- The number and size of the peaks vary with the types of fibres in the particular nerve being studied.
- When less than maximal stimuli are used, the shape of compound action will depend upon the number and type of fibres stimulated.

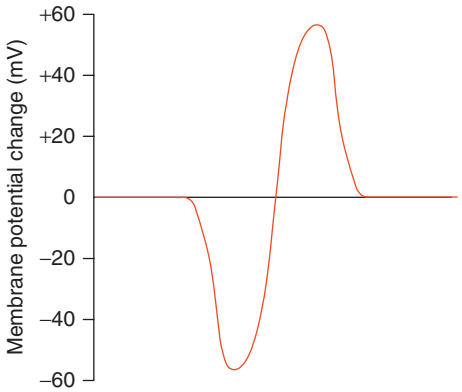


Fig. 2.1-19 A typical complete record of biphasic action potential.

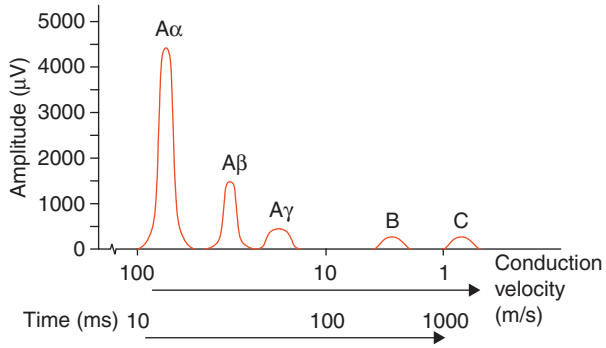


Fig. 2.1-20 Typical record of compound action potential (recorded from a mixed nerve) showing multiple peaks.

NERVE FIBRE TYPES

Various schemes for classification of the nerve fibres on the basis of their diameter and their conduction velocity have been proposed.

CLASSIFICATION OF NERVE FIBRES

1. Letter classification of Erlanger and Gasser

This is the best known classification based on the diameter and conduction velocity of the nerve fibres. The nerve fibres have been classified as follows:

'Type A' nerve fibres. The fastest conducting fibres are called type A fibres. Their diameter varies from 12–20 μm and conduction velocity from 70–120 m/s. They are myelinated fibres.

Type A fibres have been further subdivided into α, β, γ and δ. Type A fibres subserve both motor and sensory functions.

'Type B' nerve fibres. These fibres are myelinated have a diameter of less than 3 μm and their conduction velocity varies from 4–30 m/s. They form preganglionic autonomic efferent fibres, afferent fibres from skin and viscera, and free nerve ending in connective tissue of muscle.

'Type C' nerve fibres. These are unmyelinated, have a diameter of 0.4–1.2 μm and their conduction velocity varies from 0.5–4 m/s. These form the postganglionic autonomic fibres, some sensory fibres carrying pain sensations, some fibres from thermoreceptors and some from viscera.

The salient features of type A, B and C nerve fibres are summarized in Table 2.1-2.

2. Numerical classification

Some physiologists have classified sensory nerve fibres by a numerical system into type Ia, Ib, II, III and IV. A comparison

Table 2.1-2 Salient features of type A, B and C nerve fibres

Fibre type	Myelinated/ non-myelinated	Fibre diameter (μm)	Conduction velocity (m/s)	Spike duration (ms)	Absolute refractory period (ms)	Function	
						Efferent	Afferent
A	α Myelinated	12–20	70–120	0.4–0.5	0.4–1	Somatic motor	Proprioception
	β Myelinated	5–12	30–70	0.4–0.5	0.4–1	–	Touch, pressure
	γ Myelinated	3–6	15–30	0.4–0.5	0.4–1	Motor to muscle spindles	–
	δ Myelinated	2–5	12–30	0.4–0.5	0.4–1	–	Pain, cold, touch
B	Myelinated	<3	3–15	1.2	1.2	Preganglionic autonomic	–
C	Non-myelinated	0.4–1.2	0.5–2	2	2	Postganglionic autonomic	Pain, temperature, and some mechanoreceptors reflex responses

of the numerical classification and the letter classification is shown in Table 2.1-3.

3. Susceptibility of nerve fibres to hypoxia, pressure and local anaesthetics

Hypoxia. As shown in Table 2.1-4, the type B fibres are most susceptible to hypoxia. Since, the preganglionic autonomic fibres are of type B, therefore, hypoxia is associated with alteration of the autonomic functions in the body such as rise in heart rate, blood pressure and respiration.

Pressure. Type A fibres are most susceptible to pressure and type C least. Therefore, pressure on a nerve can produce temporary paralysis due to loss of conduction in motor, touch and pressure fibres (type A), while pain sensation (carried by type C fibres) remain relatively intact. This is common observation after sitting cross-legged for long periods and after sleeping with arms under the head.

Local anaesthetics. Type C fibres (conducting pain, touch and temperature sensations generated by cutaneous receptors) are most susceptible to local anaesthetics. This fact is useful for surgical interventions under local anaesthesia.

Table 2.1-3 Numerical *vis-a-vis* letter classification of sensory nerve fibres

Types of nerve fibre		Origin
Numerical classification	**Letter classification**	
Ia	Aα	Muscle spindle (annulospiral endings)
Ib	Aα	Golgi tendon organ
II	Aβ	Muscle spindle, (flower spray endings), touch and pressure
III	Aδ	Pain and cold receptors, some touch receptors
IV	C	Pain, temperature and other receptors

Table 2.1-4 Susceptibility of nerve fibres to hypoxia, pressure and local anaesthetics

Sensitivity to	Type of nerve fibre		
	Most susceptible	**Intermediate susceptible**	**Least susceptible**
Hypoxia	B	A	C
Pressure	A	B	C
Local anaesthetics	C	B	A

DEGENERATION AND REGENERATION OF NEURONS

When the axon of neuron is injured, a series of degenerative changes are seen at three levels:

- In the axon distal to injury,
- In the axon proximal to injury and
- In the cell body.

Along with the degenerative changes, the reparative process (regeneration) also starts soon if the circumstances are favourable. The effects of injury to a nerve and the occurrence of regenerative changes, thereafter will depend upon the degree and type of damage.

CAUSES AND GRADING OF NERVE INJURY

Common causes of nerve injury are: transection (through and through cut), crushing of nerve fibres, local injection of toxic substances, ischaemia due to obstruction of blood flow and effects of hyperpyrexia on the neurons.

GRADING OF NERVE INJURY

Sunderland had graded the injury to nerves fibres in order of severity into following degrees:

- ***First degree injury*** involves only transient loss of function resulting from a mild pressure on the nerve. The lost function of the nerve fibres returns within few hours to few weeks of the removal of causative pressure, since the axon is not destroyed.
- ***Second degree injury*** includes severe nerve damage with intact endoneural tube. It results from a severe and prolonged pressure on the nerve. The nerve fibre is severely damaged at the pressure point and is followed by degenerative changes. However, the regeneration and restoration of the function of the nerve are facilitated as the endoneural tube remains intact.
- ***Third degree injury*** includes severe damage to the nerve fibre with interruption of the endoneural tube.
- ***Fourth degree injury*** refers to a severe damage to the nerve fibres associated with disorganization of nerve fasciculi.
- ***Fifth degree injury*** is labelled when there occur complete transection, i.e. the nerve fibres are cut through and through. The degenerative changes are initiated very early following fifth degree injury to the nerve fibres.

STAGE OF DEGENERATION

The degenerative changes which occur in the part of axon distal to the site of injury are referred to as an anterograde degeneration or Wallerian degeneration (after the discoverer A Waller, 1862). The degenerative changes occurring in the neuron proximal to the injury are referred to as retrograde degeneration. These changes take place in the cell body and in the axon proximal to injury.

CHANGES IN THE PART OF AXON DISTAL TO INJURY

The degenerative changes start within few hours of injury and continue for about 3 months and include the following (Fig. 2.1-21):

Axis cylinder

Axis cylinder becomes swollen and irregular in shape within a few hours of injury. After a few days it breaks up into small fragments, the neurofibrils within it break down into granular debris and seen in the space occupied by axis cylinder.

Myelin sheath shows slow disintegration which starts on eighth day and continues up to 32nd–35th day. In fact, myelin sheath is converted into fat droplets containing cholesterol esters.

Neurilemmal sheath is usually unaffected but the Schwann cells start multiplying rapidly.

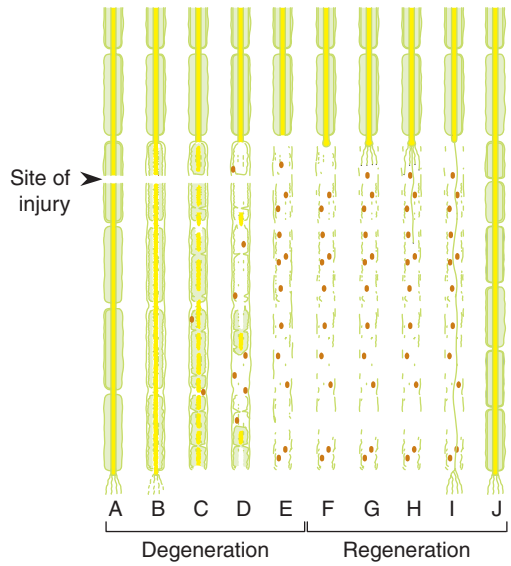


Fig. 2.1-21 Degenerative changes (Wallerian degeneration) in distal part of a nerve fibre after injury (A–E) and subsequent regenerative changes (F–J).

Macrophages invade the region and remove degenerating axons, myelin and cellular debris and thus the neurilemmal tube becomes empty.

Schwann cell's cytoplasm proliferates rapidly and fills up the empty neurilemmal tube. These cells produce a large series of membranes that help to form numerous tubes which play a vital role in the regeneration of nerve fibres.

Changes in the cell body of neuron

Changes in the cell body of injured neuron start within 48 h and continue up to 15–20 days. The changes are:

Nissl substances undergo disintegration and dissolution (chromatolysis).

Golgi apparatus, mitochondria and neurofibrils are fragmented and eventually disappear.

Cell body draws in more fluid, enlarges and becomes spherical.

Nucleus is displaced to the periphery (towards cell membrane). Sometimes the nucleus is extruded out of the cell, in which case the neuron atrophies and finally disappears completely.

STAGE OF REGENERATION

The stage of degeneration is followed by the stage of regeneration under favourable circumstances (listed below). It starts within 4 days of injury but becomes more active after 30 days and may take several months to one year for complete recovery.

FACTORS AFFECTING REGENERATION

- Regeneration occurs more rapidly when a nerve is crushed than when it is severed and the cut ends are separated.
- Chances of regeneration of a cut nerve are considerably increased if the two cut ends are near each other (gap does not exceed 3 mm) and remain in the same line.
- Presence of neurilemma is must for regeneration to occur. Therefore, axons in the CNS once degenerated never regenerate as these nerve fibres have no neurilemma.
- Presence of nucleus in the neuron cell body is also must for regeneration to occur. If it is extruded, the neuron is atrophied and the regeneration does not occur.

REGENERATIVE CHANGES

Anatomical regeneration

1. Changes in the axon

Stage of fibres formation. Axis cylinder from the proximal cut end of the axon elongates and gives out fibrils up

to 100 in number in all directions. These branches grow into the connective tissue at the site of injury in an effort to reach the distal cut end of the nerve fibre (Fig. 2.1-21F and G).

Stage of entry of fibrils into endoneural tube. Strands of the Schwann cells from the distal cut end of axon guide the regenerating fibrils to enter their axon endoneural tube, once the fibril enters the endoneural tube it grows rapidly within it. The axonal fibrils that fail to enter one of the tubes degenerate. It often happens that more than one fibril (1–25) enters the same tube. Under such circumstances, the largest fibril survives and the rest others degenerate.

Stage of active growth. The axonal fibril growing through the endoneural tube enlarges and establishes contact with an appropriate peripheral end organ. The new axon formed in this way is devoid of myelin sheath (Fig. 2.1-21H and I). The process of regeneration up to this stage takes about 3 months.

Stage of myelination. The myelin sheath is then formed by the cells of Schwann slowly. The myelination is completed in 1 year.

2. Changes in the cell body of neuron (Fig. 2.1-22)

- Nissl granules followed by the Golgi apparatus appear in the cell body.
- The cell loses excess fluid and regains its normal size.
- The nucleus occupies the central position.
- The above changes in some of the neurons start within 20 days of injury and are completed in 80 days.

Functional regeneration

The above described regenerative changes constitute the anatomical regeneration. The functional (physiological) recovery, however, occurs after a long period.

FACTORS PROMOTING NEURONAL GROWTH

Various factors affecting neuronal development, growth and survival have been isolated and studied. These can be broadly arranged into two groups:

- Neurotrophins and
- Other factors affecting neuronal growth.

NEUROTROPHINS

Neurotrophins are the proteins which provide trophic support to the neurons, i.e. they promote nerve growth and survival. They also have a role in neuronal plasticity.

Production. The neurotrophins are produced by the muscles or other structures especially the glands that the neurons innervate. Some of the neurotrophins are produced by the astrocytes.

Established neurotrophins include:

- Nerve growth factor
- Brain-derived neurotrophic factor
- Neurotrophin-3 and 4/5

1. Nerve growth factor. Nerve growth factor (NGF) is probably the first neurotrophin recognized. It promotes the growth of *sympathetic nerves and* some sensory nerves.

- It is found in various tissues in human beings and many species of animals. NGF is particularly found in high concentration in the submaxillary salivary glands of the male mice.
- NGF is made up of 2α, 2β and 2γ subunits.
  - The α subunits have trypsin-like activity.
  - The β subunits are similar in structure to insulin and possess all the nerve-growth promoting activities.

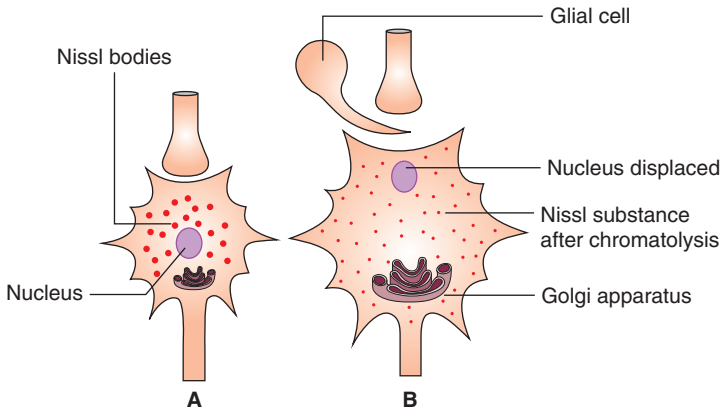


Fig. 2.1-22 Changes in cell body of a neuron after injury (B) and normal cell body of a neuron (A).

  - The γ subunits are serine proteases. Their functions are unknown.
- Receptor of NGF is Trk A (tyrosine kinase activity A).
- NGF is also present in the brain and appears to be responsible for the growth and maintenance of cholinergic neurons in the basal forebrain and striatum.
- Immunosympathectomy, i.e. complete destruction of the sympathetic ganglia is produced when antiserum against NGF is injected in a newborn animal.

2. Brain-derived neurotrophic factor. Brain-derived neurotrophic factor (BDNF) has a growth promoting role for peripheral sensory nerves. It has been seen that BDNF-deficient mice lose peripheral sensory nerves and have severe degenerative changes in their vestibular ganglia. Receptor for BDNF is Trk B.

3. Neurotrophin-3

- The neurotrophin-3 (NT-3) plays a growth promoting role in cutaneous mechanoreceptors, since its disruption by gene knockout gives rise to marked loss of these receptors.
- Receptors for NT-3 are Trk C (mainly) and Trk A and B (to some extent).

4. Neurotrophin-4/5

- Receptors for neurotrophin-4/5 is Trk B.

OTHER GROWTH PROMOTING FACTORS

Factors, other than neurotrophins, which promote nerve growth include:

- Ciliary neurotrophic factor
- Leukaemia inhibitory factor
- Insulin-like growth factor-I
- Transforming growth factor
- Fibroblast growth factor
- Platelet-derived growth factor

Chapter 2.2

Neuromuscular Junction

STRUCTURE OF NEUROMUSCULAR JUNCTION

- Terminal button
- Presynaptic membrane
- Synaptic cleft
- Postsynaptic membrane

NEUROMUSCULAR TRANSMISSION

- Release of acetylcholine by the nerve terminals
- Effect of acetylcholine on the postsynaptic membrane
- Development of end plate potential
- Removal of acetylcholine

DRUGS AFFECTING AND DISORDERS OF NEUROMUSCULAR JUNCTION

- **Drugs affecting neuromuscular junction**
  - Neuromuscular blockers
  - Neuromuscular stimulators
- **Disorders of neuromuscular junction**
  - Myasthenia gravis
  - Lambert–Eaton syndrome

STRUCTURE OF NEUROMUSCULAR JUNCTION

Neuromuscular junction refers to the intimate contact of the nerve endings with the muscle fibre to which they innervate. Characteristics of the nerve and muscle fibre at and near the neuromuscular junction are as given (Fig. 2.2-1):

Terminal button. The axon of a neuron supplying a skeletal muscle loses its myelin sheath and divides into a number of fine branches which end in small swellings (knobs) called *terminal buttons* or end feet which forms a neuromuscular junction, at the centre of muscle fibre (Fig. 2.2-1A). The nerve terminal or the so-called synaptic knob contains a large number of vesicles (about three lac) containing acetylcholine and mitochondria. The acetylcholine is synthesized by the mitochondria and is stored in the vesicles (Fig. 2.2-1B).

Presynaptic membrane. This refers to the axonal membrane lining the terminal buttons of the nerve endings.

Synaptic cleft. It is a 50–100 nm wide space between the presynaptic membrane and the postsynaptic membrane. It is filled by extracellular fluid with reticular fibres forming the matrix.

Postsynaptic membrane (Fig. 2.2-1B). This is the name given to the muscle fibre membrane (sarcolemma) in the region of neuromuscular junction. The muscle membrane in this region is thickened and depressed to form the *synaptic trough* in which the terminal button fits. This thickened portion of the muscle membrane is also called *motor end plate.* The postsynaptic membrane contains *receptor sites* for acetylcholine called the *nicotinic receptors.* The matrix of cleft contains enzyme cholinesterase which hydrolyzes acetylcholine.

Synthesis and storage of acetylcholine

Mitochondria in the terminal buttons synthesize acetylcholine from choline, Acetyl Co enzyme-A, ATP and glucose in the presence of enzyme choline acetylase (choline transferase). Acetylcholine once synthesized is stored temporarily in the vesicles in small packets called quanta (consisting of approximately 10^4 molecules of ACh).

NEUROMUSCULAR TRANSMISSION

The skeletal muscle is stimulated only through its nerve. The neuromuscular junction transmits the impulses from the nerve to the muscle. The *sequence of events* which causes transmission of impulse through the neuromuscular junction is:

- Release of acetylcholine by the nerve terminals
- Effect of acetylcholine on the postsynaptic membrane
- Development of end plate potential

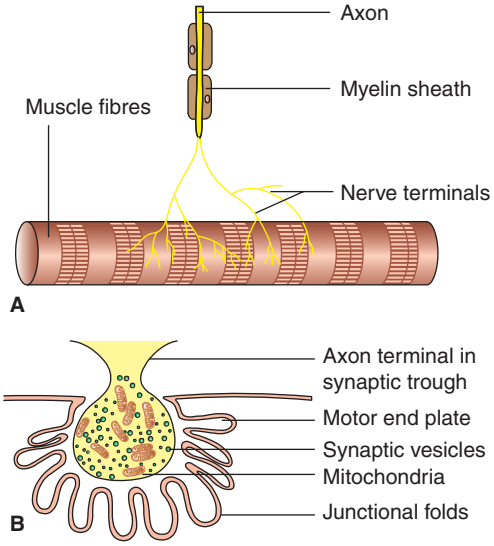


Fig. 2.2-1 Structure of neuromuscular junction: A, the axon of the neuron loses its myelin sheath and divides into fine branches and B, structure of terminal button and motor end plate.

- Miniature end plate potential
- Removal of acetylcholine by cholinesterase
- Initiation of the action potential in the muscle fibre.

Release of acetylcholine by the nerve terminals

When the nerve impulse (action potential) travelling in the nerve fibre (axon) reaches the terminal buttons, the voltage-gated Ca^{2+} channels present on the presynaptic membrane open up, increasing its permeability to Ca^{2+} ions. Consequently, the Ca^{2+} ions present in the extracellular fluid (ECF) of the synaptic cleft enter the terminal buttons. The elevated Ca^{2+} levels in the cytosol of terminal buttons trigger a marked increase in exocytosis of vesicles releasing acetylcholine in the synaptic cleft (Fig. 2.2-2).

Effect of acetylcholine on the postsynaptic membrane

The acetylcholine so released diffuses in the synaptic cleft and binds to the nicotinic–acetylcholine receptors located mainly on the junctional folds of the motor end plate (postsynaptic membrane) leading to opening up of the tubular channels.

Development of end plate potential

Due to opening of the acetylcholine-gated channels in the end plate membrane, a large number of Na^{+} ions from the ECF enter inside the muscle fibre causing a local positive potential change inside the muscle fibre membrane called the *end plate potential*.

The end plate potential is non-propagative but when a critical level of –60 mV is reached, it triggers the development of action potential in the muscle fibre (Fig. 2.2-3).

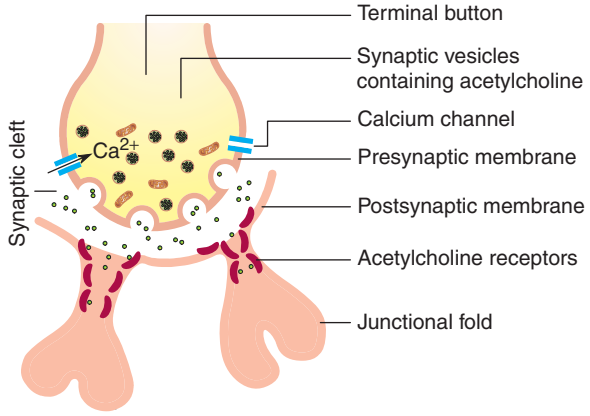


Fig. 2.2-2 Detailed structure of neuromuscular junction showing Ca^{2+} channels on the presynaptic membrane and the ACh receptors on postsynaptic membrane.

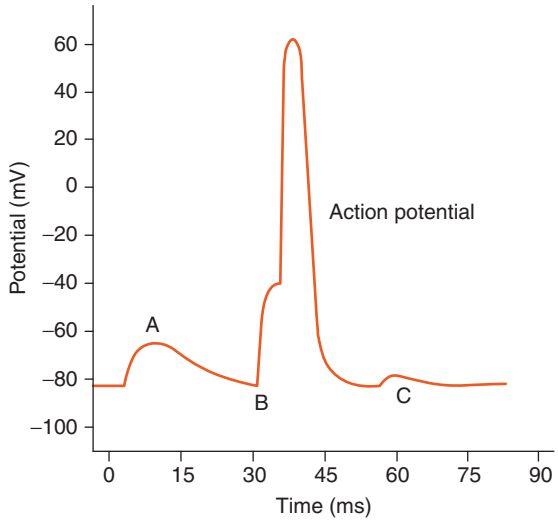


Fig. 2.2-3 End plate potential and development of action potential in muscle fibre: A, weak end plate potential (<–60 mV); B, end plate potential triggers propagating action potential (>–60 mV) and C, miniature end plate potential.

The action potentials are generated on either side of the end plate and conducted away from the end plate in both the directions along the muscle fibres thus causing muscle contraction.

Miniature end plate potential. Even at rest, small quanta of acetylcholine are released randomly from the nerve terminal because of random Brownian movement of axoplasm. Each quantum of acetylcholine produces a weak end plate potential about 0.5 mV in magnitude. This is called *miniature end plate potential.*

Giant end plate potential. Due to more release of ACh, the end plate potential rises up to 12 mV, yet not sufficient to generate action potential is known as *giant end plate potential.*

Removal of acetylcholine

The acetylcholine released in the synaptic cleft stays for a short period and is removed within 1 ms by the enzyme acetylcholinesterase which is present in the matrix of synaptic cleft. A small amount of ACh diffuses back into nerve terminals from the synaptic cleft.

It is important to note that the rapid removal of acetylcholine prevents the repeated excitation of muscle fibre.

DRUGS AFFECTING AND DISORDERS OF NEUROMUSCULAR JUNCTION

DRUGS AFFECTING NEUROMUSCULAR JUNCTION

Neuromuscular blockers

Neuromuscular blockers are the drugs that block transmission at the neuromuscular junction. Some of the common neuromuscular blockers, which are commonly used, in clinical practice and in research are:

1. ***Curare.*** Curare or the active principle of D-tubocurarine prevents the neuromuscular transmission by combining with acetylcholine receptors. The acetylcholine released thus cannot combine with the receptors and so end plate potential does not develop. The curariform drugs are called *receptor blockers* since they block the neuromuscular transmission by acting on the acetylcholine receptors.
2. ***Bungarotoxin*** found in the venom of deadly snakes also blocks neuromuscular transmission by binding with the acetylcholine receptors.
3. ***Succinylcholine and carbamylcholine*** act like acetylcholine and cause depolarization of the postsynaptic membrane. But, these are not destroyed by cholinesterase and so the muscle remains in a depolarized state for a long time. Thus, these drugs block the myoneural junction by keeping the muscle in a depolarized state.
4. ***Botulinum toxin*** is derived from the bacteria *Clostridium botulinum*. It blocks the transmission across the myoneural junction by preventing the release of acetylcholine from the terminal buttons of the nerve endings.

Neuromuscular stimulators

Drugs having acetylcholine like action. The drugs methacholine, carbachol and nicotine act like acetylcholine and produce end plate potential exciting the muscle fibre. However, these drugs are either not destroyed or are destroyed very slowly by the enzyme acetylcholinesterase. So they cause repeated stimulation and continuous action of muscle, thereby causing a state of muscle spasm.

Drugs that inactivate the enzyme cholinesterase (anticholinesterase). The drugs like neostigmine, physostigmine and diisopropyl fluorophosphate (DFP) stimulate the neuromuscular junction by inactivating the enzyme acetylcholinesterase. Once this enzyme is inactivated, the acetylcholine released at the nerve terminal cannot be hydrolysed, this leads to repeated stimulation and continuous action of muscle.

The effect of neostigmine and physostigmine lasts for several hours while that of DFP lasts for several weeks. The DFP is thus a lethal poison which can cause death due to laryngeal spasm.

DISORDERS OF NEUROMUSCULAR JUNCTION

Myasthenia gravis

Myasthenia gravis is a disorder in which the myoneural junction is unable to transmit signals from the nerve fibre to the muscle fibres, thereby causing paralysis of the involved muscles. Myasthenia is probably an autoimmune disease. In this disease, the antibodies are produced against the acetylcholine-gated channels (receptors) present on the motor end plate which destroy these channels. Thus, the acetylcholine released at the nerve terminal is not able to produce adequate end plate potential to excite the muscle fibre. If the disease is intense enough, the patient dies of paralysis—in particular, paralysis of respiratory muscles.

Lambert–Eaton syndrome

In this disease, antibodies are produced against the calcium channels present on the presynaptic membrane which destroy the channels. Consequently, Ca^{2+} influx into the nerve terminal is markedly decreased and thereby release of acetylcholine is also reduced. Scanty amount of acetylcholine is not able to produce adequate end plate potential to excite the muscle fibres producing muscular weakness.

Chapter 2.3

Skeletal Muscle

INTRODUCTION

The muscle cell, like the neuron, is an excitable tissue, i.e. an action potential is generated when it is stimulated either chemically, electrically or mechanically. Further, the muscle is a contractile tissue with a chemically stored energy which can be transformed into mechanical energy.

There are three different types of muscles in the body: skeletal muscles, cardiac muscles and smooth muscles. Based on certain distinctive features the muscles can be grouped as:

Striated versus non-striated muscles

Striated muscle cells show large number of cross-striations at regular intervals when seen under light microscope. Skeletal and cardiac muscles are striated.

Non-striated muscle cells do not show any striations. Smooth muscles or the so-called plain muscles are non-striated.

Voluntary versus involuntary muscles

Voluntary muscles can be made to contract under our will to perform the movements we desire. All skeletal muscles are voluntary muscles. These are supplied by the somatic motor nerves.

Involuntary muscle's activities cannot be controlled at will. Cardiac and all smooth muscles are involuntary muscles. These are innervated by the autonomic nerves.

Skeletal muscles

The skeletal muscles, as the name indicates, are attached with the bones of the body skeleton and their contraction results in the body movements. The skeletal muscles constitute about 40% of the total body mass.

FUNCTIONAL ANATOMY AND ORGANIZATION

STRUCTURAL ORGANIZATION OF MUSCLE

Structurally, the skeletal muscle consists of a large number of muscle fibres and a connective tissue framework organized as (Figs 2.3-1 and 2.3-2):

- Each *muscle fibre* is surrounded by a delicate connective tissue called *endomysium* which contains large quantity of elastic tissue arranged longitudinally.
- The muscle fibres are grouped into a number of bundles called *fasciculi.* Each fasciculus is surrounded by a stronger sheath of connective tissue called *perimysium.*
- All the fasciculi collectively form the *muscle belly.* The connective tissue that surrounds the entire muscle belly is called *epimysium.*
- At the junction of the muscle with its tendon, the fibres of endomysium, perimysium and epimysium become continuous with the fibres of the tendon.
- *Tendons* are fibrous terminal ends of the muscles made up of collagen fibres.

STRUCTURE OF A MUSCLE FIBRE

Each muscle fibre is basically a long (1–40 mm), cylindrical (10–100 µm in diameter) multinucleated cell. Its cell membrane is called *sarcolemma* and the cytoplasm is called *sarcoplasm.* Like any other cell, in the sarcoplasm are embedded many structures, the nuclei, Golgi apparatus, mitochondria, sarcoplasmic reticulum, ribosomes and glycogen and occasional lipid droplets. In addition, the sarcoplasm mainly contains number of *myofibrils* which form the main structure of a muscle fibre. The sarcolemma along with the sarcoplasmic reticulum forms the so-called *sarcotubular system.*

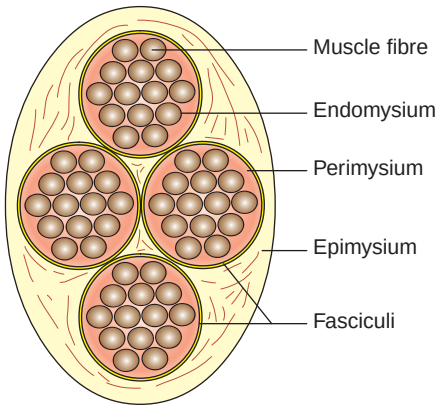


Fig. 2.3-1 Transverse section of skeletal muscle seen under light microscope.

MYOFIBRIL

Each muscle fibre consists of a large number of myofibrils which are arranged parallel to each other and running along the entire length of the muscle fibre. Myofibril is about 1–2 µm in diameter and 1–40 mm in length depending upon the length of the muscle fibre.

Each myofibril consists of many thick and thin filaments (myofilaments) made up of contractile proteins. The peculiar arrangement of these myofilaments when seen under *light microscope* gives an appearance of alternate dark and light bands (striations) as described.

STRIATIONS OF MUSCLE FIBRES

The dark and light bands result from a difference in the refractive index of its different parts. The arrangement is as below (Fig. 2.3-2D):

- The dark band is called *A band* (anisometropic to polarized light). It is 1.5 µm in length. In the area of A band the thick (myosin) filaments line up the thin filaments.

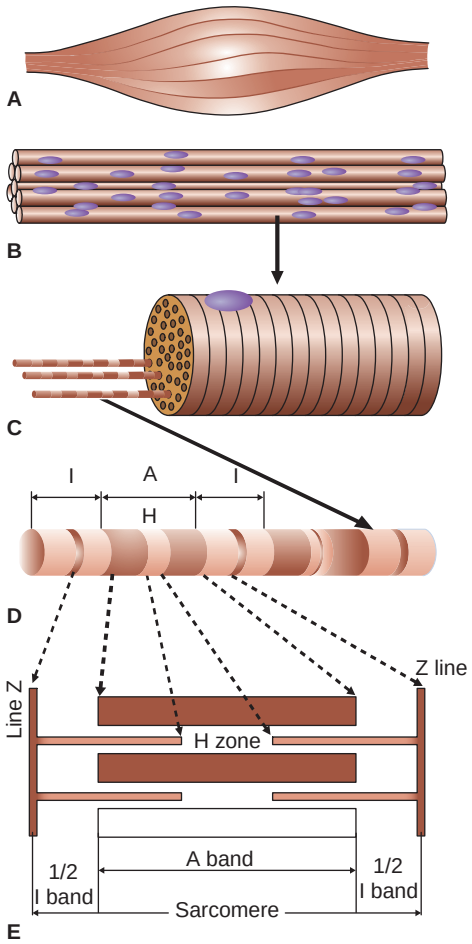


Fig. 2.3-2 Schematic diagram showing structural organization of skeletal muscle: A, muscle belly; B, muscle fibres grouped into fasciculi; C, muscle fibre; D, myofibril and E, arrangement of thick and thin filaments.

- In the centre of each A band there is a lighter *H zone* where thin filaments do not overlap the thick filaments. (The word H either represents the discoverer, Henson or the hell which in German means light).
- In the centre of each H zone is seen *M line,* which is more pronounced during muscle contraction.
- The light band is called *I band,* because it is isotropic to polarized light. It is about 1 µm in length. This area contains only thin (actin) filaments.
- Each I band is bisected by a narrow dark *Z line* (the word Z has been taken from Z Wischenscheibe which in German means between discs).
- The portion of myofibril between two successive Z lines is called a *sarcomere.* Thus a sarcomere includes ½ I band, +1A band and ½ I band, and is about 2.5 µm in length at rest. The sarcomere is the structural and functional unit of the muscle fibre. During muscle contraction the sarcomere reduces in length to 1.5 µm and during stretching of the muscle it increases in length to 3.5 µm.

THICK AND THIN FILAMENTS

The thick and thin filaments (Fig. 2.3-2E) form the contractile apparatus of a striated muscle.

Thick filament

Thick filaments are about twice the diameter of thin filaments. Each thick filament is surrounded by six thin filaments arranged in a regular hexagonal manner. A thick filament is made up of hundreds of molecules of a complex actin-binding contractile protein called myosin.

Structure of myosin molecule. The myosin molecule is 10–11 nm thick and 45 nm apart, and has a molecular weight of 4,80,000 and is made up of six polypeptide chains (two heavy chains and four light chains). The two heavy chains wrap around each other to form a double helix, which constitutes the *tail and body* of the myosin molecule (Fig. 2.3-3B). The light chains combine with the terminal part of the heavy chains to form the globular *head* of myosin molecule. The myosin molecule present in the skeletal muscle has two heads and is called *myosin-II* (Fig. 2.3-3C) (single headed myosins present in some other cells of the body is called myosin-I).

The myosin head contains an *actin binding site* and a *catalytic site* that hydrolyzes adenosine triphosphate (ATP). During muscle contraction, the head forms the *cross-bridging* (described later). Digestion with trypsin generates two fragments of myosin molecules (Fig. 2.3-4):

- *Light meromyosin* which comes from the tail part of the myosin. It does not have any ATPase activity or actin binding ability.
- *Heavy meromyosin (HMM)* contains the globular head as well as part of the tail. HMM can be further splitted by papain into two parts. The *globular HMM-SI,* which has all the ATPase activity and actin binding ability, and the fibrous HMM-S2, which has none of it (Fig. 2.3-4).

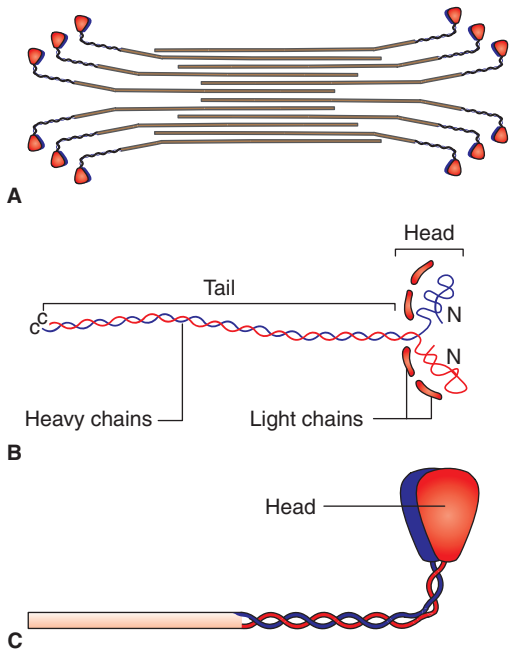


Fig. 2.3-3 Myosin molecules: A, arrangement of myosin molecules in thick filament; B, structure of myosin molecule and C, molecule of myosin II (with two heads).

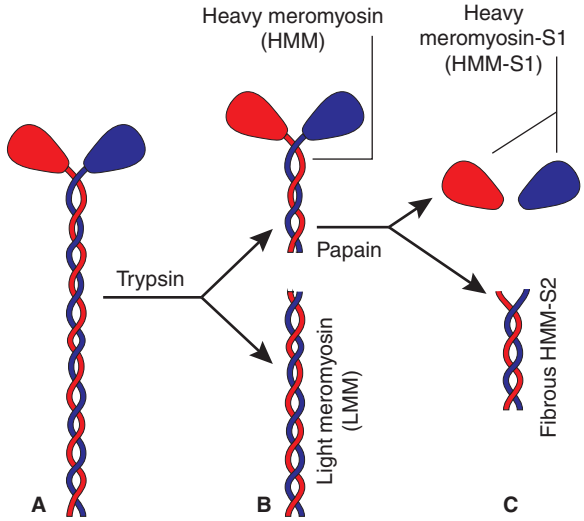


Fig. 2.3-4 Myosin molecule after digestion with trypsin and papain: A, before digestion with trypsin; B, after digestion with trypsin (form two fragments LMM and HMM) and C, heavy meromyosin fragment after digestion with papain.

Arrangement of myosin molecules in a thick filament. In a thick filament, half of the myosin molecules are oriented with their heads in one direction and the remaining half in opposite direction (Fig. 2.3-3A). Because of this arrangement

the central portion of a thick filament is devoid of the head portions of the myosin molecules. This accounts for the comparatively lighter H zone seen in the centre of dark A band.

Thin filament

Each thin actin filament is 4–5 nm in diameter and made up of contractile protein molecules (actin) and two types of regulatory protein molecules (tropomyosin and troponin) (Fig. 2.3-5).

Actin. About 300–400 actin molecules are present in each thin filament. The actin molecules form a long double helix consisting of two chains of globular units (G-actin) and the chain formed by them is designated as *fibrous actin* (F-actin).

Tropomyosin. About 40–60 tropomyosin molecules (molecular weight 70,000) are present in each thin filament. The tropomyosin molecules are long filaments which lie in the groove between the two chains of actin molecules. It covers the binding site of actin where myosin head comes in contact with actin. Thus, it is a regulatory protein which prevents the interaction between actin and myosin filaments.

Troponin. Troponin molecules are small globular units located at intervals along the tropomyosin molecule. The troponin molecule has three subunits:

- *Troponin-T* binds the other troponin components to tropomyosin.
- *Troponin-I* prevents the interaction of myosin heads with the active sites on actin.
- *Troponin-C* contains binding site for Ca^{2+} that initiates muscle contraction. Each molecule of troponin-C binds to four molecules of Ca^{2+} ions.

Arrangement of anchoring proteins of contractile apparatus

The anchoring proteins of the contractile apparatus include α-actinin, titin, nebulin and dystrophin associated glycoproteins. These are arranged as:

- *α-actinin* cross-link the actin filaments in the area of Z line (Fig. 2.3-6).
- *Titin* earlier called as *connectin* or gap filament is a large elastic filament which interconnects the Z lines. It forms the *series elastic components* (SEC) of the muscle.
- *Nebulin* is an inextensible filament which is connected at one end to the α-actinin in the area of Z line and at another end to the tropomyosin–troponin complex of thin filaments at regular intervals.

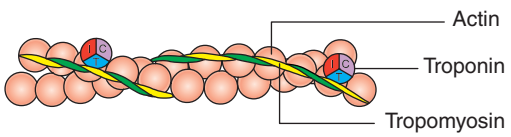


Fig. 2.3-5 Structure of a thin filament.

- *Dystrophin–glycoprotein complex.* The dystrophin–glycoprotein complex forms the best known anchor protein complex which provides structural support and strength to myofibril.

SARCOTUBULAR SYSTEM

The sarcolemma (cell membrane of muscle cell) along with the sarcoplasmic reticulum (the endoplasmic reticulum of muscle cell) forms a highly specialized system called sarcotubular system. This plays an important role in the internal conduction of depolarization within the muscle fibre. The sarcotubular system is primarily formed by a transverse tubular system (T-system) and a longitudinal sarcoplasmic reticulum (Fig. 2.3-7).

Transverse tubular system (T-system)

The T-system of transverse tubules is formed by the through and through invagination of sarcolemma into the muscle

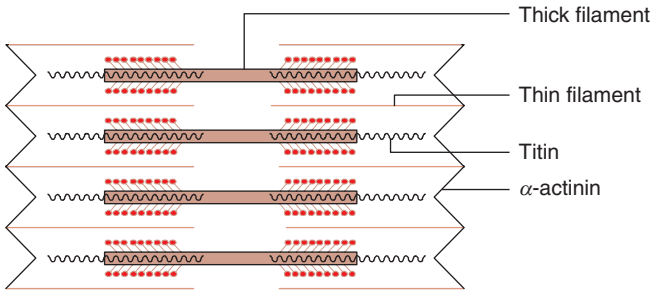


Fig. 2.3-6 Arrangement of anchoring proteins of contractile apparatus: A, α-actinin (which cross link the thin filament in area of Z lines) and titin (interconnects two Z lines).

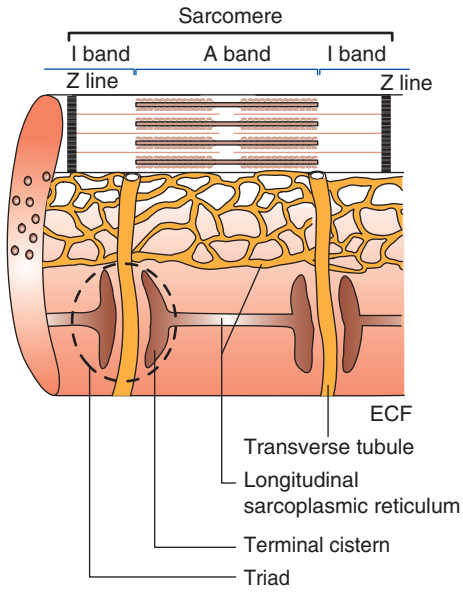


Fig. 2.3-7 Sarcotubular system showing transverse tubules and longitudinal sarcoplasmic reticulum.

fibre in the region of junction of A and I bands (Fig. 2.3-7). Since, the T-tubules are formed by the invagination of sarcolemma, their lumen contains extracellular fluid (ECF) which surrounds the muscle cell. The membrane of T-tubules contains voltage-gated Ca^{2+} channels called *dihydropyridine* (DHP) receptors (as they get blocked by the drug DHP) through which they activate the longitudinal sarcoplasmic reticulum.

Longitudinal sarcoplasmic reticulum (L-tubules)

The longitudinal sarcoplasmic reticulum is the name given to the sarcoplasmic tubules of the sarcoplasmic reticulum which run in long axis of the muscle fibre forming a closed tubular system around each myofibril. These L-tubules do not open to the exterior like T-tubules. The longitudinal sarcoplasmic tubules on either side of the T-tubule are dilated to form the so-called terminal cisterns. A, T-tubule with the two *terminal cisterns* lying in close proximity (contiguity) constitute a triad which is found at the junction of A and I bands. Thus there are two triads in each sarcomere. The longitudinal tubules store a large quantity of calcium ions.

PROCESS OF MUSCLE EXCITABILITY AND CONTRACTILITY

As we know, the muscle is an excitable tissue, i.e. when stimulated an action potential is produced (*electrical phenomenon*). The skeletal muscle responds to stimulus by contracting (mechanical phenomenon). The events which link the electrical phenomenon with the mechanical phenomenon is called *excitation–contraction coupling phenomenon.* These three phenomena which mark the excitability and contractility of the muscle when stimulated by the nerve innervating it are discussed.

PROCESS OF MUSCLE EXCITATION

As discussed in the transmission across neuromuscular junction (page 63), when end plate potential (EPP) reaches a threshold level, produces an *action potential* which propagates over muscle fibre surface and into the muscle fibre along the transverse tubules.

Essential features of electrical phenomena

Essential features of electrical phenomena which occur in the muscle fibre (resting membrane potential and action potential) are similar to those occurring in a nerve fibre. However, there are some quantitative differences between the electrical phenomenon occurring in a skeletal muscle fibre and a nerve fibre which are summarized in Table 2.3-1.

PROCESS OF EXCITATION–CONTRACTION COUPLING

The sequence of events by which an action potential in the plasma membrane of a muscle fibre leads to cross-bridge activity (excitation–contraction coupling) is as:

- When the action potential reaches the tip of T-tubule, it activates the voltage-gated channels called DHP receptors which are located in the T-tubule membrane (Fig. 2.3-8).

Table 2.3-1 Essential features of electrical phenomena in the skeletal muscle fibre and the nerve fibre

Features	Skeletal muscle fibre	Nerve fibre
1. Resting membrane potential	–90 mV	–70 mV
2. Initial excitation threshold level	30–40 mV	15 mV
3. Magnitude of action potential	120–130 mV	100–105 mV
4. Duration of spike potential	2–4 ms	0.4–2 ms
5. Absolute refractory period	1–3 ms	0.4–2 ms
6. Maximum number of impulses conducted	Less (100–200/s)	More (1000/s)
7. The excitability	Less (Chronaxie is longer)	More (Chronaxie is shorter)
8. Conduction velocity of action potentials	Low (3–5 m/s)	Variable, directly proportional to its diameter. In a myelinated nerve fibre it is up to 120 m/s
9. Equilibrium potential for different ions		
Na^+	+65 mV	+60 mV
K^+	–95 mV	–90 mV
H^+	–32 mV	–25 mV
Cl^-	–90 mV	–70 mV
HCO_3^-	–32 mV	–25 mV

- Activated DHP receptors in turn trigger the opening of Ca^{2+} release channels located on the terminal cisterns, the so-called *ryanodine receptor* (RYR). This is possible because the lateral cisterns are located very close to the tips of T-tubules and the protein channels of the cisterns (DHP and RYR) are mechanically interlocked. Thus, in short, when the DHP is activated by the depolarization of T-tubules, it undergoes conformational changes which result in the opening of RYR (being actually pulled open).
- Due to opening of calcium release channels (RYR), calcium ions diffuse into the cytoplasm. The concentration of Ca^{2+} in the intracellular fluid is increased by some 2000 times, i.e. from 10^{-7} moles/L to 2×10^{4} moles/L.
- The Ca^{2+} ions get attached to troponin-C and start a chain of events (discussed below) which produce contraction. Hence, the calcium ions are said to form the basis of excitation–contraction coupling.

PROCESS OF MUSCLE CONTRACTION

Molecular basis of muscle contraction

The process of muscle contraction is initiated by the calcium ions as discussed above. A. F. Huxley and H. E. Huxley in 1954 put forward the *sliding theory* or *ratchet theory* to explain how the actin filaments slide over myosin filaments forming the actin–myosin complex during muscular contraction.

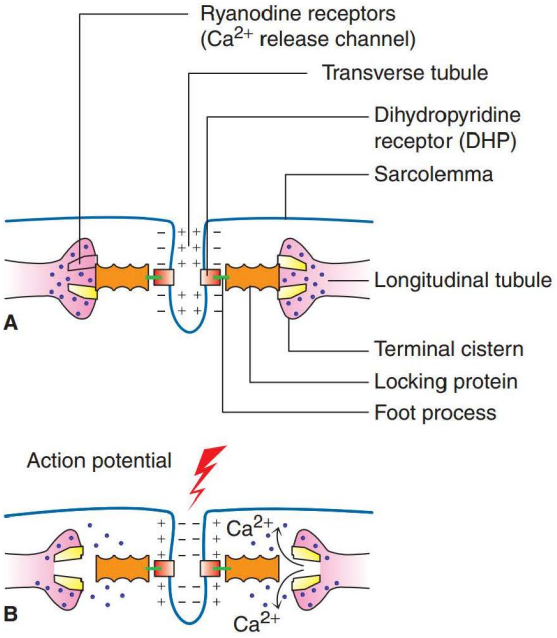


Fig. 2.3-8 Mechanism of release of calcium ions from terminal cistern of longitudinal tubules: A, in resting state Ca^{2+} release channels (RYR) remain closed due to mechanical interlocking between DHP and RYR and B, during activation state (depolarization of T-tubule). Conformational change in DHP results in opening of RYR and release of Ca^{2-}.

This theory explains that the sliding of filaments is brought about by a repeated cycle of formation of the *cross-bridges* between the head of myosin and actin molecules.

Steps of cross-bridge cycling

1. Initiation of cross-bridge cycling. During resting stage, troponin-I is lightly bound to actin and the tropomyosin molecules are located in the groove between the strands of actin filaments in such a way that they block the myosin binding sites on actin. Thus, during resting stage, no actin–myosin cross-bridges are formed. Thus, the troponin–tropomyosin complex so-called relaxing proteins which inhibit the interaction between actin and myosin (Fig. 2.3-9A). When activation takes place, the Ca^{2+} ions released into the cytosol from the terminal cisterns of the sarcoplasmic reticulum get attached to troponin-C subunit of the protein troponin. It results in a conformation change which causes the tropomyosin molecule to move laterally, uncovering the binding sites on the actin molecules for head of the myosin molecules. Seven myosin binding sites on the actin filament are uncovered for each molecule of troponin that binds a Ca^{2+} ion. Thus

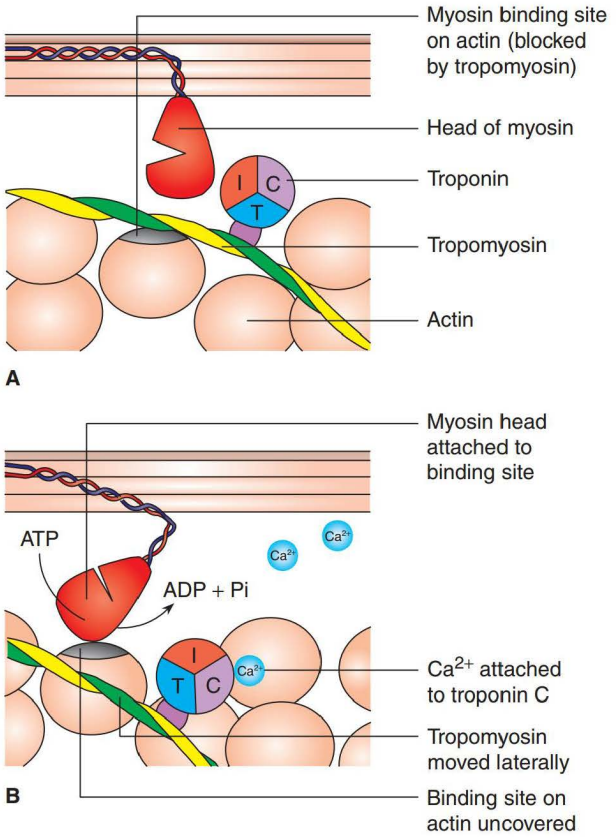


Fig. 2.3-9 Initiation of cross-bridge cycling: A, resting state (myosin binding site on actin is covered by the troponin–tropomyosin complex) and B, on activation Ca^{2+} binds to troponin C subunit which results in conformational change and lateral displacement of tropomyosin causing uncovering of binding site for myosin (head of myosin) on actin (initiation of cross-bridge cycle).

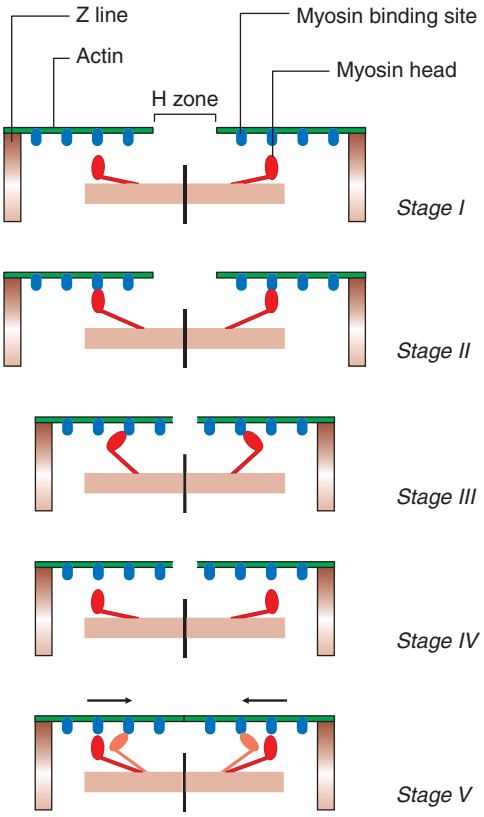


Fig. 2.3-10 Stages of cross-bridge cycling.

the cross-bridge cycle is switched on (initiated) by the lateral movement of the tropomyosin (Fig. 2.3-9B).

2. Formation of actin–myosin complex (i.e. attachment of myosin head to active site of actin filament). Then the head of myosin molecule binds with adenosine triphosphate (ATP). The ATPase activity of myosin head immediately causes breaking of ATP to adenosine diphosphate (ADP) and Pi cleavage products which remains bound to the myosin head. The head of myosin therefore becomes energized. The activated myosin head extends perpendicularly (at 90° conformation) towards the actin filament and gets attached to actin filament (Stages I and II) (Fig. 2.3-10).

3. The power stroke. Formation of the *actin–myosin–ADP Pi* complex triggers simultaneously the following two events:

- Release of the Pi and ADP from the complex and
- A conformational change in the myosin head causing it to flex towards the arm of the cross-bridge. The flexion of the myosin head from the high energy 90° conformation to low-energy 45° conformation generates mechanical force (the power stroke) (Fig. 2.3-10, Stage III).

4. Detachment of myosin head of a cross-bridge from the active site of an actin filament. The release of ADP and Pi allows a fresh ATP molecule to bind to the myosin head. The *myosin–ATP complex* has a low affinity for actin, and therefore, it results in the dissociation of myosin head from the actin filament (Fig. 2.3-10) (Stage IV).

5. Reactivation of myosin head. The freshly bound ATP molecule splits again and myosin head is reactivated for the next cycle to begin. The energized head extends perpendicularly towards the actin filament and gets attached to the new active site for repeating the cycle.

Thus with each cross-bridge cycle, there is movement of the actin filament towards the centre of myosin to a small degree. Repeated cross-bridge cycling causes the movement of actin filaments of either side towards the centre of myosin filament of the sarcomere leading to muscle contraction (Fig. 2.3-10) (Stage V).

IMPORTANT NOTE

During each cross-bridge cycle muscle shortens by 1% and maximum shortening is up to 30% of the total length of muscle.

Steps in muscle relaxation

Within a few milliseconds of the action potential, the calcium pump transports Ca^{2+} ions present in the sarcoplasm during contraction, back into the longitudinal portion of the sarcoplasmic reticulum, from where the Ca^{2+} ions are discharged into the terminal cistern for storage. Removal of calcium from troponin restores blocking action of the troponin–tropomyosin complex. Myosin cross-bridge cycle closes and muscle relaxes.

Functions of ATP in skeletal muscle contraction and relaxation:

1. *Hydrolysis of ATP* by myosin energizes the cross-bridges providing the energy for force generation.
2. *Binding of ATP to myosin* causes dissociation of cross-bridges allowing the bridges to repeat their cycle of activity.
3. *Hydrolysis of ATP by Ca^{2+} ATPase* in the sarcoplasmic reticulum provides the energy for active transport of calcium back into the cisternae, lowering cytoplasmic calcium, ending the contraction and allowing muscle fibre to relax.

APPLIED ASPECTS

Rigor mortis refers to shortening and rigidity of all the body muscles which occurs some hours after death. The rigidity occurs because of fixation of cross-bridges of myosin head to actin filaments due to loss of all the ATPs (which is normally required for detachment of cross-bridges of myosin heads from the actin filaments causing relaxation). The rigidity disappears after some hours due to destruction of the muscle proteins by enzymes released from the cellular lysosomes. The appearance and disappearance of rigor mortis is used by the forensic experts in fixing the time of death.

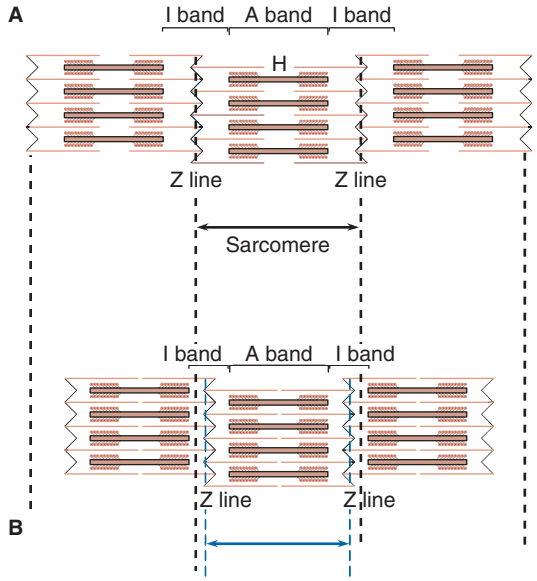


Fig. 2.3-11 Changes produced in a sarcomere by sliding of thin filament (actin) over thick filament (myosin) during muscle contraction: A, relaxed state and B, contracted state.

Changes produced by sliding of thin filaments over thick filaments during muscle contraction

Figure 2.3-11 shows the following changes produced by sliding of thin filaments over thick filaments during muscle contraction.

1. The width of A band remains constant
2. H zone disappears
3. I-band width decreases
4. The Z lines move closer
5. The sarcomere shortens
6. The actin filaments from the opposite end of the sarcomere approach each other and when the muscle shortening is marked, these filaments apparently overlap.

SEQUENCE OF EVENTS DURING MUSCLE CONTRACTION AND RELAXATION WHEN STIMULATED BY A NERVE

The events which occur during contraction and relaxation in a skeletal muscle, when excited by a nerve are summarized sequence wise:

Nerve excitation

- Stimulation of motor neuron

↓

- Initiation of action potential in motor neuron's axon

↓

Nerve conduction

- Propagation of action potential in the motor nerve

↓

- Impulse reaching at nerve ending (at synaptic button)

↓

Neuromuscular transmission

- Increased permeability of presynaptic membrane to Ca^{2+} ions

↓

- Inflow of Ca^{2+} ions from ECF into the nerve terminals

↓

- Release of ACh from the microvesicles present at the nerve terminal

↓

- Diffusion of ACh into the synaptic cleft

↓

- Binding of ACh to receptors on the motor end plate (postsynaptic membrane)

↓

- Opening of ACh-gated channels in the motor end plate membrane

↓

- Entry of mainly Na^{+} ions and to a lesser extent Ca^{2+} ions through these channels into the muscle fibre

↓

- Development of EPP

Muscle excitation

- Local EPP when reaches a threshold magnitude, voltage-gated Na^{+} channels are opened up at the site

↓

- Generation of action potential (AP) in the muscle fibre by the end plate depolarization

↓

- Propagation of AP in muscle fibre along the surface and into the fibre along the T-tubules

↓

Excitation–contraction coupling

- Release of Ca^{2+} ions from terminal cistern

↓

- Diffusion of Ca^{2+} ions into the sarcoplasm

↓

- Binding of Ca^{2+} ions to troponin C

↓

Muscle contraction (molecular theory)

- Uncovering of binding sites for myosin on actin

↓

- Cross-bridge formation between myosin head and actin

↓

- Angular movement of cross-bridges (Power-stroke)

↓

- Sliding of thin filaments over thick filaments

↓

- Initiation of muscle contraction

↓

Muscle relaxation

- Active transport of Ca^{2+} into sarcoplasmic reticulum
  ↓
- Decreased concentration of Ca^{2+} in sarcoplasm
  ↓
- Removal of Ca^{2+} ions from troponin-C
  ↓
- Cessation of cross-bridge cycling
  ↓
- Relaxation of muscle fibre

CHARACTERISTICS OF MUSCLE CONTRACTILITY

CONTRACTILITY

To understand contractile response and its characteristics, it is essential to have knowledge about the following elementary aspects in relation to skeletal muscle:

- Contractile and elastic components of a muscle,
- Concept about muscle length,
- Motor unit and
- Contractile response.

CONTRACTILE AND ELASTIC COMPONENTS OF A MUSCLE

To understand certain facts associated with muscle contraction such as shortening, contraction without shortening and effect of passive stretch, a *three-component model* has been proposed. According to this model, the skeletal muscle as a whole consists of three components (Fig. 2.3-12):

- Contractile component,
- Series elastic component and
- Parallel elastic component.

1. Contractile component

The contractile component (CC) represents the thick (myosin) and thin (actin) filaments present in the myofibrils. It is considered to be viscous in nature, i.e. it offers no resistance to stretch and is unable to return to its original length after it has shortened (Fig. 2.3-12A). The CC comprises 60% (3/5th) of the total muscle proteins.

2. Series elastic component

The series elastic component (SEC) refers to that elastic tissue of the muscle which is present in series with the CC of the muscle. It consists of the elastic tendon of the muscle. In resting condition, the SEC offers resistance to passive stretch and explains how muscle is able to contract even when its external length does not change, i.e. isometric contraction

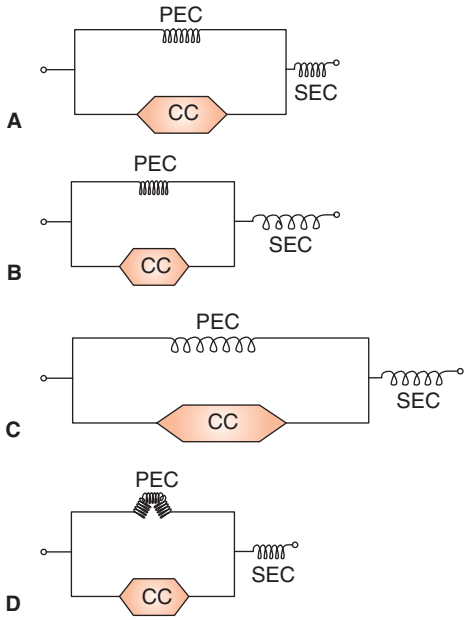


Fig. 2.3-12 Three component model of skeletal muscle consisting of contractile component (CC), series elastic component (SEC) and parallel elastic component (PEC): A, when muscle is at normal length; B, during isometric contraction; C, when muscle is passively stretched and D, in isotonic contraction.

(Fig. 2.3-12B). It also explains how the muscle regains its original length after contracting isometrically.

3. Parallel elastic component

The parallel elastic component (PEC) refers to the elastic tissue of the muscle which is attached parallel to the CC. The PEC is represented by the structural elastic tissue of the muscle such as connective tissue sheaths of the muscle, sarcolemma and filaments. Presence of this component explains why the muscle regains its original length after it is passively stretched (Figs 2.3-12C and D). In isotonic contraction this component gets folded up. It also offers some resistance to passive stretch. The SEC and PEC combinedly form 40% (2/5th) of total muscle proteins.

CONCEPTS ABOUT MUSCLE LENGTH

Following concepts about muscle length will be useful in understanding certain characteristics about muscle contraction:

Optimum length refers to that length of the muscle at which it will develop *maximum active tension*.

Resting length of a muscle represents the length of the muscle during relaxed state under natural conditions in the body. The resting length of many muscles in the body is optimum length.

Equilibrium length refers to the length of a relaxed muscle cut free from its bony attachments.

Initial length is the length of the muscle before it contracts.

MOTOR UNIT

Motor unit is the functional unit of muscle contraction in the intact body. It consists of the single motor neuron cell body, its axon fibres and the muscle fibres innervated by it (Fig. 2.3-13). The cell bodies of the motor neurons (α motor neuron) supplying the skeletal muscle fibres lie in the ventral horn of the spinal cord or the motor cranial nerve nuclei.

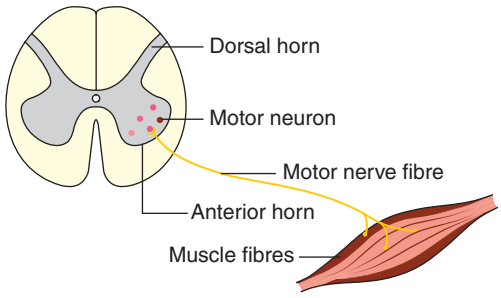


Fig. 2.3-13 Structure of a motor unit.

Type of motor units. Each motor neuron innervates only one type of muscle fibre. In other words, in a single motor unit the muscle fibres supplied by it are of the same type. Therefore, depending on the type of muscle fibres, the motor units are also of two types:

- *Type I* (Red or slow) motor units and
- *Type II* (White or fast) motor units. The characteristic features of each type of motor units are given in Table 2.3-2.

CONTRACTILE RESPONSE

Contractile response is the characteristic feature of a skeletal muscle. When stimulated, an action potential is developed in the muscle fibres which is followed by the muscle

Table 2.3-2 Characteristic features of type I and type II motor units

Characteristics	Motor unit	
	Type I	**Type II**
1. Muscle fibre type	The muscle fibres of type I motor units are: slow, red, and involved in tonic activity.	The muscle fibres of type II motor unit are: fast, white and involved in phasic activity.
2. Motor unit innervation ratio	High (120–160 muscle fibres/axon), e.g. postural muscles.	Low (6 muscle fibres/axon) e.g. extraocular muscles.
3. Metabolism	Aerobic, low glycolytic and high oxidative capacity.	Anaerobic, high glycolytic and low oxidative capacity.
• Mitochondria number	High	Low
• Glycogen contents	Low	High
• Capillary density	High	Low
• Blood supply	High	Normal
• Myoglobin content	High	Low
• Enzymes:	High	Low
NADH dehydrogenase	Low	High
Phosphorylase activity	Low	High
Myosin ATPase activity	High	Low
4. Axon diameter	Small	Large
5. Axon conduction velocity	Slow	Fast
6. Twitch duration of the muscle	Long	Brief
7. Tetanic tension	Small	Large
8. Type of movements	These are adapted for tonic contraction, i.e. for posture maintenance and are first to be recruited during muscle contraction.	These are adapted for phasic contractions, e.g. fine and skilled movements and these remain inactive during contraction and recruited only when brief and powerful contraction is required.
9. Fatiguability	Fatigue resistant	Easily fatiguable
10. Further types	–	Type II motor units are further of four types IIa, IIb, IIc and IIm. IIa: are fast, fatigue resistant and glycolytic. IIb: are fast, fatiguable and glycolytic. IIc: contain muscle fibres found in fetal stage. IIm: are superfast, having unique myosin structure and present mainly in the jaw muscles.

contraction. The muscle contraction is manifested by either shortening (isotonic contraction) or development of tension (*isometric contraction*) or both. The contractile response can be studied in an isolated *nerve–muscle preparation.* In experimental studies, frog's gastrocnemius-sciatic nerve preparation is used to demonstrate the different characteristics of the contractile response. The contractile response of a muscle to a single stimulus through its nerve can be recorded using a suitable lever system on kymograph or physiograph.

A typical contractile response consists of a brief contraction followed by relaxation and is referred to as *single muscle twitch.* The contractile response of a skeletal muscle can be discussed under following headings:

- Isometric versus isotonic contraction,
- Single muscle twitch and
- Factors affecting force of contraction.

Isometric versus isotonic contraction

Isometric contraction

As the name indicates (iso = same, metric = measure, i.e. length), in this type of contraction, the length of muscle remains same but tension is developed in the muscle. Thus there is no movement of the object. Since work done is the product of *force* × *distance,* therefore, in isometric contraction *no external work is done.*

How the muscle length remains same in isometric contraction is depicted in Fig. 2.3-12B. As shown in this figure the shortening produced by the CC of the muscle is compensated by the stretching of the SEC.

Examples of isometric contraction of muscles

- Contraction of muscles which help in maintaining posture against gravity and
- Contraction of arm muscles when trying to push a wall.

Isotonic contraction

As the name indicates (iso = same, and tonic = tone or tension), in this type of contraction, the tension in the muscle remains same whereas its length decreases. Since the muscle length is shortened, so the *external work is done* in isotonic contraction. As shown in Fig. 2.3-12D, in isotonic contraction, the CC and PEC are shortened, but the SEC does not stretch further, producing a visible shortening of the muscle.

Examples of isotonic muscle contraction

- Contraction of leg muscles during walking and running,
- Contractions of muscles while lifting a weight and
- Contraction of muscles during flexion of arm.

Single muscle twitch

As mentioned earlier, the single muscle twitch or also known as the simple muscle twitch refers to the typical contractile response of a skeletal muscle to the single stimulus.

Phases of single muscle twitch

A single muscle twitch recorded under isotonic conditions from a frog's gastrocnemius-sciatic preparation. It shows (Fig. 2.3-14):

- *Point of stimulation (PS)* denotes the time when the stimulus is applied,

The total duration of the muscle twitch is 0.1 s and it shows three phases: latent period, contraction phase and relaxation phase.

1. Latent period (LP). As shown in Fig. 2.3-14, the contraction occurs after a brief gap of stimulation. This time interval between the PS and the point of start of contraction (PC) is called the latent period.

Causes of latent period

The latent period includes:

- Time taken by the impulse to travel from PS on the nerve to the neuromuscular junction,
- Time taken by the impulse for neuromuscular transmission,
- Time taken by the excitation–contraction coupling phenomenon,
- Time taken by the chemical events to cause muscle contraction (sliding phenomenon),
- Time taken for the development of tension in the muscle, and
- Time taken by the inertia of recording lever.

2. Contraction phase. Contraction phase extends from the PC to the point of maximum contraction and is recorded as upward movement of the lever. During this phase, the muscle shortens by about 20% of its resting length. The magnitude of contraction is affected by many factors (see factors affecting force of contraction at page 77).

3. Relaxation phase. The contraction phase is followed by the relaxation phase during which the muscle is stretched back to its original length. It is recorded as downward

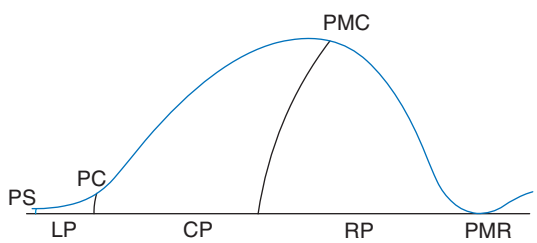


Fig. 2.3-14 Typical single muscle curve recorded from frog's gastrocnemius-sciatic nerve preparation. (PS = point of stimulation; PC = point of contraction; PMC = point of maximum contraction; PMR = point of maximum relaxation; CP = contraction period; RP = relaxation period.)

movement of the recording lever system. In general, the relaxation phase is longer than the contraction phase.

Duration of twitch

The total duration of twitch (*contraction time*) varies with the *type of muscle fibres:*

- *Fast (white) muscle fibres* (e.g. extraocular muscles and muscle of hands for fine movements) have shorter contraction time (about 0.025 s).
- *Slow (red) muscle fibres* (e.g. back muscles) have longer contraction time (0.1 s).

The total duration of twitch also varies from species to species. In human skeletal muscle the duration of twitch is 30–50 ms in comparison to 100 ms in frog's (amphibian's) muscle.

Factors affecting contractile response

The factors which can affect the contractile response (force of contraction) of a skeletal muscle are:

- Strength of stimulus,
- Frequency of stimulus,
- Load on the muscle (pre-load and after-load),
- Initial length of muscle and
- Temperature.

1. Strength of stimulus

A single muscle fibre obeys the all or none law, i.e.

- a subthreshold stimulus evokes no response and
- with threshold, maximal and supramaximal stimuli the contractile response remains constant.

The whole muscle, however, when stimulated with different intensity stimuli the response obtained (force of contraction) is graded one.

In the intact body the whole muscle gets stimulus from the activity of anterior horn cells through their axons to the muscle fibres supplied by that particular neuron, i.e. by *recruitment of motor units.* With minimum activity, only a few motor units are recruited for activity. With increasing activity, more and more motor units (from the motor neuron pool of a muscle) are recruited into activity. This phenomenon is called multiple *motor unit summation.*

2. Frequency of stimulus

The effect of repeated stimuli on the contractile response of a skeletal muscle depends upon the number of stimuli (frequency).

That is the response obtained will depend upon where the next stimulus falls:

- after the first twitch, or
- in relaxation phase of first twitch, or
- in contraction phase or to second half of latent period of first twitch.

Based on the above facts following types of responses are observed due to multiple stimuli:

(i) ***Discrete responses.*** When the frequency of stimulation is such that the next successive stimulus falls after completion of relaxation phase of the previous twitch then the succeeding contraction obtained, with brief intervals between them, are complete individual twitches (with contraction and relaxation phases). Such a response is called *discrete response.* Further, each successive twitch has increased force of contraction (due to beneficial effect of previous twitch) till a maximal beneficial effect is achieved. This phenomenon is called the staircase effect or treppe (a German word for staircase).

(ii) ***Incomplete tetanus or clonus.*** When the frequency of multiple stimuli is such that the next successive stimulus falls on the relaxation phase of the previous twitch then the succeeding contraction obtained will be superposed over the previous twitch due to incomplete summation of waves. This state is referred to as state of subtetanus or incomplete tetanus or clonus.

(iii) ***Complete tetanus.*** When the frequency of multiple stimuli is such that the next successive stimulus falls from second half of latent period to contraction phase of the previous twitch, (i.e. before relaxation begins) than due to complete summation effect, the muscle will remain in a state of sustained, smooth and forceful contraction called tetanus or tetanic contraction. During complete tetanus, the tension developed in the muscle is four times greater than that developed during the individual muscle twitch.

Ionic basis of tetanus. The Ca^{2+} ions released in the sarcoplasm during single twitch are removed quickly and relaxation occurs. When the muscle is stimulated in rapid succession, there occurs a progressive accumulation of Ca^{2+} ions in the sarcoplasm. The longer stay of Ca^{2+} ions in the sarcoplasm increase the duration of active state (due to continuous recycling of myosin heads). This increases the amount of stretch on the SEC and the tension developed rises to tetanic levels.

3. Effect of load

Load is the force exerted by the weight of an object on the muscle. The force exerted by the contracting muscle on the object is known as *muscle tension*. Thus, muscle load and muscle tension are two apposing forces. The load acting on the muscle is of two types: free-load (or pre-load) and after-load.

Effect of free-load. A load which starts acting on a muscle before it starts to contract is called free-load (or pre-load). Example of a free-load on the muscles in an intact body is filling water from a tap by holding a bucket in the hand.

The free-load increases the force of contraction and work efficiency of the muscle. The free-load stretches the muscle passively producing a *passive tension* across the muscle. This passive tension increases the force of muscle contraction in two ways:

- by increasing the initial length of the muscle to its resting length at which maximum force is generated and
- by adding an elastic recoil force to the muscle during its contraction.

Effect of initial length on force of contraction. According to Starling's law, the force of contraction is a function of the initial length of muscle fibres. Therefore, up to physiological limits, the greater the initial length, greater is the force of contraction.

Length–tension relationship. When a muscle is removed from the body, it shortens because muscles in the body are in a state of slight stretch. The length of the muscle when it is detached from its bony attachments is called *equilibrium length.*

The *length–tension relationship graph* can be plotted by measuring isometric tension at different muscle lengths in an isolated muscle preparation. For this, the isolated muscle is attached to an isometric lever, which does not allow the shortening of muscle to occur. The tension developed is recorded through the force transducer (Fig. 2.3-15). The length of the muscle is varied by changing the distance between its two attachments and the recording is made as:

- At each length the *passive tension* is measured. The passive tension is due to stretching of parallel and series elastic components of the muscle (PEC and SEC).
- The muscle is then electrically stimulated at each length and the tension developed is recorded. The *total tension* so recorded includes both the passive tension and the active tension developed due to contraction of the contractile component (CC) of the muscle.
- The *active tension* is thus denoted by the difference between the total tension and passive tension at any length.

Length–tension relationship graph is then plotted with increase in muscle length (in cm) along horizontal axis and tension (in kg) along the vertical axis (Fig. 2.3-16). The following inferences can be drawn from this graph:

- Passive tension due to stretching of elastic components (PEC and SEC) of the muscle increases with the passively increased muscle length.
- At each length the *passive tension* is measured.
- Active tension de*veloped is maxim*um at the optimum length of the muscle (position B in Fig. 2.3-16) which is equivalent to the resting length in intact body.
- Total tension is contributed by the active tension and the passive tension at different muscle lengths (Fig. 2.3-16).

Molecular basis of length–tension relationship. During isometric contraction, the tension developed in the muscle is proportional to the cross-bridges formed between actin and myosin filaments. The effect of muscle length on the tension produced during contraction can be explained by the sliding filament theory of muscle contraction as:

- *At optimum length*, there is optimum overlapping between the actin and myosin filaments, so maximum cross-bridges are formed between them.
- *At muscle length shorter than optimum* length (at position A in Fig. 2.3-17), the thin filaments overlap each other and thus reduce the cross-bridges between the actin and myosin filaments and so the active tension produced is less.
- *When the muscle is overstretched* (position C in Fig. 2.3-17) the Z lines are pulled apart and the overlapping

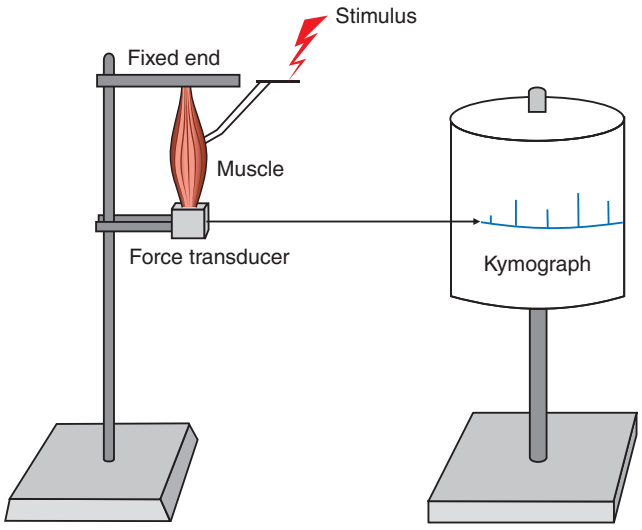


Fig. 2.3-15 Arrangement for recording isometric contraction in an isolated nerve muscle preparation.

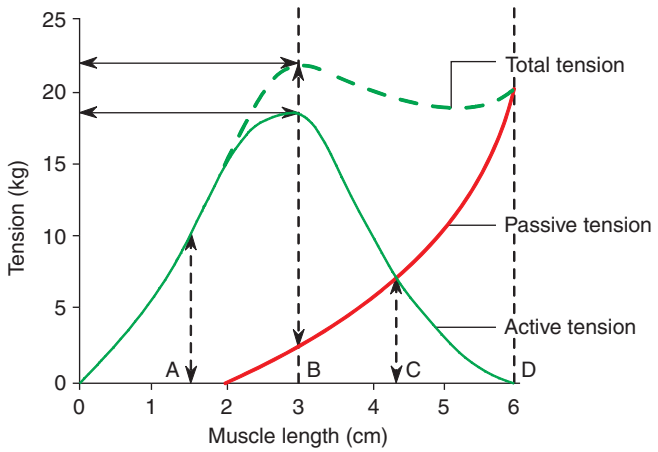


Fig. 2.3-16 Length–tension relationship graph.

between the actin and myosin filaments is markedly reduced and so no active tension is developed at this level.

Effect of after-load. After-load refers to the load which acts on the muscle after the beginning of muscular contraction. Thus, the after-load opposes the force produced by muscle contraction. The work done in an after-loaded muscle is less than that of a free-loaded muscle. An example of after-load in an intact body is lifting any object from the ground.

Force–velocity relationship. The force–velocity curve (Fig. 2.3-18) is plotted by noting the velocity of shortening of muscle with progressively increasing load on the muscle.

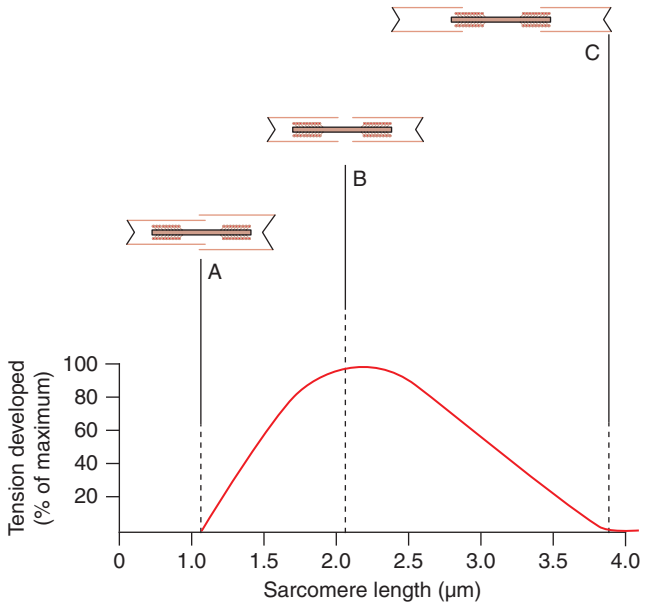


Fig. 2.3-17 Molecular basis of length-tension relationship.

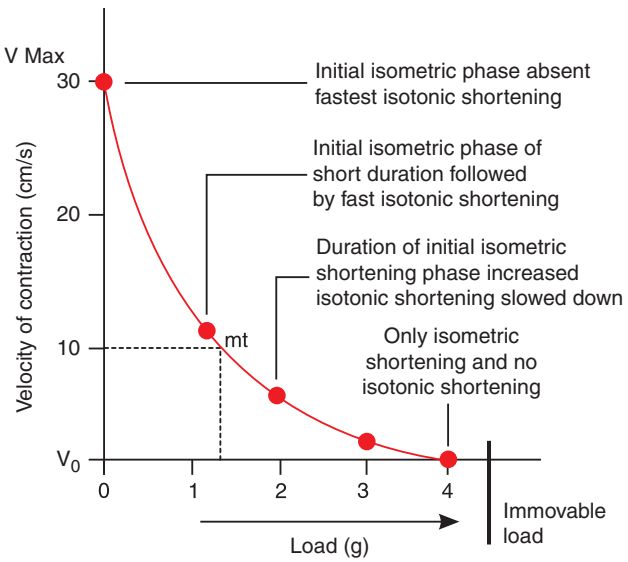


Fig. 2.3-18 Force–velocity curve plotted by recording velocity of the shortening of skeletal muscle with progressively increasing load.

Following inferences can be drawn from the force–velocity curve:

- When load is zero, the muscle contracts rapidly and the velocity of muscle shortening is maximum (V_{max}).
- As the load increases the velocity of shortening decreases. With further increases in the load, a stage comes when the muscle is unable to lift the load. At this point muscle contracts isometrically.
- Between the two extremes of zero load and immovable load, all contractions have variable durations of isometric and isotonic contractions.
- In the force–velocity curve, **mt** is the point of maximum efficiency of the muscle. It lies about 1/3rd of the abscissa and 1/3rd from the ordinate.

4. Effect of temperature

The contractile response is altered due to the effect of temperature. At moderately high temperature (say 40°C) there occurs an increase in amplitude of muscle curve occurs due to increase in isotonic shortening of muscle. This occurs due to decrease in the internal viscoelastic resistance. Velocity of contraction also increases.

At low temperature (say 5–10°C) the reverse changes occur; however, the effects of cold are reversible. So, if the muscle is gradually re-warmed, excitability is regained.

At high temperature (above 50–60°C) there occurs coagulation of the muscle proteins leading to stiffness and shortening of the muscle fibres. This condition is called heat rigor. It is an irreversible phenomenon.

Other types of rigors. Some other types of rigors (other than the heat rigor) are also described here because of the similar changes:

- *Cold rigor.* It occurs following exposure to severe cold. It is a reversible phenomenon.
- *Calcium rigor.* It occurs due to increased calcium content. It is also a reversible phenomenon.
- *Rigor mortis (see page 72).*

SOME CHARACTERISTICS OF THE SKELETAL MUSCLES IN THE INTACT BODY

MUSCLE TONE

Muscle tone is the state of slight contraction with certain degree of vigour and tension. All the skeletal muscles exhibit muscle tone. However, it is more marked in the anti-gravity muscles, viz. extensors of the lower limbs, trunk muscles and muscles of the neck.

Maintenance of muscle tone. Muscle tone is a state of partial tetanus of the muscle maintained by asynchronous discharge

of impulses from gamma motor neurons in the anterior grey horn of the spinal cord concerned with the motor nerve supply of the muscles. The gamma motor neurons in turn are controlled by some higher centres in brain (see page 833).

NATURE OF MUSCLE CONTRACTIONS IN THE BODY

The simple muscle twitch is not a physiological event. Basically, all contractions in the body are tetanic in nature. Weak contractions result from low frequency of firing (5–10/s) of the motor units. The expected jerkiness and the disadvantage of incomplete tetanus are overcome by the asynchronous discharge (out of step firing) of groups of motor units. When one group is firing, the others are silent and vice-versa. Algebraic summation occurs, the individual variations are evened out, and a smooth contraction results. The degree, to which the motor neuron discharge is asynchronous, is related both to the force and duration of contraction.

With increasing firing rates and, of course, with more recruitment of motor units, contractions become stronger until, at and beyond the tetanizing rate, sustained and powerful contractions result.

GRADATION OF FORCE OF MUSCLE CONTRACTION IN THE INTACT BODY

For performing different kinds of work, e.g. picking up a pen from the table, or lifting 10 kg weight, same muscles are involved but on these different situations, muscles can generate different degree of power. This property of skeletal muscles is known as gradation of force of contraction. Gradation of muscle power in muscles is made possible by certain factors which affect the force of contraction. These are:

1. Recruitment of motor units. The force of contraction produced in a muscle depends upon the number of motor units recruited (see page 77).

2. Frequency of nerve impulses. The motor control system in the brain can vary the force of contraction by varying the frequency of nerve impulses stimulating the muscle. As the impulse frequency increases, its effects are summated (wave summation) and the muscle tension increases.

3. Synchronization of impulses. At any one time, the motor units are in different phases of activity, i.e. some are contracting and others are relaxing. Due to algebraic summation, the muscle gives a steady but weak pull. With increasing synchronization of the motor units, the force of contraction increases.

MUSCLE FATIGUE

Failure of a muscle to maintain tension as a result of previous contractile activity is known as muscle fatigue. If the muscle is allowed to rest after the onset of fatigue it recovers its ability to contract.

Fatigue refers to the fatigue of the most of muscles that develops after prolonged general exercise such as marathon running, and competitive football match playing.

Onset and recovery of fatigue depends on:

- Intensity and duration of exercise and
- Type of muscle fibres. Fast glycolytic fibres fatigue early and also recover rapidly from fatigue. Slow oxidative fibres do not fatigue early but they also require longer time of rest (up to 24 h) for complete recovery.

Site, causes and mechanism of fatigue. In human body, the sites of fatigue are in the following order:

- Fatigue of *synapses of central nervous system* due to slight hypoxia occurs first of all. It is particularly of high intensity in short-duration exercises.
- Second site of fatigue is *motor neurons* (anterior grey horn cells) of spinal cord.
- *Motor end plate* in the neuromuscular junction may also be fatigued.
- Changes in muscles also contribute to development of fatigue:
  - In short duration, high-intensity exercises such as weight lifting and short distance running there occurs increased acidity in the muscle cells which accompanies rise in lactic acid (formed due to anaerobic glycolysis). H^+ ion concentration directly inhibits cross-bridge cycles and therefore force generated by them. The second cause is decrease in release of Ca^{2+} ions from the sarcoplasmic reticulum.
  - In long duration, low intensity exercises such as marathon race, depletion of muscle glycogen is an important contributing factor.

Psychological fatigue. Lack of motivation due to failure of cerebral cortex to send excitatory signals to motor neurons causes an individual to stop exercising. The psychological fatigue (feeling of weariness) is different from the *physiological fatigue* of the muscles. The muscles are actually not fatigued (therefore, it is not a true muscle fatigue). Athlete's performance, therefore, depends not only on physical status of appropriate muscle but by psychological fatigue also.

ELECTROMYOGRAPHY AND COMMON DISORDERS OF MUSCLES

ELECTROMYOGRAPHY

Electromyography refers to the technique of recording the total electrical activity of the motor nerve and the muscle

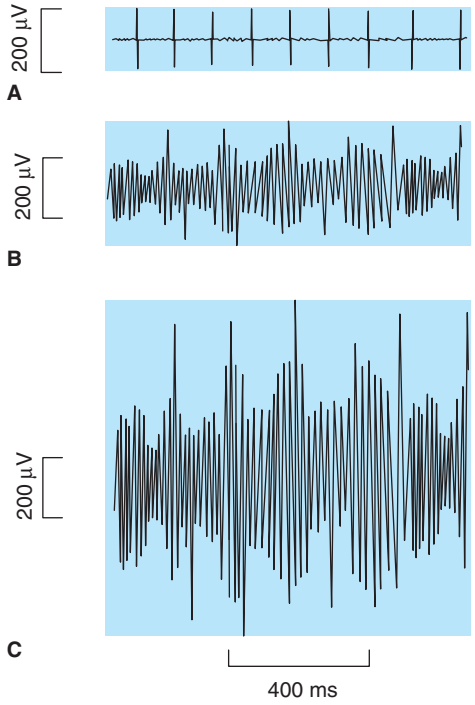


Fig. 2.3-19 Electromyogram during voluntary muscle activity: A, on minimal contraction; B, on moderate contraction (recruitment pattern) and C, on maximum contraction (normal interference pattern).

under study. The machine used to record the said electrical activity is called electromyograph and the record obtained is called the electromyogram (Fig. 2.3-19).

ELECTROMYOGRAM

Components of a normal electromyogram (EMG) along with the common abnormalities which can be detected are given.

Spontaneous activity at rest

Normally, at rest there is complete electrical silence and no spontaneous activity is recorded, except for:

1. The insertion activity. When a needle electrode is inserted into a muscle it evokes a discharge of action potentials (due to mechanical stimulus). These potentials are of short duration and small amplitude. The insertion activity is:

- prolonged in denervated muscle and
- absent when muscle tissue is not viable.

2. End plate noise. After the cessation of insertion activity, no other spontaneous activity is seen but for a monophasic negative potential in the end plate region called as end plate noise.

Voluntary activity during muscle contraction

The potential changes recorded during muscle contraction is called motor unit potential (MUP). The electromyographic records obtained *during different grades* of muscle contraction have following characteristics:

1. ***On minimal voluntary contraction,*** only a single or two smaller motor units in the vicinity of needle electrode give off electrical discharge (Fig. 2.3-19).
2. ***With the progressive increase in the voluntary contraction,*** the firing rate of small units increases until it reaches a certain frequency, when larger units are recruited. Figure 2.3-19 shows the recruitment pattern with moderate force of contraction.
3. ***During maximal contraction*** so many motor units are recruited and thus the so many rhythmically recurring MUPs become superimposed upon one another on the oscilloscope screen that it is impossible to determine their individual characteristics. The resulting appearance of the EMG is designated as normal interference pattern (Fig. 2.3-19C).

Abnormal spontaneous activities

Abnormal spontaneous activities which can be recorded during resting phase are:

1. ***Fasciculation potentials*** resemble MUPs and represent the involuntary contraction of muscle fibres of single motor unit.
2. ***Fibrillation potentials*** are of very short duration and low amplitude.

DISORDERS OF SKELETAL MUSCLES

1. Muscular dystrophy

Muscular dystrophy is a syndrome which occurs due to genetic mutation and is characterized by progressive muscle weakness.

2. Myotonia

Myotonia is a disorder which occurs due to abnormalities of the sodium and chloride channels caused by abnormal genes on chromosomes 7, 17 or 19. It is characterized by an abnormally prolonged muscle relaxation after voluntary contraction.

3. Myasthenia gravis

It is a disorder of neuromuscular junction, characterized by a grave weakness of the muscles (see page 65).

4. Abnormal muscle tone

The normal muscle tone of a muscle may be increased or decreased constituting abnormal muscle tone.

5. Fibrillation and denervation hypersensitivity

The denervation of skeletal muscles in lower motor neuron lesions causes flaccid paralysis of the muscle, fibrillation and denervation hypersensitivity.

- ***Fibrillation*** is characterized by fine, irregular contractions of individual muscle fibres.
- ***Denervation hypersensitivity*** refers to when the muscle becomes highly sensitive to acetylcholine.

When motor nerve to skeletal muscle is cut, then muscle gradually becomes hypersensitive to acetylcholine. In degenerative hypersensitivity, the motor end plate is not only depolarized by acetylcholine but the surrounding area is also depolarized and whole muscle becomes sensitive. This may be because of increase in number of active receptors for acetylcholine (up-regulation of receptors) or decreased uptake of released neurotransmitter.

SOURCE OF ENERGY AND METABOLIC PHENOMENON DURING MUSCLE CONTRACTION

ENERGY SOURCE FOR MUSCLE CONTRACTION

The muscle contraction requires lot of energy. In fact muscle has been labelled as a *machine for converting chemical energy into mechanical work*. The immediate source of energy is ATP and the ultimate source is the intermediary metabolism of carbohydrate and lipids.

Hydrolysis of ATP

The hydrolysis of ATP provides energy for muscle contraction [for details see molecular basis of muscle contraction (page 50)]. ATP stored in the muscle initiates the contractile activity but is consumed after a few twitches. In about 3 seconds, all the ATP stored in the muscle cell is depleted. Thus, there is need for resynthesis of ATP.

Resynthesis of ATP

There are three ways in which muscle fibre can resynthesize ATP from ADP during contractile activity:

1. Phosphorylation of ADP by creatine phosphate. Immediately after the depletion of ATP stores of the muscle, ATP is regenerated using the energy released by the dephosphorylation of creatine phosphate reserves of the muscle fibre.

   Creatine phosphate + ADP ⇔ creatine + ATP.

2. Glycolysis. After depletion of creatine phosphate reserves, the next important source of energy which is used to reconstitute both ATP and phosphocreatine is glycogen (previously stored in the muscle cell) by the process of glycolysis which can sustain muscle contraction for about 1 min.

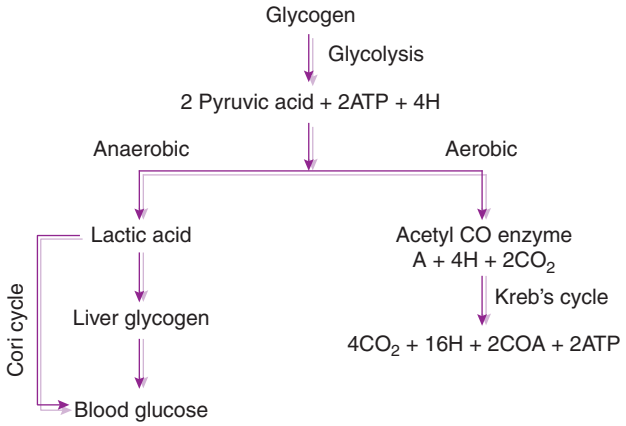


Fig. 2.3-20 Schematic diagram showing breakdown of glycogen stored in a muscle cell.

As shown in Fig. 2.3-20, each molecule of glycogen after glycolysis produces two molecules of pyruvic acid and two molecules of ATP. Further changes in pyruvic acid depend upon the availability of oxygen.

- *In the absence of oxygen,* the pyruvic acid is converted into lactic acid which is released into the blood. From the blood, lactic acid is taken up by the kidney and liver where it is reconverted into glucose and released back into circulation (Cori cycle).
- *If oxygen is available,* the pyruvic acid enters into the *Krebs' cycle.*

A total of 38 ATP molecules are formed during breakdown of each glycogen molecule.

3. Oxidative metabolism. Oxidative metabolism, i.e. combining of oxygen with various cellular foodstuffs to liberate ATP is the final source of energy during muscle contraction. This source contributes more than 95% of all energy used by the muscles for sustained long-term contraction. Foodstuffs used in oxidative metabolism include fats, carbohydrates and proteins.

Oxygen demand, consumption and debt

Oxygen demand increases with the intensity of exercise, i.e. intensity of muscle contraction. It has been reported that in a sprint lasting for ½ min, the oxygen demand is around 20 L/min.

Oxygen consumption or oxygen utilization is the volume of oxygen which has been actually consumed during the exercise. The maximum amount of oxygen that can be consumed by a person while performing severe exercise (irrespective of the demand) is VO_2 max. A world class

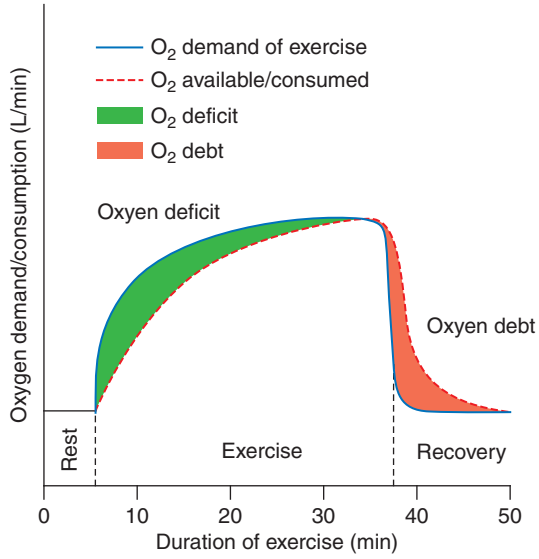


Fig. 2.3-21 Oxygen debt.

sprinter is expected to have a VO_2 max around 75 mL/kg/min (about 4 L/min).

Oxygen debt. During intense exercise, the maximum oxygen consumed is much less than the oxygen demand. So the energy requirement is met with by the *anaerobic pathway.* After the period of exercise, extra O_2 is consumed to remove the excess lactate collected due to anaerobic glucose breakdown, replenish the ATP and phosphoryl creatine store and replace the small amounts of oxygen that have come from the myoglobin. This amount of extra oxygen consumed is called O_2 debt and is proportionate to the extent to which energy demands during exercise exceeded the capacity of aerobic synthesis of energy store, i.e. extent to which *oxygen deficit occurred* during exercise. Oxygen debt can be measured experimentally by determining oxygen consumption after exercise until constant basal consumption is reached and then subtracting the basal oxygen consumption from the total oxygen consumed during this period (Fig. 2.3-21).

APPLIED ASPECTS

To avoid excessive O_2 debt early in the race, the experienced long-distance runners begin the race very slowly to allow the cardiorespiratory system to gear up to the energy demands of muscular activity once a steady state is attained due to cardiorespiratory readjustment, the oxygen supply balances the oxygen requirements of the muscles. This state of oxidative metabolism to provide energy for the muscles can continue for several hours without producing excessive oxygen debt. This prevents too much anaerobic metabolism and accumulation of lactic acid which hamper the efficiency.

Mechanical efficiency of muscle

During contraction the efficiency of muscle is about 25%. Mechanical efficiency is equal to output/input. Therefore, mechanical efficiency is equal to

$$\frac{\text{Work done by the muscle (W)}}{\text{Oxygen consumption } (VO_2)}$$

OR

$$\frac{\text{W (kilopond meter/min/426.7)}}{VO_2 \text{ (L/min)} \times 5}$$

(Since 426.7 kilopond meter/min = 1 kcal, and 1 L/min VO_2 at STPD produces 5 kcal of energy) [STPD means standard temperature (0°C), pressure (760 mm Hg) and dry]

- Therefore, mechanical efficiency in isotonic contraction is approximately 25% of the energy expenditure and rest 75% is degraded as heat and
- During isometric contraction as no external work is done, therefore mechanical efficiency is nil and 100% energy expenditure is disappeared as heat.

CHANGES IN pH DURING MUSCLE CONTRACTION

Changes occurring in the pH and in the reaction of the muscle during contraction are as follows:

- *During resting condition,* the reaction of muscle is alkaline with a pH of 7.3.
- *During onset of muscle contraction,* due to dephosphorylation of ATP to ADP, the pH of muscles becomes acidic.

THERMAL CHANGES DURING MUSCLE CONTRACTION

Thermal changes in the muscle during different phases of contraction are:

1. Resting heat. Resting heat is the heat generated when the muscle is at rest, i.e. not contracting. Resting heat is the external manifestation of the basal metabolic process of the muscle.

2. Initial heat. Initial heat refers to the heat generated in excess of resting heat during muscular contraction. It is made up of following components:

- *Activation heat* refers to initial rapid liberation of energy before the actual contraction of the muscle. It is mostly due to the heat liberated while calcium ions are released from the L-tubules of sarcoplasmic reticulum and the myosin ATPase is activated.
- *Shortening heat* is produced when the muscle contracts isotonically. It is produced due to various structural changes in the muscle fibre like movement of cross-bridges

and myosin heads and breakdown of glycogen. It is absent during the isometric contraction.

- *Maintenance heat* is generated during the isometric contraction, when no actual shortening of the muscle fibre takes place. Its cause is complicated and mostly obscure.

3. Recovery heat. Recovery heat refers to the heat produced in excess of resting heat following muscle contraction. It continues for about 30 min of the cessation of muscle contraction. This heat is generated by the metabolic processes that restore the muscle to its precontraction state. The recovery heat is approximately equal to initial heat (heat produced during contraction).

4. Relaxation heat. This is the extra heat, in addition to the recovery heat, which is produced during relaxation of the isotonically contracted muscle.

Fenn effect: Fenn effect states that the heat produced is directly proportional to the work done. When the work done is more, the expenditure of ATP will also be more Therefore, Fenn effect can be considered to state that more work done causes more expenditure of energy.

Chapter 2.4

Smooth Muscle and Cardiac Muscle

SMOOTH MUSCLE

FUNCTIONAL ANATOMY AND ORGANIZATION

Smooth muscles (nonstriated muscles), as the name indicates, are characterized by absence of the typical cross-striated pattern seen in the skeletal muscles. Because of their spontaneous activity or activity through the autonomic nervous system, they are also called *involuntary muscles.*

The smooth muscle cells are long fusiform in shape and are aggregated to form bundles or *fasciculi.* The fasciculi are aggregated to form layers of variable thickness. Thus, smooth muscles exist either in sheet or bundles of fibres. In each layer the cells are so arranged that thick central part of one cell is opposite the thin tapering ends of adjoining cells (Fig. 2.4-1).

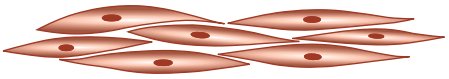


Fig. 2.4-1 Arrangement of smooth muscle fibres.

TYPES OF SMOOTH MUSCLES

Smooth muscles are of two types: single unit and multiunit smooth muscles.

1. Single unit smooth muscles (Fig. 2.4-2)

Single unit smooth muscles are also called *visceral smooth muscles* since they are present in the walls of hollow viscera such as gastrointestinal tract, uterus, ureters, urinary bladder and respiratory tract.

Salient features of single unit smooth muscles are:

- These are arranged in the form of large sheets and has low resistance bridges between individual muscle cells and function in a *syncytial fashion* and that is why they are called single unit muscles (Fig. 2.4-2). The low

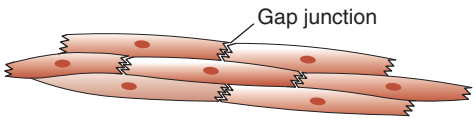


Fig. 2.4-2 Single unit smooth muscle fibre showing gap junctions between two adjacent cells.

resistance intercellular bridges or the so-called gap junctions are in abundance and have high conductance for the ions. Therefore, syncytium contracts as a single unit in many large areas.

- These muscles have their own rhythmic contractility *myogenic tone* that is independent of the nerve supply. The rate of contraction may be determined by the pacemaker regions present within the muscles. The nervous influence only modulates their activity, i.e. the role of the nerves is to increase or decrease the rate of rhythmic contraction.
- Contraction of this kind of smooth muscles is also stimulated by *stretching.* The muscles of smaller blood vessels are mainly of this kind and their contraction in response to stretch is involved in the autoregulation of blood flow.
- In addition to the autonomic nervous system, their contractile activity is also influenced by some non-neural stimuli, e.g. hormones and local tissue factors (such as temperature and pH).

2. Multiunit smooth muscles

Multiunit smooth muscles, as the name indicates, are made up of multiple individual units without interconnecting bridges, i.e. *non-syncytial in* character (Fig. 2.4-1). These are located in most blood vessels, epididymis, vas deferens, iris, ciliary body and piloerector muscles.

Salient features of multiunit smooth muscles are:

- These muscles are made up of multiple individual units of muscle fibres each innervated by a single nerve ending.
- These fibres do not exhibit spontaneous contraction, i.e. *no pacemaker activity.*
- Since the *gap junctions* are not present, the excitation remains localized within the motor unit.
- These muscles do not respond to stretch.

INNERVATION AND NEUROMUSCULAR JUNCTION OF SMOOTH MUSCLES

Nerve supply

Smooth muscles are innervated by the autonomic nerves, both sympathetic as well as parasympathetic. The two have opposite effects. In some organs sympathetic stimulation causes contraction and parasympathetic stimulation causes relaxation of smooth muscles. While in some other organs a reverse action is seen.

Neuromuscular junction

The postganglionic nerve fibres, as approach the smooth muscles branch extensively and come in close contact with large number of smooth muscle fibres (Fig. 2.4-3). The neuronal network so formed has a beaded appearance due to the large enlargements called *varicosities.* These *varicosities* contain the chemical neurotransmitter (acetylcholine or norepinephrine).

In the smooth muscle, the nerve fibres are not ending in *motor end plates* (as seen in skeletal muscles), i.e. the nerve fibres do not make any direct contact with the muscle fibres. Instead, the nerve fibres release its neurotransmitter from each varicosity into the interstitial fluid close to the muscle fibres. The neurotransmitter so released diffuses into a large number of cells and causes activation of all muscle fibres up to where it is forming a syncytium.

Excitatory junctional potential (EJP) or inhibitory junction potential, i.e. either a depolarizing or a hyperpolarizing response may be recorded from a smooth muscle in response to an appropriate nerve stimulus. These potentials summate with repeated stimuli.

STRUCTURE OF SMOOTH MUSCLE FIBRE

Each smooth muscle fibre is a long spindle-shaped cell (myocyte) having a broad central part and tapering ends (Fig. 2.4-4A). The length of smooth muscle fibre is highly

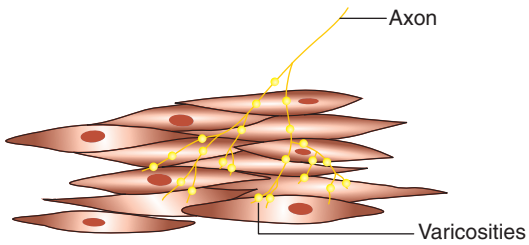


Fig. 2.4-3 The nerve supplying to smooth muscle showing varicosities (beaded appearance).

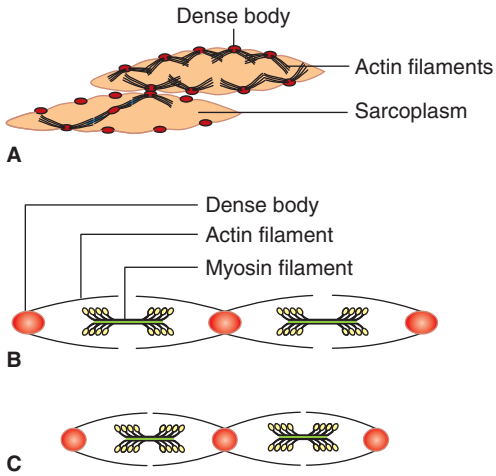


Fig. 2.4-4 Structure of smooth muscle: A, arrangement of thin (actin filaments attached to dense bodies) and thick (myosin) filaments; B, position of dense bodies in relaxed state and C, the dense bodies are drawn closer to each other in contracted state.

variable (15–500 µm) depending upon the organ in which they are present. For example,

- *digestive tract* fibres are 30–40 µm long and 5 µm in diameter,
- fibres in *blood vessels* are 15–20 µm long and 2–3 µm in diameter and
- fibres in *uterus* are 300 µm long and 10 µm in diameter.

Salient features of structure of a smooth muscle fibre

Plasma membrane which binds the smooth muscle is surrounded by an external lamina. Adjacent smooth muscle cells communicate through gap junctions.

Nucleus is oval or elongated and lies in the central part of the cell.

Sarcoplasm, in addition to a single nucleus contains other cell organelles like mitochondria (source of energy), a Golgi complex, some granular endoplasmic reticulum and free ribosomes. Apart from these, sarcoplasm also contains myofibrils and intermediate filaments.

Myofibrils are made of contractile proteins, the myosin and actin filaments. The *longitudinal striations* seen on light microscopy are due to these myofibrils. The salient differences from the skeletal muscle are:

- Smooth muscles contain relatively less *thick filaments* and more *thin filaments.*
- *Z line* is not well defined in smooth muscles.
- *Myosin* is chemically different from that seen in skeletal muscles. It binds to actin only if its light chain is phosphorylated. Thus, phosphorylation of myosin is necessary for the contraction of smooth muscles.
- *Thin actin filaments* are also different from those in skeletal muscle due to absence of the troponin protein molecules.
- *Dense bodies* (Fig. 2.4-4B) attached to the cell membrane and scattered all over the body of the fibres are seen under electron microscope. The actin filaments are attached to these dense bodies. In between the actin filaments, the thick myosin filaments are situated. There are cross-bridges between actin and myosin, which help in sliding mechanism of muscle contraction. When the muscle contracts the points on the cell membrane, where dense bodies are attached, are drawn closer to each other. This converts an oblongated smooth muscle in one that is oval (Fig. 2.4-4C).

PROCESS OF EXCITABILITY AND CONTRACTILITY

- Process of muscle excitation
- Process of excitation–contraction coupling
- Process of muscle contraction.

PROCESS OF MUSCLE EXCITATION

Process of muscle excitation basically includes the electrical activity in smooth muscles which differs in a multiunit smooth muscle than that in a single unit muscle; and so is discussed separately.

Electrical activity in single unit (visceral) smooth muscles

Resting membrane potential

The resting membrane potential in a visceral smooth muscle ranges between –50 and –75 mV. Sometimes it may reach as low as –25 mV. Thus, the peculiarity of the resting membrane potential is its *unstability,* i.e. there is no true resting value rather it keeps on oscillating between –55 and –35 mV (Fig. 2.4-5). These oscillations in the resting membrane potential occur due to superimposition by the *pacemaker potentials* which in turn occur due to rhythmic changes in either Ca^{2+} channel permeability and/or the activity of Na^{+}–K^{+} pump.

Action potential

When depolarization reaches threshold potential, an action potential is generated which is transmitted to the adjacent muscle cells through the gap junction. Three types of action potentials are known to occur in the visceral smooth muscle fibres viz. spike potential, spike potential superimposed over pacemaker potential and action potential with plateau.

1. Spike potential. A typical spike potential similar to that seen in the skeletal muscles is also observed in most, if not all, single unit smooth muscles (Fig. 2.4-6A).

2. Spike potential superimposed over slow wave potentials. The slow wave rhythms, also called pacemaker waves, are seen in many visceral smooth muscles such as muscles of gut (Fig. 2.4-6B) and cause rhythmic contractions of the self-excitatory smooth muscles. When the potential of slow waves rises above the level of about –35 mV (the approximate threshold for eliciting action potential in most

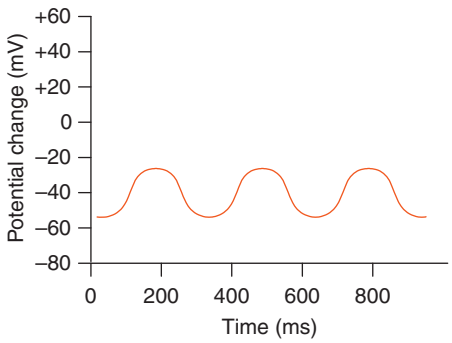


Fig. 2.4-5 Resting membrane potential (fluctuating type) in visceral smooth muscle.

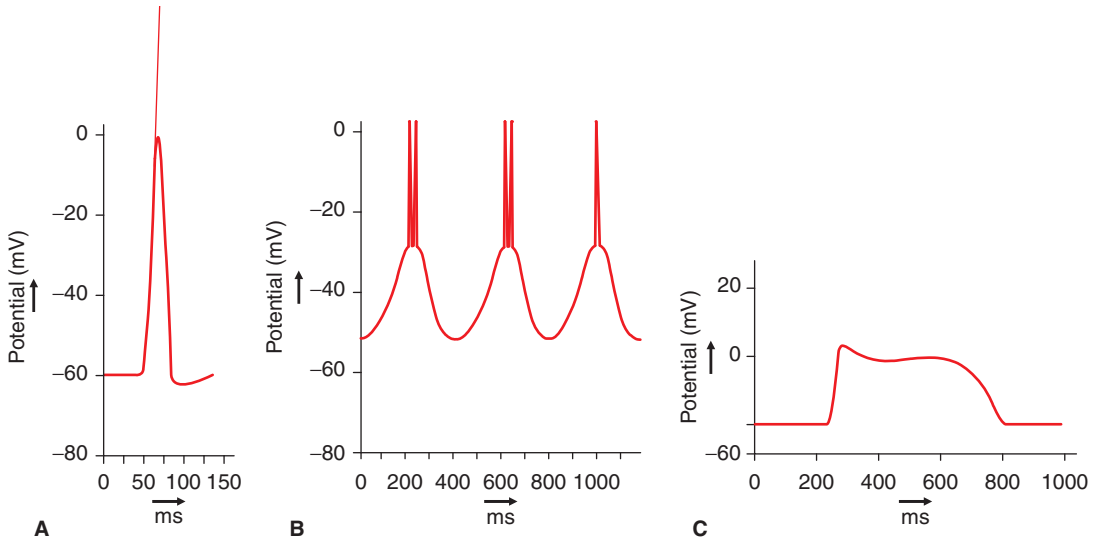


Fig. 2.4-6 Three types of action potentials recorded from smooth muscles: A, spike potential; B, spike potential superimposed over slow wave potentials and C, action potential with plateau.

visceral smooth muscles), an action potential develops and spread over the muscle mass. Such a spike potential appears rhythmically at a rate of about one or two spikes superimposed at the peak of each slow wave (Fig. 2.4-6B).

3. Action potential with plateau is seen in some tissues, such as ureter, the uterus under some conditions and some types of vascular smooth muscles as shown in Fig. 2.4-6C. This prolonged depolarization accounts for the sustained contraction of certain smooth muscle fibres. However, like skeletal muscle the repolarization does not occur immediately, but is delayed by 100–1000 ms. This prolonged depolarization accounts for the sustained contraction of certain smooth muscle fibres.

Ionic basis of action potential

In smooth muscles, the depolarization occurs due to entry of Ca^{2+} ions from the extracellular fluid (ECF) to inside the cell rather than Na^{+} ions (as seen in skeletal muscles). The smooth cell membrane has far more *voltage-gated calcium channels* than does skeletal muscle but few voltage-gated sodium channels. Unlike sodium channels, the calcium channels open and close slowly. This accounts for the prolonged action potential observed in smooth muscles. The calcium ions, in addition to causing depolarization, also produce contraction of smooth muscles by directly acting on the contractile mechanism.

Electrical activity in multiunit smooth muscles

The multiunit smooth muscles (such as the muscles of iris and piloerector muscles) usually respond to nerve stimuli. The nerve endings secrete neurotransmitter (acetylcholine or norepinephrine), which causes depolarization of the smooth muscle membrane. Since fibres are too small, they do not generate action potential. The local depolarization called the excitatory junctional potential (*EJP).* However, the EJP spreads electrotonically over the entire fibre and is sufficient to cause the muscle contraction.

PROCESS OF EXCITATION–CONTRACTION COUPLING

Excitation–contraction coupling refers to the sequence of events by which excited plasma membrane leads to cross-bridge cycling by increasing cytosolic Ca^{2+} concentration. Since a smooth muscle can be excited by so many possible ways, so there are different ways of excitation–contraction coupling as well. Three different mechanisms of excitation–contraction coupling known in smooth muscles are:

1. ***Electro-mechanical coupling*** occurs when the smooth muscle is excited through sarcolemmal depolarization.
2. ***Pharmaco-mechanical coupling*** occurs when the smooth muscle is excited by some chemical agent.
3. ***Mechano-mechanical coupling*** occurs when the muscle is excited by a stretch.

PROCESS OF SMOOTH MUSCLE CONTRACTION

The molecular mechanism of a smooth muscle contraction by the cross-bridge cycling and sliding of filaments is similar to the skeletal muscle. However, since the smooth muscle does not contain the regulatory protein tropomyosin and troponin, so its regulation is different. In smooth muscle, one of the light chains of the myosin filament located in the neck region serves the function of tropomyosin and thus is called the *regulatory chain of myosin.* Similarly, the Ca^{2+} binding protein *calmodulin* plays the role of troponin.

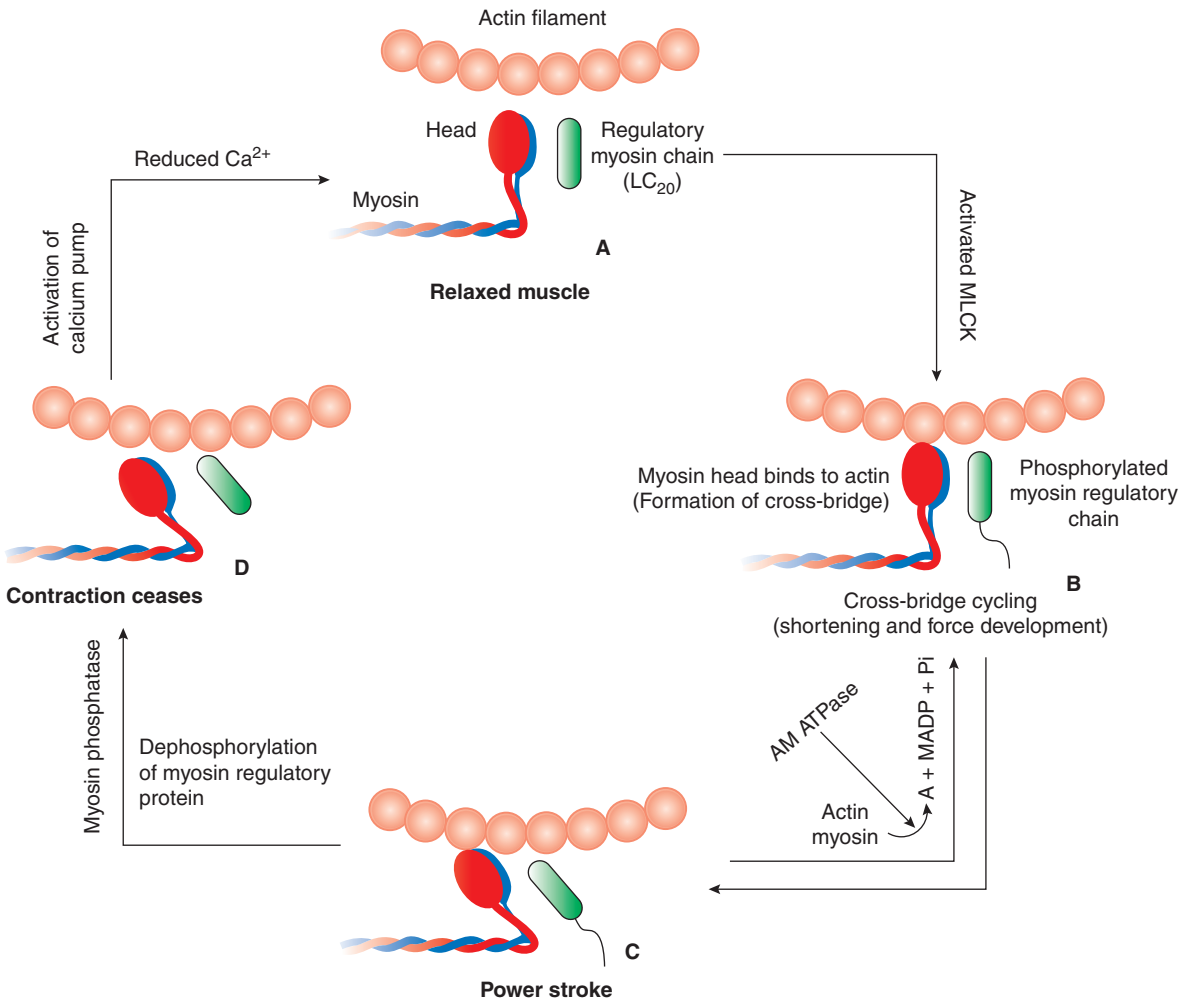


Fig. 2.4-7 Steps of cross-bridge cycling in smooth muscle: A, relaxed muscle; B, activation of myosin light chain kinase (MLCK) catalyzes phosphorylation of myosin regulatory chain and initiate cross-bridging between myosin head and actin filament; C, power stroke, triggered by conformational change in myosin head due to formation of actin-myosin ADP-Pi, complex and D, dephosphorylation of myosin regulatory protein resulting in cessation of cross-bridging.

Steps of cross-bridge cycling

Steps of cross-bridge cycling in a smooth muscle are summarized (Fig. 2.4-7):

1. Activation of the enzyme myosin light chain kinase (MLCK). The enzyme MLCK is activated by the Ca^{2+}–calmodulin complex.

2. Phosphorylation of the myosin regulatory chain. The activated enzyme MLCK uses ATP to phosphorylate the myosin regulatory chain.

3. Cross-bridging. When the myosin regulatory chain is phosphorylated, the head of myosin filament acquires the capability to bind with actin filament to form the cross-bridge (Fig. 2.4-7B).

4. Power stroke. Formation of the actin–myosin ADP-Pi complex triggers a conformational change in the myosin head causing it to flex towards the arm of the cross-bridge. The flexion of the myosin head generates mechanical force (*the power stroke)* (Fig. 2.4-7C).

Due to power stroke, the actin filament slides over the myosin filament producing contraction. As shown in Fig. 2.4-4C the *dense bodies* play a role in the contraction of a smooth muscle fibre. In fact, the dense bodies of smooth muscles serve the same role as the Z-disc (Z line) in the skeletal muscle.

5. Relaxation of smooth muscle. To cause relaxation of the smooth muscle contraction, it is necessary to remove the calcium ions from the sarcoplasm. This is accomplished by the calcium pump which pumps the calcium from intracellular fluid (ICF) to ECF and also from ICF into the sarcoplasmic reticulum. When the cytoplasmic Ca^{2+} falls to the resting level, the processes involved in the contraction of a smooth muscle automatically reverse except for the phosphorylation of the myosin head. Reversal of this occurs when the enzyme *myosin phosphatase* causes

dephosphorylation of the myosin regulatory chain. After this the cross-bridge cycling stops and the contraction ceases (Fig. 2.4-7D). The time required for the relaxation of contracted muscle, therefore, is determined to a great extent by the amount of active myosin phosphatase in the cell.

The calcium pumps operating in the smooth muscles are slow-acting in comparison with the fast-acting sarcoplasmic reticulum pump in skeletal muscles. Therefore, the duration of smooth muscle contraction is prolonged (in seconds) as compared to skeletal muscles (from 1/100th to 1/10th of a second).

CHARACTERISTICS OF SMOOTH MUSCLE EXCITATION AND CONTRACTION

Certain characteristic features of smooth muscle excitation and contraction are as follows:

1. Slow excitation–contraction coupling

A smooth muscle starts contracting approximately 200 ms after the start of the spike potential, (i.e. 150 ms after the spike is over). The peak of contraction is reached after 500 ms of the spike. Thus, the excitation–contraction coupling is very slow in smooth muscles as compared to the skeletal muscles.

2. Plasticity

A smooth muscle exhibits the property of plasticity, i.e. it can readjust its resting length (the length at which a muscle generates maximum active tension). Thus, the smooth muscle defies the usual *length–tension relationship* that is valid for striated muscles (skeletal as well cardiac muscles), when a smooth muscle is passively stretched, it first exerts increased tension, which gradually reduces to prestretch level (even when the stretch is maintained). Therefore, the *length–tension relationship curve* in a smooth muscle is not a smooth curve but a jagged line (Fig. 2.4-8).

3. Latch phenomenon

Latch phenomenon is another characteristic exhibited by smooth muscles. It refers to the mechanism by which a smooth muscle can maintain a high tension without actively contracting. This phenomenon allows long-term maintenance of tone in many smooth muscle organs. In such a state, muscle cannot generate active tension but can effectively resist passive stretching. The latch phenomenon can be explained by the fact that when the myosin kinase and myosin phosphatase enzyme are both strongly activated, the cycling frequency of the myosin heads and the velocity of contraction are great. Then, as the activation of the enzymes decreases, the cycling frequency decreases, but at the same time, the lower activation of the enzymes causes the myosin heads to remain attached to the actin filaments for a larger and longer proportion of the cycling period. Therefore, the number of heads attached to the thin filament at any given time remains large. Because the number of heads attached to the actin determines the static force of contraction, tension is maintained, yet little energy is used by the muscle because ATP is not degraded to ADP, except on the rare occasion when head detaches.

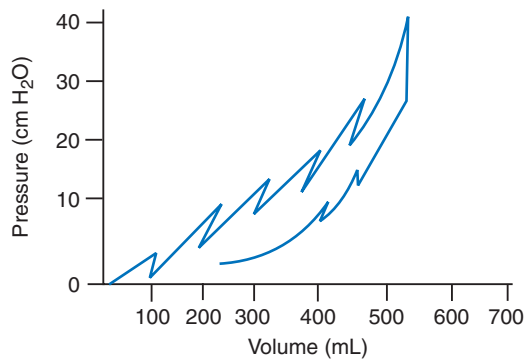


Fig. 2.4-8 Plasticity in smooth muscle fibre.

4. Marked shortening of a smooth muscle during contraction

Marked shortening of a smooth muscle during contraction is another characteristic of smooth muscle that makes it different from the skeletal muscle.

5. Energy required to sustain smooth muscle contraction

Energy required to sustain smooth muscle contraction is much less than required by a skeletal muscle.

EXCITATION AND INHIBITION OF SMOOTH MUSCLES

Excitation of smooth muscles

Multiunit smooth muscles are stimulated only through nerves.

Single unit smooth muscles can be excited by several ways:

- Through nerves (i.e. by neurotransmitter, e.g. Ac-h),
- By hormones,
- Through pacemaker (spontaneous excitation),
- Stretching (due to stretch receptor) and
- Cold temperature.

Inhibition of smooth muscles

- *Through nerves,* (i.e. by neurotransmitter epinephrine) by sympathetic stimulation, e.g. in case of intestinal smooth muscle.

Table 2.4-1 Comparison of skeletal, smooth and cardiac muscles

Feature	Skeletal muscle	Cardiac muscle	Smooth muscle
Structural features			
• Striations	Present	Present	Absent
• Size of fibres			
– Length	1–40 mm	80–100 μm	50–500 μm
– Diameter	50–500 μm	15 μm	2–10 μm
• Shape of the muscle fibre	Cylindrical	Ribbon like	Spindle shaped
• Branching of fibres	Absent	Present	Absent
• Connection between fibres	Absent	Functional connections present forming functional syncytium	In single unit muscle, functional connections are present. In multiunit muscles, no connections
• Nucleus	Single or multiple at periphery	Single, central with many nuclei	Single
• Sarcoplasmic reticulum (SR)	Very well developed	Well developed but not as in skeletal muscles	Moderately developed
• Sarcotubular system	Well developed, two triads per sarcomere, T-tubule present at A–I junction	Present, one triad per sarcomere T-tubule present at Z line	Present, but not well developed
• Thick and thin filaments	Arranged regularly	Arranged regularly	Not arranged regularly
• Sarcomere	Present	Present	Not present
• Regulating protein	Troponin	Troponin	Calmodulin
• Calcium store and calcium pump in SR	High	Moderate	Low
• Sodium channels in the membrane	Fast voltage-gated Na^+ channels	Fast voltage-gated Na^+ channels with slow voltage-gated Na^+–Ca^{2+} channels	Mainly slow voltage gated Na^+–Ca^{2+} channels. Very few fast voltage-gated Na^+channels
• Mitochondria	Few	Many	Few
Nerve supply and control			
• Nerve supply	Somatic nerves	Autonomic nerves – Sympathetic: excitatory (transmitter nor epinephrine) – Parasympathetic: inhibitory (transmitter acetylcholine)	Autonomic nerves – Sympathetic: inhibitory – Parasympathetic: excitatory
• Control	Voluntary	Involuntary	Involuntary
Electrical features			
• Resting membrane potential	–90 mV	–90 mV	–55 mV
• Action potential shape and duration	Spike potential of 5 ms duration	Plateau potential of 100–300 ms	– Single unit muscle: variable, plateau potential of 100–1000 ms and spike potential also seen of 10–50 ms duration – Multiunit muscle: spike potential
• Stimulated by	Somatic nerves	Autonomic nerves	Autonomic nerves, hormones and local tissue factors
• Excitability	High	Moderate	Low
• Conductivity	Fast	Slow	Slow
• Absolute refractory period	1–3 ms	180–200 ms	Not defined
• Autorhythmicity	Not present	Present	Present in single unit muscle
Excitation–contraction coupling			
• Speed of phenomenon	Rapid	Very rapid	Very slow
• Site of calcium attachment	Troponin	Troponin	Myosin
• Mechanism of Ca^{2+} mobilization	T-Tubule is depolarized	Ca^{2+} induced Ca^{2+} release	Inositol triphosphate increases release of Ca^{2+}
• Dependence on concentration of ECF calcium concentration	Not dependent	Partly dependent	Almost totally dependent

(Contd)

Table 2.4-1 Continued

Feature	Skeletal muscle	Cardiac muscle	Smooth muscle
Contractility characteristics			
• Rate of contraction	Fast	Fast	Slow
• Rate of relaxation	Fast	Fast	Slow
• Duration of muscle twitch	– In fast fibres: 7.5 ms – In slow fibres: 100 ms	1½ times the total duration of action potential	About 1000 ms
• All or none law	Obeyed by single muscle fibre	Obeyed by whole muscle	– Single unit muscle: obeyed by whole muscle – Multiunit muscle: obeyed by single muscle fibre
• Multiple fibre (quantal) summations	Possible	Not possible, as it is a functional syncytium	Not possible
• Tetanus (wave summation)	Possible	Not possible due to long refractory period	Not possible, as the process of contraction is long
• Fatigue	Possible	None, since long refractory period ensures recovery and also due to presence of more blood supply	Possible but difficult to demonstrate
• Length–tension relationship	Maximum tension is developed at optimal length	Maximum tension developed at optimal length	Shows property of plasticity
Chemical composition, blood supply, oxygen consumption and muscle energetics			
• Protein	Maximum	Less	Less
• Glycogen	Less	More	Less
• ATP & phosphagen	Present	Present	Present
• Fats	Mainly neutral fats	More phospholipids and cholesterol than others	Mainly neutral fats
• Blood supply	840 mL/min (3–4 mL/100 g/min)	Abundant, 250 mL/min (80 mL/100 g/min)	350 mL/min (1.4 mL/100 g/min)
• Oxygen consumption	Moderate	High	Low
Energy utilization under basal state by			
• Fats	20%	60%	Mainly
• Carbohydrates	60%	35%	Very few
• Proteins	20%	5%	Very few

- *Through hormones* e.g., progesterone decreases the activity of the uterus by acting on pacemaker potential.

CARDIAC MUSCLE

Functional anatomy

- Structural organization of cardiac muscle
- Structure of a cardiac muscle fibre
- Sarcotubular system.

Process of excitability and contractility

- Electrical potentials in cardiac muscle
- Excitation–contraction coupling phenomenon
- Process of cardiac muscle contraction.

Properties of cardiac muscle

- Automaticity
- Rhythmicity
- Conductivity
- Excitability
- Contractility.

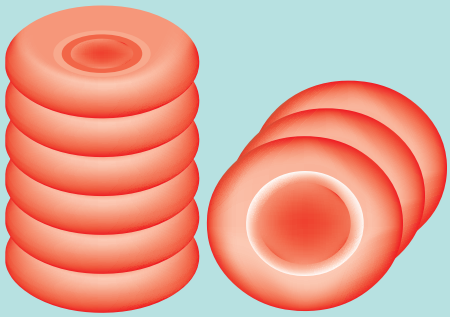
Functional anatomy and physiology of cardiac muscle is discussed in Chapter 4.1 of 'Cardiovascular System' (see page 177).

COMPARISON OF SKELETAL, SMOOTH AND CARDIAC MUSCLES

The salient features of skeletal, smooth and cardiac muscles are shown in Table 2.4-1.

SECTION 3

Blood and Immune System



Blood is a fluid connective tissue which transports substances from one part of the body to another. It provides nutrients and hormones to the tissues and removes their waste products. Blood, confined in the cardiovascular system, constitutes a major part of the extracellular fluid of the body.

Some of the important physical characteristics of blood are:

- *Colour* of the blood is opaque red due to the pigment haemoglobin in the red blood cells (RBCs). The arterial blood is bright red and venous blood is dark red in colour.
- *Volume* of blood in an average adult is about 5–6 L (8% of the body weight or 80 mL/kg body weight).
- *Viscosity* of blood is five times more than that of water.
- Specific gravity of blood is 1.050–1.060. Specific gravity of RBC is greater (1.090) than that of plasma (1.030).
- *pH* of blood is about 7.4 (ranges from 7.38 to 7.42), i.e. it is alkaline in nature. In acidosis, pH of blood falls below 7.38 and in alkalosis, pH is more than 7.42.

Blood is composed of two main components, plasma and cellular elements.

Plasma constitutes about 55% of the blood volume. It is a clear straw coloured fluid portion of blood. *Plasma proteins,* an important constituent of plasma, form about 7% of its volume.

Cellular elements of blood are about 45% of the total blood volume and constitute the so-called packed cell volume. Blood cells are:

- Erythrocytes or RBCs (5 million/μL)
- Leucocytes or white blood cells (4000–11,000/μL)
- Platelets or thrombocytes (1.5–4 lac/μL).

FUNCTIONS OF BLOOD

1. *Nutritive function.* Blood carries the nutritive substances like glucose, amino acids, fatty acids, vitamins, electrolytes and others from the gut to the tissues where they are utilized.
2. *Respiratory functions.* Blood picks up oxygen from the lungs and delivers it to the various tissues. Most important function of the blood is the uninterrupted delivery of O_2 to the heart and the brain. It also carries away CO_2 from the tissues to the lungs from where it is expelled out in the expired air.
3. *Excretory function.* Blood transports various metabolic waste products, such as urea, uric acid and creatinine to excretory organs (kidney, skin, intestine and lungs) for their disposal.
4. *Transport function.* The various hormones produced by the endocrine glands, the biological enzymes and antibodies are transported by the blood to the target tissue to modulate metabolic process.
5. *Protective function.* Blood plays an important role in the defence mechanism of the body:
   - Neutrophils and monocytes engulf the microorganisms entering the body by phagocytosis.
   - Lymphocytes and γ globulins initiate immune response.
   - Eosinophils accomplish detoxification, disintegration and removal of foreign proteins.
6. *Homeostatic function.* Blood plays an important role in maintaining the internal environment of the body (homeostatic function):
   - The water content of blood is freely interchangeable with the interstitial fluid and helps in maintaining the water and electrolyte balance of the body.
   - Plasma proteins and haemoglobin act as buffers and help in maintaining the acid–base balance and pH of the body fluids.
7. *Maintenance of body temperature.* Blood plays an important role in regulation of the body temperature, as described:
   - *Specific heat* of blood is high, which is useful in buffering the sudden changes in the body temperature.
   - *High heat conductivity* of blood renders it possible for distribution of heat from deep organs to the skin and lungs for dissipation.
   - Due to *high latent heat* of evaporation of blood, a large amount of heat is lost from the body by evaporation of water from the lungs and skin.
8. *Storage function.* Blood serves as a ready-made source of substances stored in it (such as glucose, water, proteins and electrolytes for use in emergency conditions like starvation, fluid loss and electrolyte loss).

IMMUNE SYSTEM

The immune system which constitutes the body's defence system consists of immunological cells distributed in two main components: mononuclear phagocytic system and lymphoid component. The immune system of the body responds to an antigen by two ways:

- *Humoral or antibody-mediated immunity* which is mediated by antibodies produced by the plasma cells.
- *Cell-mediated immunity* which is mediated directly by the sensitized lymphocytes.

Chapter 3.1

Plasma and Plasma Proteins

PLASMA

Plasma is the clear straw coloured fluid (with dissolved solid substances) portion of the blood minus its cellular elements. It constitutes about 55% of the blood volume (about 5% of body weight).

COMPOSITION

Plasma contains the following constituents:

Water. Water is the main constituent of plasma forming 91% of it.

Solids. The solids dissolved in the plasma constitute a total of 9% of the plasma. The solid constituents of plasma are given below:

Plasma proteins form 7% of the solids in plasma. Their normal value ranges from 6.4 to 8.3 g/dL. They include albumin, globulins, fibrinogen and others.

Other organic molecules which form 1% of the solids include the following:

- *Carbohydrates,* mainly glucose (100–120 mg/dL).
- *Fats* are neutral fats (30–150 mg/dL), phospholipids (150–300 mg/dL) and cholesterol (150–240 mg/dL).
- *Non-protein nitrogenous substances* (28–40 mg/dL) are ammonia (traces), amino acids, creatine (1–2 mg/dL), creatinine (0.6–1.2 mg/dL), xanthine (traces), hypoxanthine (traces), urea (20–40 mg/dL) and uric acid (2–4 mg/dL).
- *Hormones, enzymes and antibodies.*

Inorganic substances which constitute 1% of the solids in plasma include sodium, potassium, calcium, magnesium, chloride, iodide, iron, phosphates and copper.

Gases present in the plasma are oxygen, carbon dioxide and nitrogen.

Serum

Plasma from which fibrinogen and clotting factors (II, V and VIII) have been removed is called serum. Serum is formed when the blood is allowed to clot in a test tube and the clot is retracted. Serum has a higher serotonin (5HT) content because of the breakdown of platelets during clotting.

PLASMA PROTEINS

CLASSIFICATION OF PLASMA PROTEINS

Plasma proteins form the major solid constituent of the plasma. The total plasma protein concentration is 7.4 g/dL (ranges from 6.4 to 8.3 g/dL). Presently, more than 100 types of plasma proteins have been identified. The original classification is based on the classical method of precipitation of salts as described by Howe (1922). By electrophoretic techniques the globulins have been further subclassified. Based on these, the important fractions of plasma proteins are given below:

- Albumin (4.8 g/dL),
- Globulins (2.3 g/dL) include:
  - α_1 globulin
  - α_2 globulin
  - β globulin and
  - γ globulin
- Fibrinogen (0.3 g/dL).

Electrophoretic protein patterns. On the basis of paper electrophoresis, following classes of serum proteins (Fig. 3.1-1) are identified:

- Albumin (55%),
- α_1 globulins (5%),
- α_2 globulins (9%),
- β globulins (13%) and
- γ globulins (11%).

PROPERTIES OF PLASMA PROTEINS

1. Molecular weight. Plasma proteins are large molecules with the following molecular weight:

- Albumin: 69,000,
- Globulins: From 90,000 to 1,56,000 and
- Fibrinogen: 5,00,000.

Thus the fibrinogen has got highest molecular weight. Relative size and shape of different plasma proteins are shown in Fig. 3.1-2.

2. Osmotic pressure. The plasma proteins exert an oncotic pressure of about 25 mm Hg.

3. Specific gravity. The specific gravity of plasma proteins is 1.026.

4. Isoelectric point. Proteins can ionize either as acids or as bases owing to the fact that the side chains of their constituent amino acids contain a selection of amino group (NH_2) and carboxyl groups (–COOH). At an intermediate pH (specific for each protein), the protein molecules carry equal number of positive and negative charges and hence have a zero net charge. This pH value for electrical neutrality of the molecule is known as the *isoelectric point.*

5. Electrophoretic mobility. The proteins act as *anions* in alkaline solutions and as *cations* in acidic solutions. Because of this property they possess electrophoretic mobility.

6. Precipitation by salts. Proteins can be precipitated by different concentrations of salts. This property of proteins is utilized for their separation by the precipitation method.

Ammonium sulphate solution is utilized to precipitate the plasma protein fractions in different strengths:

- *Albumin* is precipitated by full saturation.
- *Globulins* are precipitated by half saturation. Among the globulins, there is a fraction which can be precipitated by one-third saturation with ammonium sulphate and is termed *euglobulin.* The rest is called *pseudoglobulin.*
- *Fibrinogen* is separated by one-fifth saturation with ammonium sulphate.

7. Water solubility. The protein molecules are soluble in water because of the presence of polar residues like NH_2 and COOH.

8. Amphoteric nature. Protein molecules are amphoteric in nature because of the presence of NH_2 and COOH groups. By virtue of their amphoteric nature the plasma proteins act as efficient buffers.

FEATURES OF INDIVIDUAL FRACTION OF PLASMA PROTEINS

1. Albumin

- ***Plasma levels*** are 4.8 g/dL (range 3–5 g/dL).
- ***Molecular weight*** of prealbumin is 60,000 and of albumin is 69,000.

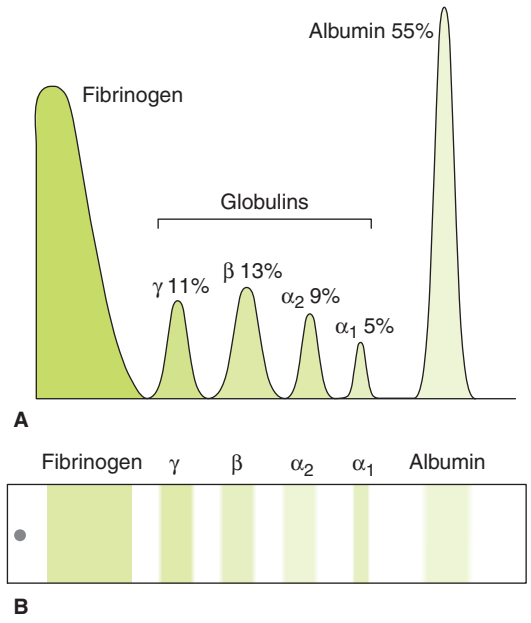


Fig. 3.1-1 Paper electrophoresis showing: A, relative amount of plasma proteins and B, bands of different plasma proteins.

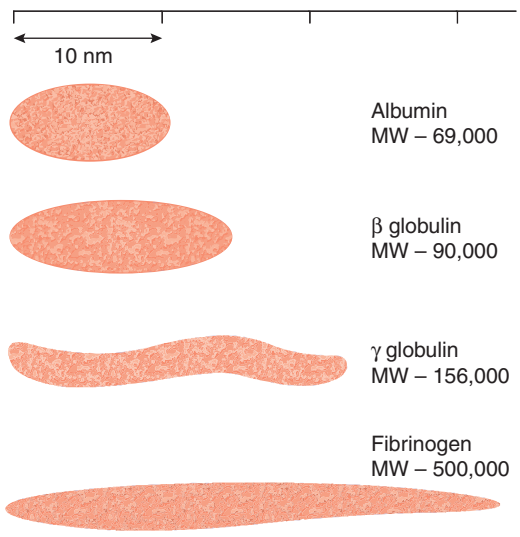


Fig. 3.1-2 Molecular weights and shapes of different plasma proteins.

- ***Synthesized*** in liver.
- ***Half-life*** is about 10 days.

2. Globulins

- ***Plasma levels*** are 2.3 g/dL (range 2 to 3 g/dL).
- ***Molecular weight*** varies from 90,000 to 156,000.

Types include α_1, α_2, β_1, β_2 and γ globulins.

Forms of globulins are described below:

- *Glycoproteins* consist of carbohydrates and protein
- *Lipoproteins* consist of α_2 globulin and lipids. It has got the following subtypes:
  - *High density lipoproteins (HDL).* These are α lipoproteins which contain 50% protein with large amount of cholesterol and phospholipids.
  - *Low density lipoproteins (LDL).* These are β lipoproteins and contain large amount of glycerides.
  - *Very low density lipoproteins.* These are also β lipoproteins and have higher proportion of fat in the form of triglycerides or cholesterol.
  - *Chylomicron* contains 98% triglycerides. It is synthesized in the intestine following a meal.
- *Transferrin* is an α_2-β globulin having a molecular weight of 90,000. It has the specific property of iron binding and thus helps in its transport and storage. Each molecule of transferrin binds two atoms of ferric iron.
- *Haptoglobin* is an α_2 globulin having a molecular weight of 90,000. It forms stable complexes with free haemoglobin.
- *Ceruloplasmin* is an α_2-β globulin having a molecular weight of 16,000. It binds with copper and helps in its transport and storage. Its deficiency causes Wilson's disease (hepatolenticular degeneration), in which liver and brain are damaged due to high levels of free copper.
- *Fetuin* is a growth promoting protein seen in infants and newborn.
- *Immunoglobulins* are γ globulins which play role in immunity.
- *Angiotensinogen* is an α_2 globulin.
- *Haemagglutinins* are antibodies against the RBCs antigens.

3. Fibrinogen

- ***Plasma levels*** are 0.3 g/dL.
- ***Molecular weight*** is 500,000.
- ***Synthesized*** in the liver.
- ***Chemical structure.*** Protein part of the molecule is made up of six polypeptide chains (α_2, β_2, γ_2) joined by disulphide bonds.
- ***Functions*** as a clotting protein.

4. Prothrombin

- ***Plasma levels*** are 40 mg/dL.
- ***Molecular weight*** is 68,000.
- ***Synthesized*** in the liver. Synthesis is promoted by vitamin K.

APPLIED ASPECTS

Electrophoretic separation is very useful in clinical diagnosis. It helps in knowing:

- The change in relative concentration of different proteins.
- The presence of abnormal proteins.
- The absence of normal proteins.

FUNCTIONS OF PLASMA PROTEINS

1. Exert osmotic pressure. The protein molecules are unable to pass across the capillary membrane and consequently exert colloid osmotic pressure of about 25 mm Hg on the capillary membrane. About 70–80% of the osmotic pressure is contributed by the albumin fraction. The colloid osmotic pressure plays an important role in exchange of water between the blood and tissue fluid.

2. Contribution to blood viscosity. Fibrinogen and globulins are significant contributors to blood viscosity because of their asymmetrical shape. The blood viscosity plays an important role in the maintenance of blood pressure by providing resistance to flow of blood in blood vessels.

3. Role in coagulation of blood. The fibrinogen, prothrombin and other coagulation proteins present in the plasma play an important role in the coagulation of blood.

4. Role in defence mechanism of the body. The γ globulins are antibodies which play an important role in the immune system meant for defence of the body against the microorganisms.

5. Role in maintaining acid–base balance of the body. Plasma proteins act as buffers and contribute for about 15% of the buffering capacity of blood. Because of their amphoteric nature, plasma proteins can combine with acids and bases as explained below:

- *In acidic pH,* the NH_2 group of the proteins acts as base and accepts proton and is converted to NH_4.
- *In alkaline pH,* the COOH group of the proteins acts as acid and can donate a proton and thus becomes COO^-.
- *At normal pH of blood,* proteins act as acids and combine with cations (mainly sodium).

6. Transport function. Plasma proteins combine easily with many substances and play an essential role in their transport as explained below:

- *Carbon dioxide* is transported by the plasma proteins in the form of carbamino compound.
- *Thyroxine* is transported by an α globulin called thyroxine-binding protein.
- *Cortisol* is transported by transcortin which is a mucoprotein.
- *Vitamin A, D and E* are transported by the high and low density lipoproteins (HDL and LDL).
- *Vitamin B_{12}* is bound to transcobalamin for transport.
- *Bilirubin* is associated with albumin and also with fractions of the α globulin.
- *Drugs* of various types are transported after combining with the albumin.
- *Calcium* of the plasma is partly (50%) bound to the proteins for transport.
- *Copper* is bound to ceruloplasmin (α_2 globulin) for transport.
- *Free haemoglobin* in the vessels is bound by haptoglobin and carried to reticuloendothelial system.

7. Role as reserve proteins. Plasma proteins serve as reserve proteins and are utilized by the body tissues during conditions like:

- Fasting,
- Inadequate protein intake and
- Excessive catabolism of body proteins.

8. Role in suspension stability of RBCs. Suspension stability refers to the property of RBCs by virtue of which they remain uniformly suspended in the blood. Globulins and fibrinogen accelerate this property.

9. Fibrinolytic function. The enzymes of the fibrinolytic system digest the intravascular clot (thrombus) and thus save from the disastrous effects of thrombosis.

10. Role in genetic information. Many plasma proteins *exhibit polymorphism.* Polymorphism is a Mendelian trait that exists in the population with differing prevalence.

11. Role of nourishment of tissue cells. The plasma proteins are utilized by the leucocytes to produce the substances known as *trephones* or *carrel* which are essential for the nourishment of tissue cells.

SYNTHESIS OF PLASMA PROTEINS

Site of synthesis

In embryo, the plasma proteins are synthesized by the mesenchymal cells. First, the albumin is produced and then the other proteins are synthesized.

In adults, plasma proteins are synthesized as described below:

- The albumin and fibrinogen are synthesized mostly by the reticuloendothelial cells of the liver.
- α and β globulins are synthesized by the liver, spleen and bone marrow.
- γ globulins are synthesized by the B lymphocytes.

Factors affecting synthesis of plasma proteins

1. Dietary proteins

Dietary proteins play the most essential role in the synthesis of plasma proteins. The relation of plasma proteins to diet was studied in the plasma protein depleted dogs, first of all by Whipple by an experimental procedure called *plasmapheresis.*

Plasmapheresis. In this experiment, the dog is rendered hypoproteinaemic by repeatedly withdrawing whole blood and injecting back the cellular elements of the blood (suspended in the Ringer–Locke solution). This process is repeated daily till the level of plasma proteins falls to 4 g/100 mL. Thereafter, different standard diets are given and their effects on protein synthesis are studied. Following conclusions have been drawn from these experiments:

- *Dietary proteins* are essential for the synthesis of plasma proteins.
- *Essential amino acids* must be present in the diet for the satisfactory synthesis of plasma proteins.
- *Dietary proteins of animal origin* favour albumin synthesis.
- *Dietary proteins of plant origin* favour globulin synthesis.

2. Other factors

Other factors which affect plasma protein synthesis in the body are the following:

- *Presence of infection* in the body reduces plasma protein synthesis.
- *Exposure to some antigen* stimulates formation of antibodies.
- *Inflammatory conditions* promote the synthesis of a number of proteins.
- *Interleukin-1,* a material released by the activated macrophages in the body, stimulates the synthesis of many acute phase proteins in the liver.
- *Prostaglandins* are also reported to increase the synthesis of acute phase proteins, possibly through stimulation of macrophage release of interleukin-1.

CHANGES IN PLASMA PROTEINS IN HEALTH AND DISEASE

Physiological variations

- *In infants,* the total protein level is low (about 5.5 g/dL) due to low γ globulins.

- *In old age,* there is a tendency for the albumin level to fall and the total globulin level to rise.
- *In pregnancy,* during first six months, the albumin and globulin levels decrease while the fibrinogen level increases.

Abnormalities of plasma protein levels

Hypoproteinaemia

Hypoproteinaemia refers to generalized decrease in the levels of plasma proteins.

Causes of hypoproteinaemia include the following:

- *Dietary deficiency* and starvation are associated with hypoproteinaemia.
- *Malabsorption syndrome* due to intestinal diseases such as sprue is associated with hypoproteinaemia.
- *Liver diseases* like hepatitis and cirrhosis cause hypoproteinaemia due to reduced synthesis of proteins in the liver.
- *Renal diseases* like nephrotic syndrome cause hypoproteinaemia due to more loss of proteins in the urine.
- *Haemorrhage and extensive burns* are associated with acute hypoproteinaemia.
- *Hereditary analbuminaemia* is an inborn defect in the genetic level where there is no synthesis of albumin.
- *Congenital afibrinogenaemia* is a rare condition characterized by defective blood clotting.

Effects of hypoproteinaemia. Low levels of plasma proteins are associated with a decrease in the plasma osmotic pressure which causes water retention and oedema of the body tissue.

Hyperproteinaemia

Hyperproteinaemia, i.e. increase in the plasma protein levels, is seen in following conditions:

- *Acute inflammatory conditions* are associated with increased synthesis of the so-called *acute phase proteins* which include C-reactive proteins, α antitrypsin, haptoglobin, fibrinogen and ceruloplasmin.
- *Chronic inflammation and malignancies* are also associated with raised levels of C-reactive proteins.
- *Multiple myeloma* is associated with increased levels of the so-called *Bence Jones* proteins and myeloma globulin due to their abnormal formation in the bone marrow.

Reversal of normal A/G ratio

The normal albumin: globulin (A/G) ratio (1.7:1) is reversed in the following conditions:

- When the albumin synthesis is decreased as occurs in liver diseases (globulin levels being normal because many globulins are synthesized by the B lymphocytes).
- When the globulin levels are increased (as occurs in most of the conditions) associated with hyperproteinaemia.

Chapter 3.2

Red Blood Cells and Anaemias

CHARACTERISTIC FEATURES OF RED BLOOD CELLS

FUNCTIONAL MORPHOLOGY

The red blood cells (mature erythrocytes) form one of the important constituent of the cellular elements of the blood. Each red blood cell (RBC) like any other cell in the body is bounded by a *cell membrane* but is *non-nucleated* and lacks the usual cell organelles. The cytoplasm of the RBC contains a special pigmented protein called the *haemoglobin* (Hb) which forms 90% of the weight of the erythrocytes. The red colour of the RBCs and thus of the blood is due to the presence of Hb.

NORMAL SIZE, SHAPE AND COUNTS OF RBCs

Normal size

- *Diameter* of each RBC is 7.2 µm (range 6.9–7.4 µm),
- *Thickness* in the periphery is 2 µm and in the centre 1 µm,
- *Surface area* of each RBC is about 120–140 μm^2 and
- *Volume* is about 80 μm^3 (range 78–86 μm^3).

Normal shape

The RBCs are circular, biconcave discs (Fig. 3.2-1).

Advantages of biconcave shape are:

- It renders the red cells quite flexible so that they can pass through the capillaries whose minimum diameter is 3.5 µm (Fig. 3.2-2).
- The biconcavity provides greater surface area as compared to volume which allows considerable alterations in the cell volume. Thus, the RBC can withstand considerable changes of osmotic pressure. In this way, the RBCs can resist haemolysis to certain extent when placed in the hypotonic solution.
- Greater surface area allows easy exchange of O_2 and CO_2 and rapid diffusion of other substances.

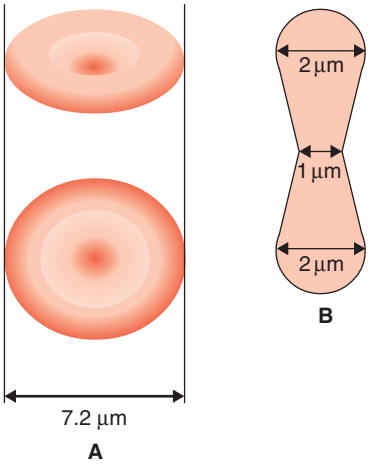


Fig. 3.2-1 Size and shape of a normal red blood cell: A, biconcave disc (diameter 7.2 µm) and B, thickness (2 µm at the periphery and 1 µm in the centre).

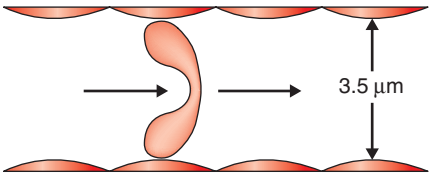


Fig. 3.2-2 Diagram showing how flexibility of red blood cell allows it to pass through smaller capillaries (diameter 3.5 µm).

Normal counts

Clinically, a count of 5 million/µL is considered as 100%.

VARIATIONS IN SIZE, SHAPE AND COUNTS OF RBCs

Variations in size

Variation in size is called *anisocytosis:*

- *Microcytosis*, i.e. decrease in the size of RBCs occurs:
  - In iron deficiency anaemia,
  - During prolonged forced breathing and
  - When osmotic pressure of the blood is increased.
- *Macrocytosis,* i.e. increase in the size of RBCs occur:
  - In megaloblastic anaemia,
  - During muscular exercise and
  - When osmotic pressure of the blood is decreased.

Variations in shape

Variation in shape is called *poikilocytosis.* Abnormal shapes of the RBCs are given below:

- *Spherocyte*
- *Elliptocytes*
- *Sickle cell*
- *Poikilocytes*

Variations in counts

Physiological increase in the RBC count (physiological polycythaemia) is seen in the following circumstances:

- *Age.* At birth, the RBC count is 6–7 million/µL of blood. After about 10 days of birth the count decreases due to the destruction of cells. This is the cause of *physiological jaundice* of newborn. In infants, the RBC count is slightly more than the adults.
- *Sex.* RBC count in the adult females (average 4.8 million/µL) is lower than the adult males (average 5.5 million/µL).
- *High altitude.* Individuals residing in high altitude areas (above 10,000 feet from the sea level) have high RBC count (7 million/µL) because of the hypoxic stimulation of erythropoiesis.
- *Excessive exercise.* Mild hypoxia and spleen contraction causes temporary increase in the RBC count.
- *Emotional conditions* like anxiety are associated with the temporary increase in the RBC count due to sympathetic stimulation.
- *After meals,* the RBC count is raised slightly.

Polycythaemia or pathological increase in the RBC count (above 7 million/µL) is of two types:

- *Primary polycythaemia or polycythaemia vera* occurs in myeloproliferative disorder like malignancies of the bone marrow. The RBC count is persistently above 14 million/µL and is always associated with high white blood cell (WBC) count.
- *Secondary polycythaemia* occurs due to certain conditions producing a state of chronic hypoxia in the body such as:
  - congenital heart disease and
  - chronic respiratory disorders like emphysema.

Physiological decrease in RBC count is seen in the following conditions:

- At high barometric pressure,
- After sleep and
- In pregnancy.

Anaemia. In anaemia, there may occur marked reduction in the RBC count or the Hb level or both (see page 117).

PACKED CELL VOLUME AND RED CELL INDICES

Packed cell volume

Packed cell volume (PCV) refers to the percentage of the cellular elements (RBCs, WBCs and platelets) in the whole blood. Since the volume of WBCs and platelets is very less, so for all practical purposes the PCV is considered equivalent to the volume of packed red cells or the so-called *haematocrit value.* The normal values of the PCV in males

are about 45% and in females about 42%. The PCV is increased in polycythaemia and decreased in anaemia.

Red cell indices

The red cell indices defined below are calculated taking normal values of the RBC count 5 million/μL, PCV 45% and Hb level of 15 g/dL.

1. Mean corpuscular volume

The mean corpuscular volume (MCV) refers to the average volume of a single RBC. It is calculated by dividing the PCV by the red cell count.

$$\text{MCV} = \frac{\text{PCV in 1000 mL of blood, i.e. (PCV} \times 10)}{\text{RBC count/μL}}$$

$$= \frac{45 \times 10}{5} = 90\,\mu m^3$$

- Normal value of the MCV is 90 μm^3 (range 78–94 μm^3) when the MCV value is normal is referred as *normocytosis.*
- Decreased value of the MCV occurs in *microcytosis.*
- Increased value of the MCV occurs in *macrocytosis.*

2. Mean cell haemoglobin

Mean cell haemoglobin (MCH) refers to the average weight of the Hb contained in each RBC. It is calculated by dividing the amount of Hb in 1 L of blood by the red cell count in 1 L of blood.

$$\text{MCH} = \frac{\text{Hbg/L}}{\text{RBC count/L}} = \frac{\text{Hb g\%} \times 10}{\text{RBC count/μL} \times 10^{12}}$$

$$= \frac{15 \times 10}{5 \times 10^{12}} = 30 \times 10^{-12}\ \text{g}$$

$$= 30\ \text{pg [since } 10^{-12}\ \text{g} = \text{one picogram (pg)]}$$

- *Normal value of MCH* is 30 (range 27–33) pg.
- *Increased values of MCH* occur in the spherocytosis and in the megaloblastic anaemia.

3. Mean cell haemoglobin concentration

The MCH concentration (MCHC) refers to the amount of Hb expressed as a percentage of the volume of a RBC. It is calculated by dividing the amount of Hb in g/dL by the volume of packed cells in 100 mL of blood and then multiplying by 100.

$$\text{MCHC} = \frac{\text{Hb g\%}}{\text{PCV/100 mL}} \times 100 = \frac{15}{45} \times 100 = 33.3\%$$

- *Normal value of MCHC* is 33.3% (range 30–38%). RBCs with normal values of MCHC are called *normochromic.*
- *In hypochromic RBCs* the values of MCHC are less than the normal, as is seen in the iron deficiency anaemia.
- *Hyperchromia* is very rare, high levels of MCHC (>38%) cannot occur, since the RBCs cannot hold Hb beyond the saturation point.
- Since MCHC is independent of the RBC count and the size of RBCs, it is considered to be of greater clinical significance as compared to other absolute values.

4. Colour index

The colour index (CI) refers to the ratio of Hb to RBC. For calculating CI, Hb of 14.8 g/dL is taken as 100% and RBC count of 5 million/μL is taken as 100%.

$$\text{CI} = \frac{\text{Percentage of normal Hb}}{\text{Percentage of normal RBC count}} = \frac{100}{100} = 1$$

For example, if Hb level is 14.8 g/dL, i.e. 100% of normal values and RBC count is 4.5 million, i.e. $4.5 \times 100 = 90\%$ of the normal values,

$$\text{then CI} = 100/90 = 1.11$$

- Normal values of CI vary from 0.85 to 1.15.
- CI is insignificant because normal range of RBC is very wide. Therefore, it has been long abandoned and is not used for any diagnostic purposes.

ROULEAUX FORMATION AND ERYTHROCYTE SEDIMENTATION RATE

Rouleaux formation

- Rouleaux formation refers to the tendency of the RBCs to pile one over the other like a pile of coins (Fig. 3.2-3). The discoid shape and protein coating of red cells play a major role in the rouleaux formation. Rouleaux formation does not occur in the normal circulation under physiological conditions, as the moving cells show little or no tendency to adhere.
- This is a reversible phenomenon, but it promotes sedimentation of the RBCs.
- Albumin decreases the rouleaux formation; while, fibrinogen, globulin and other products of tissue destruction increase rouleaux formation.

Erythrocyte sedimentation rate

Erythrocyte sedimentation rate (ESR) is the rate at which the RBCs sediment (settle down) when the blood containing an anticoagulant is allowed to stand in a vertically placed tube. It is expressed in millimetre at the end of first hour.

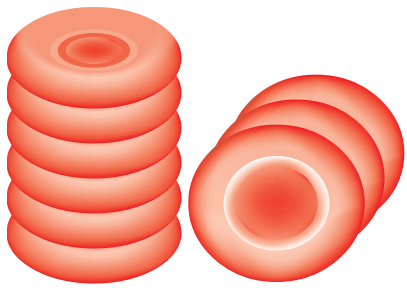


Fig. 3.2-3 Rouleaux formation.

Clinical significance of ESR

- *Normal values* of ESR by Westergren's method in males vary from 3 to 7 mm and in females from 5 to 9 mm in first hour.
- Values of ESR are raised in a large number of pathological conditions, so it has got no specific diagnostic value. However, raised levels of ESR do suggest presence of some chronic inflammatory condition in the body.
- Estimation of ESR is more useful as a *prognostic test*, i.e. to judge the progress of the disease in patients under treatment.

Factors affecting ESR

- *Rouleaux formation.* Increased tendency of the rouleaux formation raises the ESR. Fibrinogen and the proteins, which enter the plasma in inflammatory (globulins) and neoplastic diseases, favour rouleaux formation and thus increase the ESR. Increase in MCV, decrease in MCH and spherocytosis retard rouleaux formation and thus decrease the ESR.
- *Size of the RBCs.* Increase in the size of the RBCs (macrocytosis) raises the ESR.
- *Number of RBCs.* When the number of RBCs increased, the ESR is decreased and when the number of RBCs is decreased (as in anaemia), the ESR is increased.
- *Viscosity of blood.* ESR is increased when the viscosity of blood is decreased and vice versa.

Physiological variations in ESR

- *Age.* ESR is less in infants and old people as compared to young adults.
- *Sex.* ESR is greater in the females (5–9 mm) than the males (3–7 mm).
- *Menstruation.* ESR is slightly raised during menstruation in the females.
- *Pregnancy.* ESR is raised in pregnancy from third month to parturition and returns to normal after 3–4 weeks of delivery.

Pathological variations in ESR

Increase in the ESR is seen in the following pathological conditions:

- Tuberculosis,
- Malignant diseases,
- Collagen diseases,
- All anaemias (except sickle cell anaemia) and
- Chronic infections.

Decrease in the ESR occurs in the following pathological conditions:

- Polycythaemia,
- Decreased fibrinogen levels,
- Sickle cell anaemia and
- Allergic conditions.

RED CELL MEMBRANE, COMPOSITION AND METABOLISM OF RBCs

RED CELL MEMBRANE

Structure

- Red cell membrane is a *trilaminar structure* having a bimolecular lipid layer interposed between the two layers of proteins (Figs 3.2-4 and 1.2-2).
- *Important lipids* of the cell membrane are glycolipids, phospholipids and cholesterol.
- *Proteins* in the cell membrane are present as peripheral proteins and integral proteins spanning the whole membrane.
  - The outer peripheral protein surface is rich in *lecithin and sphingomyelin.*
  - The important membrane spanning integral proteins include an anion exchange protein (*band 3)* and *glycophorins* which contain a number of polysaccharide *blood group antigens.*
  - The inner surface of the cell membrane contains more phosphatidylserine and phosphatidyl ethanolamine. The peripheral proteins like spectrin, ankyrin and actin present on the inner surface of the membrane help in maintaining the shape and flexibility of the RBC. *Spectrin* is the major protein of the cytoskeleton. Protein 4.1 binds both to spectrin and actin and also interacts with certain phospholipids (thereby connecting the cytoskeleton to the lipid layer).

Permeability

The red cell membrane is a *semi-permeable membrane*, allowing some substances to pass through and preventing some.

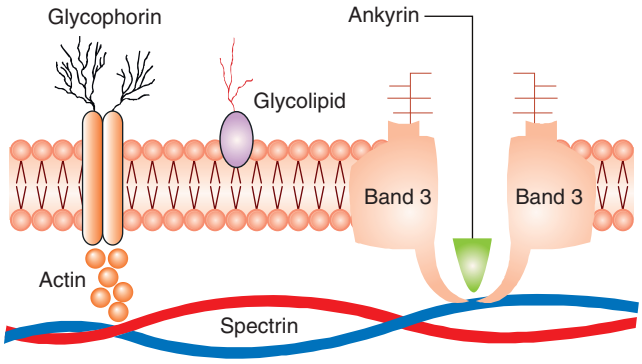


Fig. 3.2-4 Schematic diagram showing ultra structure of red cell membrane.

- *Impermeable* to sodium, calcium and barium ions, fats and sugars,
- *Slightly impermeable* to amino acids and
- *Freely permeable* to all anions like Cl^-, SO_4^- and HCO_3^-, and to urea, ammonia, aldehyde, alcohol and bile salts.

COMPOSITION OF RBCs

The body of the RBC bounded by the cell membrane contains a sponge-like stroma which is composed of the following structures:

- *Water* constitutes 60% of the wet weight of the RBC.
- *Hb,* held in the meshes of stroma, constitutes 35% of the wet weight and 90% of the dry weight of the RBC.
- *Lipids* form the major constituents of the rest of 5% of stroma. These include cephalin, lecithin and cholesterol.
- *Proteins* include glutathiones and an albumin-like insoluble protein. These act as reducing agents preventing damage to the Hb.
- *Lipoproteins.* Almost half of the lipids are bounded to the protein forming a lipoprotein complex known as elenin (Calvin).
- *Enzymes* of the glycolytic system, catalase, carbonic anhydrase and other enzymes and inorganic salts are also present in the RBC.
- *Glucose and amino acids* are present in a small amount.
- *Ions.* Anions of the plasma (Cl^-, PO_4^{3-}, HCO_3^-) are present often in large amounts. The cations Na^+ and Ca^{2+} are either present in a very small amount or are absent. Cation K^+ is present in a sufficient amount inside the RBC.
- *Non-protein nitrogenous (NPN) substances.* Urea, NH_4, creatine and uric acid have a higher concentration inside the RBC than the plasma.

METABOLISM OF RBCs

Glucose, taken up by the facilitated diffusion, is the only fuel utilized by the RBC. The mature RBC has a low respiratory quotient and consumes very little oxygen. The glucose metabolism and its special significance in a RBC is described below:

Embden–Meyerhof pathway is responsible for 90% of the glycolysis. Two molecules of ATP are generated from each molecule of glucose. Special significance of this pathway is given below:

2,3 Diphosphoglycerate (2,3 DPG) synthesized in a side reaction in this pathway influences the oxygen affinity of Hb and thus plays an important role in the red cell physiology.

Hexose monophosphate (HMP) shunt oxidizes about 10% of glucose. *Glucose-6 phosphate dehydrogenase (G6PD)* is the key enzyme in the HMP shunt. Inherited deficiency of this enzyme leads to the compromise of RBC function and viability in the face of such oxidant stress; thus, causing a number of disorders characterized by susceptibility to haemolysis.

Utilization of ATP

- A major portion of ATP obtained through glycolysis is utilized in maintaining the Na^+–K^+ ATPase pump.
- Some of the energy is utilized in maintaining the integrity of the red cell membrane.
- Some of the energy is spent in maintaining the Hb iron in the reduced form (Fe^{2+}).

FORMATION OF RED BLOOD CELLS

Formation of RBCs is a part of the process of development of blood cells (RBCs, WBCs and platelets) called *haemopoiesis* which includes:

- Erythropoiesis, i.e. development of RBCs,
- Leucopoiesis, i.e. development of WBCs and
- Thrombopoiesis or megakaryocytopoiesis, i.e. development of platelets.

SITES OF HAEMOPOIESIS

- *In the first two months of gestation*, the *yolk sac* is the main site of haemopoiesis (mesoblastic stage).
- *From third months of gestation,* liver and spleen become the main sites of blood formation and continue to do so till birth. Spleen makes small contribution as compared to the liver (hepatic stage).
- *From 20th week of gestation,* haemopoiesis begins in the bone marrow and by seventh or eighth month it becomes the main site (myeloid stage).
- The active haemopoietic bone marrow is red in colour due to marked cellularity and hence is called *red bone marrow.* However, during this period there occurs a progressive fatty replacement throughout the long bones converting red bone marrow into the so-called *yellow bone marrow.*
- *In adults,* therefore, haemopoietic (red) bone marrow is confined to the *axial skeleton* (skull, vertebrae, sternum, ribs, sacrum and pelvis) and the *proximal ends of long bones* (humerus, femur and tibia). Differences between red and yellow bone marrow are summarized in Table 3.2-1.

Table 3.2-1 Differentiating features of red and yellow bone marrow

Red bone marrow	Yellow bone marrow
1. Red bone marrow is red in colour and active haemopoietic tissue.	1. Yellow bone marrow is yellowish in colour and inactive.
2. Cellularity is marked. It contains different stages of all types of developing blood cells.	2. Cellularity is very very less and replaced by fatty tissue.
3. Red bone marrow is present in axial skeleton (skull, vertebrae, sacrum, sternum and pelvis) and long bones in children.	3. In adults except ends of long bones and axial skeleton, all other bones contain yellow bone marrow.

BLOOD CELL PRECURSORS

Stem cells

The *monophyletic theory* of haemopoiesis is now widely accepted, according to which all blood cells originate from the *pluripotent* or *multipotent stem cell*. Stem cells possess two *fundamental properties*:

- *Self-replication,* i.e. stem cells are capable of cell division to give rise to more stem cells and
- *Differentiation and commitment,* i.e. the stem cells have ability to differentiate into specialized cells called progenitor cells.

Progenitor cells

The stem cells after a series of divisions differentiate into progenitor cells:

Pluripotent progenitor cells which can give rise to any type of blood cells.

Lymphoid (immune system) stem cells which ultimately develop into lymphocytes.

Myeloid (trilineage) stem cells which later differentiate into three types of cell lines:

- *Granulocyte–monocyte progenitors* which produce all leucocytes except the lymphocytes.
- *Erythroid progenitors* which produce RBCs.
- Megakaryocyte progenitors which produce platelets.

Features of progenitor cells

Progenitor cells possess the ability to give rise to *clones* (group of cells), so they are also called *colony forming cells* or *colony forming units* (CFU). The three types of progenitor cells are given:

- *CFU-GEMM* (colony forming unit–granulocyte, erythroid, megakaryocyte and macrophage) refers to a multipotent progenitor cell, i.e. myeloid progenitor cells.
- *BFU-E* (burst forming unit–erythroid) form large colonies of erythroid series.
- *CFU-E* (colony forming unit–erythroid) develop into erythrocytes.
- *Ba–CFU* refers to the basophil colony forming units.
- *Eo–CFU* are the eosinophil colony forming units.
- *M–CFU* refers to the monocyte colony forming units.
- *G–CFU* are the neutrophil forming units.

The broad outlines of haemopoiesis discussed above are summarized in Fig. 3.2-5. Further details of erythropoiesis are discussed in this chapter and details of development of other blood cells are discussed in the relevant chapters.

Bone marrow examination. The red bone marrow contains the stem cells, progenitor cells, colony forming cells and various types of blood cells in different stages of development, which can be observed in a stained smear of red bone marrow obtained from the iliac bone or sternum (Fig. 3.2-6). Normally, in the red bone marrow the haemopoietic stem cells comprise only 0.01–0.5% of the total bone marrow population. About 75% of the cells are immature white cells and about 25% of the cells are immature red cells, thus forming a ratio of 3:1; while in peripheral blood, ratio of white and red cells is 1:600. This vast difference is because of the fact that the life span of red cells is far greater than that of the white cells.

CONTROL OF HAEMOPOIESIS

The growth of different blood cells from the stem cells is controlled and regulated by the haemopoietic growth factors, which in general are called *cytokines.* Cytokine is a general term used to denote the proteins released by the cells that act as intercellular mediators. The cytokines which control the formation of different types of blood cells are called *colony stimulating factors (CSF)* which are given below:

- *G-CSF* stimulates the granulocytic precursors,
- *M-CSF* stimulates the monocytic precursors,
- *GM-CSF* stimulates both the granulocytic and monocytic precursors.
- *Interleukins (IL)* refer to the cytokine stimulating lymphocytic precursor, for example, IL-1, IL-3, etc.
- *Erythropoietin* refers to the cytokine stimulating the erythroid series of cells.

STAGES OF ERYTHROPOIESIS

The RBCs develop from the burst forming unit–erythrocyte (BFU-E) and colony forming unit–erythrocyte (CFU-E) which are derived from the committed progenitor cells.

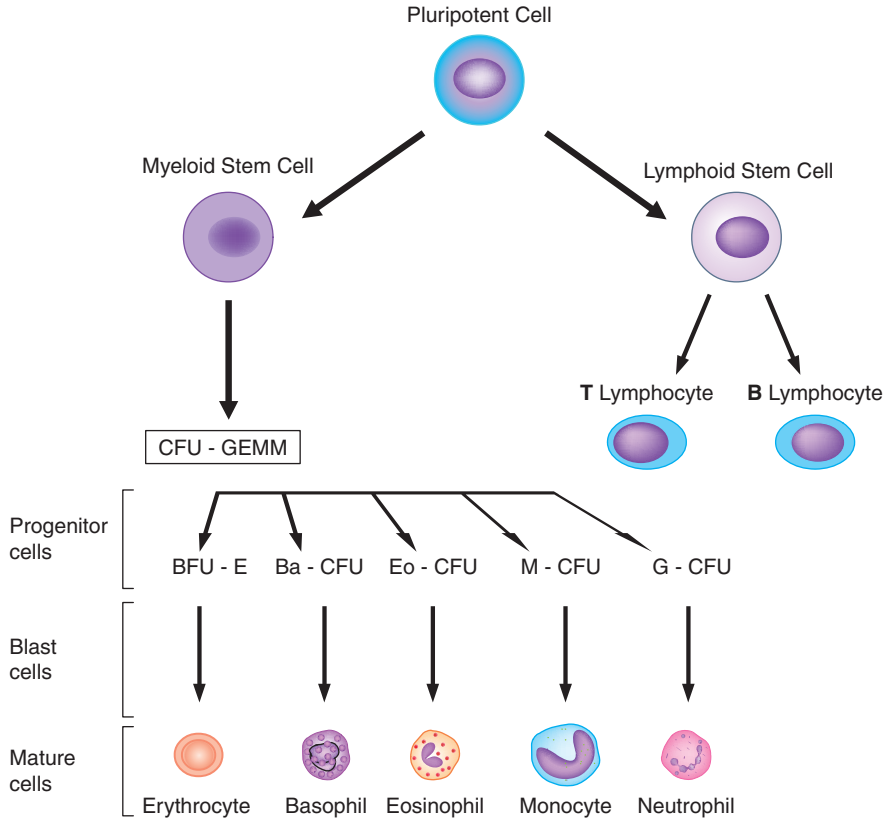


Fig. 3.2-5 Schematic broad outline of haemopoiesis.

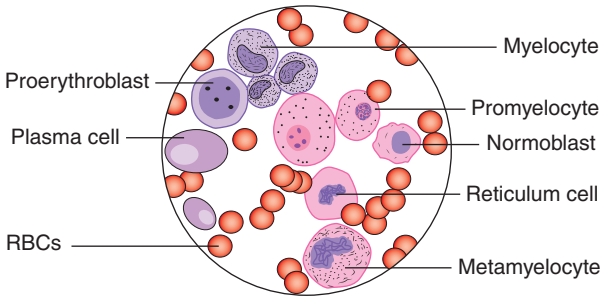


Fig. 3.2-6 Smear from red bone marrow showing various types of cells at different stages of development.

The characteristic features of different stages of erythropoiesis (Fig. 3.2-7) are summarized in Table 3.2-2.

Maturation of a reticulocyte into an erythrocyte

A reticulocyte spends 1–2 days in bone marrow and circulates for 1–2 days in the peripheral blood before maturing in the spleen, to become a biconcave red cell. The reticulocytes are also found normally in the peripheral blood. Normal range of the reticulocytes in healthy adults is 0.5–2% and in infants is 2–6%. Abnormal increase in the circulating reticulocytes is called *reticulocytosis.* This is seen when the rate of erythropoiesis is very high as occurs in the haemolytic anaemia and following the treatment of deficiency anaemias.

The reticulocytes in the peripheral blood are distinguished from the mature red cells by slightly basophilic hue in the cytoplasm similar to that of an orthochromatic normoblast. Reticulocytes can be counted in the laboratory by vital staining with dyes, such as new methylene blue or brilliant cresyl blue.

Summary of changes occurring in the cells of erythroid series during maturation

It takes seven days for the formation and maturation of RBCs. Till the stage of reticulocyte, it takes five days and to become matured red cell from reticulocyte, it takes two days. Various changes which occur during maturation from the stage of pronormoblast to erythrocyte are summarized below:

Size of the cell (from 15–20 µm of pronormoblast) goes on decreasing with subsequent stages till it reaches about 7 µm.

Nucleus first condenses, then becomes pyknotic and finally disappears at the stage of reticulocyte formation.

Hb synthesis starts at the stage of intermediate normoblast and then its content increases progressively.

Cytoplasm staining. Initially, before the appearance of Hb the cytoplasm is basophilic. When Hb starts appearing cytoplasm becomes polychromatic, i.e. stained both by acidic and basic dyes. In the stage of late normoblast when

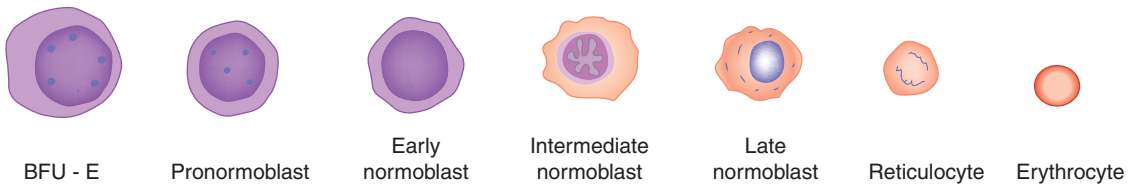


Fig. 3.2-7 Stages of erythropoiesis.

Table 3.2-2 Characteristic features of cells at different stages of erythropoiesis

Stage	Size (μm)	Nucleus	Cytoplasm		Mitosis
			Hb	Staining	
1. Homocytoblast (stem cell)	19–23	• Very large (almost occupying whole of the cell), • deep basophilic • containing 4–5 nucleoli	Absent	Deep basophilic	Present++
2 Pronormoblast (proerythroblast)	15–20	• Large (central) • Deep basophilic • Fine reticular chromatin • 2–3 nucleoli	Absent	Scanty and deep basophilic	++
3. Early normoblast	12–16	• Large • Chromatin strand becomes thicker and coarser • Nucleoli disappears	Absent	Still basophilic	++
4. Intermediate normoblast (polychromatic normoblast)	10–14	• Nucleus becomes condense, coarse and basophilic • Nucleoli absent	Appears	Acidophilic with basophilic hue (polychromatic)	+
5. Late normoblast (orthochromatic normoblast)	8–10	• Nucleus small, pyknotic with dark chromatin (cart-wheel appearance) • Nucleoli absent	Increased in amount	Acidophilic	Absent
6. Reticulocyte	7–7.5	• Nucleus absent • With supravital stain (brilliant cresyl blue) remnants of RNA appears in the form of reticulum in the cytoplasm	Increased in amount	Acidophilic	Absent

Hb synthesis is almost completed, cytoplasm is stained by an acidic dye.

Mitosis is seen up to the stage of intermediate normoblast. During these stages, 3–5 cell divisions occur. In this way, each pronormoblast gives rise to 8–320 late normoblasts. From the stage of late normoblast onwards, the mitosis ceases and cell only matures.

REGULATION OF ERYTHROPOIESIS

Erythropoietin

Erythropoietin is a hormone, which regulates the process of erythropoiesis. It is a glycoprotein having molecular weight of 34,000. It is mainly produced by the juxta glomerular apparatus of kidney. Whenever, there is hypoxia or decrease in the number of RBCs (e.g. after haemorrhage or in haemolytic anaemia), there occurs a release of renal erythropoietic factor from the juxta glomerular cells of kidney. Renal erythropoietic factor acts on the plasma α globulin called erythropoietinogen to form the erythropoietin. Thus the levels of erythropoietin vary with degree of hypoxia or number of circulating RBCs. This explains how polycythaemia (increased RBC count) is observed in the hypoxic states such as in normal individuals residing at high altitude or in the patients suffering from cardiopulmonary disorder.

Actions of erythropoietin. Erythropoietin increases erythropoiesis by acting at the site of erythropoiesis (it may be yolk sac, liver, spleen and bone marrow depending upon the age). It also promotes every stage of maturation from pronormoblast to the mature red cells.

Erythropoietin promotes erythropoiesis because of its following actions:

- Erythropoietin exerts its chief effect on the stem cells causing them to differentiate.
- It promotes Hb synthesis by increasing globlin synthesis and potentiating δ-amino laevulinic acid synthetase.
- It also promotes every stage of maturation from pronormoblast to the mature red cells.
- Erythropoietin also promotes release of RBCs from bone marrow into the peripheral circulation.

Factors increasing erythropoietin secretion. The degree of oxygenation and number of RBCs in circulation act as feedback mechanism to control the secretion of erythropoietin, i.e. depending upon the condition they either increase or decrease erythropoietin secretion to normalize the erythropoiesis. Other factors which increase secretion of erythropoietin are:

1. *Hormones* which increase erythropoietin secretion are the following:
   - *Androgens* (male sex hormones) enhance erythropoietin secretion. This explains the greater RBC count in males as compared to females.
   - *Thyroxine* also promotes erythropoiesis. This explains the occurrence of polycythaemia in hyperthyroidism.
   - *Other hormones* which increases erythropoietin secretion are growth hormone, prolactin, ACTH and adrenocortical steroids.
2. *Haemolysates,* i.e. the products released following RBC destruction, also increase erythropoietin secretion.
3. *Nucleotides* which enhance erythropoietin secretion include cAMP, NAD and NADP.
4. *Vasoconstrictor drugs* produce renal hypoxia, which in turn affects erythropoietin secretion.

Factors decreasing erythropoietin secretion are the following:

- *Adenosine antagonists,* e.g. theophylline and
- *Oestrogen* decreases erythropoietin secretion by:
  - decreasing the synthesis of globin in liver and
  - depressing the erythropoietic response to hypoxia.

FACTORS NECESSARY FOR ERYTHROPOIESIS

Factors necessary for erythropoiesis can be divided in three groups:

- General factors,
- Special maturation factors and
- Factors necessary for haemoglobinization.

I. General factors

The main general factors necessary for the process of erythropoiesis are optimum level of the hormone *erythropoietin* and the efficient *feedback mechanism* controlling the secretion of *erythropoietin* have been discussed in the regulation of erythropoiesis.

II. Special maturation factors

Special factors which are essential for maturation of a RBC include vitamin B_{12}, intrinsic factor of Castle and folic acid.

Vitamin B_{12} and intrinsic factor of Castle

Vitamin B_{12} (cyanocobalamin), also known as extrinsic factor, is essential for maturation of red cells.

Daily requirement of vitamin B_{12} in adults is 1–2 µg. Since its deficiency causes pernicious anaemia, it is also called antipernicious factor.

Role of vitamin B_{12}. It is required for the synthesis of DNA and maturation of nucleus and cell. Its interaction with folic acid is shown in Fig. 3.2-8. Deficiency of vitamin B_{12} leads to:

- Failure of maturation of nucleus.
- Cells remain large (megaloblasts) and become more fragile.
- There occurs reduction in the cell division.

Folic acid

Folic acid (pteroylglutamic acid) and related compounds are known as folates and play an important role in the synthesis of DNA along with vitamin B_{12}.

Daily requirement of folate for a normal healthy adult is 100 µg.

Role of folic acid in DNA synthesis. In the plasma, folate appears as methyl tetrahydrofolate which is changed to tetrahydrofolate (THF) by a pathway for which vitamin B_{12} is essential (Fig. 3.2-8). Without this, active folate co-enzymes are poorly formed. For synthesis of DNA, 5,10 methylene THF is the essential form. Dihydrofolate from this step is reconverted to the THF by dihydrofolate reductase, an enzyme inhibited by the folate antagonist (methotrexate). Formyl THF (folinic acid) will bypass both the metabolic blocks created by vitamin B_{12} deficiency or methotrexate and acts as an antidote to this drug.

Folate deficiency causes megaloblastic anaemia (see page 119).

III. Factors necessary for haemoglobinization

Various factors necessary for Hb formation in the RBCs are described on page 113.

HAEMOGLOBIN

The cytoplasm of erythrocytes (RBCs) contains an *oxygen-binding protein* called haemoglobin. Erythrocyte precursors

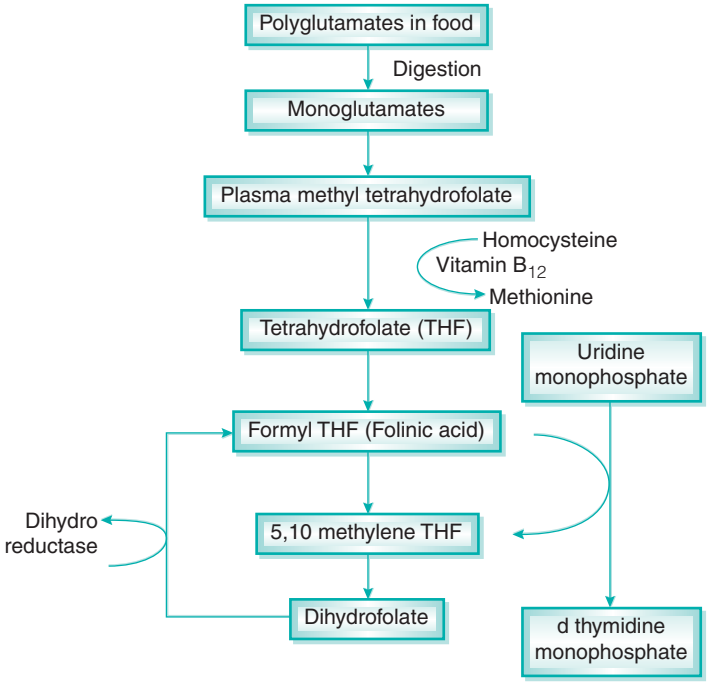


Fig. 3.2-8 Metabolic pathway showing interaction of vitamin B_{12} and folate in the synthesis of DNA.

synthesize Hb; while the mature erythrocytes lose the property of synthesizing Hb. The inclusion of Hb within the erythrocytes is most effective for functional purposes since it avoids the following *disadvantages* which would have occurred if the Hb was present in the plasma as free Hb:

- Increase in blood viscosity causing a rise in blood pressure,
- Increase in the osmotic pressure,
- Rapid destruction of Hb by the reticuloendothelial system and
- Excretion of Hb by kidney (Hb urea).

In pathologic states, e.g. acute haemolytic disorder, Hb appears in the plasma and may lead to above-mentioned consequences.

NORMAL BLOOD HAEMOGLOBIN

The normal blood haemoglobin concentration in adult males is 15.5 g/dL (range 14–18 g/dL) and in adult females the mean Hb concentration is 14 g/dL range 12–15.5 g/dL).

The normal blood Hb concentration at different ages is given below:

- *In fetus*, just before birth, the Hb concentration of blood from the umbilical cord ranges from 16.5 to 18.5 g/dL.
- *After birth*, the Hb concentration increases rapidly and may reach up to 23 g/dL. This occurs due to:
  - the transfusion of cells from the placenta to infant and
  - haemoconcentration by reduction of plasma volume.
- *At the end of 3 months*. After two days of birth, the Hb levels start falling and stabilize at the end of 3 months to 10.5 g/dL.
- *At 1 year of age*. The concentration then rises gradually to reach 12 g/dL at 1 year of age.

IMPORTANT NOTE

- The normal Hb becomes 100% saturated when blood is equilibrated with 100% oxygen (PO_2, 760 mm Hg).
- One gram of Hb when fully saturated combines with 1.34 mL oxygen. Thus Hb concentration is an index of oxygen carrying capacity of blood. Thus, normal values of oxygen carrying capacity in males is 1.34×15.5 = about 21 mL% and in females is 1.34×14 = about 18.5 mL%.
- *Clinically*, irrespective of the age, a level of 14.8 g/dL is considered as 100% Hb.

STRUCTURE OF HAEMOGLOBIN

Haemoglobin is a globular molecule having a molecular weight of 68,000. It consists of the protein *globin* combined with iron containing pigment called *haem.*

Structure of globin

The protein globin, present in the Hb, is made of four polypeptide chains. Haemoglobin A (HbA) consists of the following four polypeptide chains:

- *Two α chains,* each containing 141 amino acid residues and
- *Two β chains,* each containing 146 amino acid residues.

Therefore, the normal adult haemoglobin A is written as HbA ($\alpha_2\beta_2$).

Structure of haem

The haem is an iron–porphyrin complex called *iron–protoporphyrin IX,* i.e. it consists of a porphyrin nucleus and the iron. The structural characteristics of the haem (iron–protoporphyrin IX) are given below (Fig. 3.2-9):

Porphyrin nucleus

- The porphyrin nucleus consists of four *pyrrole rings* numbered I, II, III and IV, i.e. porphyrins are tetrapyrroles.
- The pyrrole rings are joined together by four *methine bridges* (=CH–). The carbon atoms of methine bridges are labelled α, β, γ and δ.
- Eight *side chains* are attached to the pyrrole ring at positions labelled 1–8. These are:
  - Four *methyl* (H_3C) side chains at position 1, 3, 5 and 8.
  - Two *vinyl* (–CH ═ CH_2) side chains at position 2 and 4.
  - Two propionic acid (–CH_2 ═ CH_2 ═ COOH) side chains at position 6 and 7.

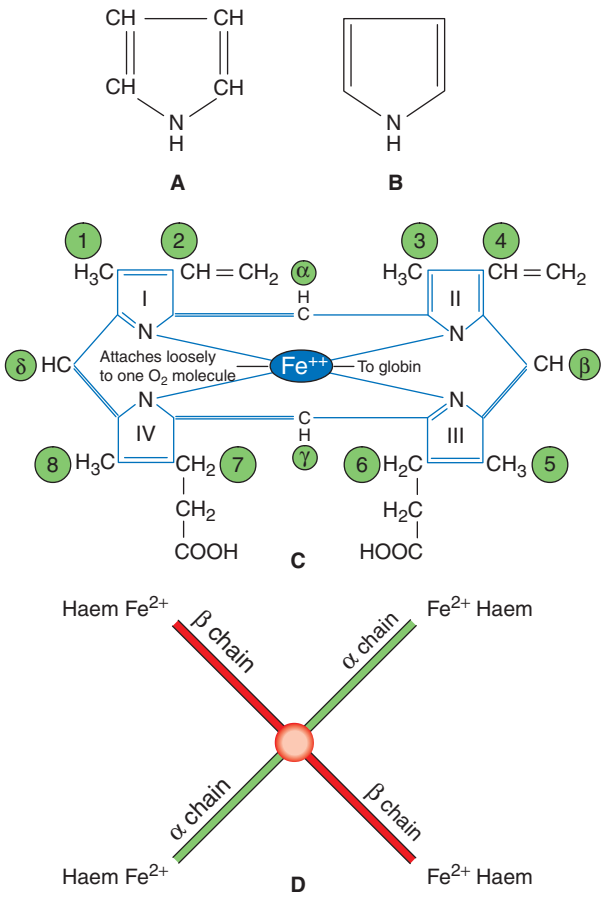


Fig. 3.2-9 Chemistry of haemoglobin: A, structure of a pyrrole ring; B, conventional outline of a pyrrole ring; C, arrangement of pyrrole rings in one unit of haem (iron protoporphyrin IX) and D, arrangement of four units of haem in one molecule of haemoglobin.

The iron

- The iron in the haem is in *ferrous* (Fe^{2+}) form.
- The iron is attached to the nitrogen atom of each pyrrole ring.
- On the iron (Fe^{2+}) a bond is available for loose union, where:
- In oxyhaemoglobin, O_2 is attached,
- In carboxyhaemoglobin, CO is attached, and so on (see derivatives of Hb).

Attachment of haem to globin

One molecule of Hb contains four units of haem, each attached to one of the four polypeptide chains constituting globin (Fig. 3.2-9D). As there are four units of haem in one molecule of Hb, so there are four iron atoms in one molecule of Hb which can carry four molecules (eight atoms) of oxygen.

FUNCTIONS OF HAEMOGLOBIN

1. Transport of O_2 from lungs to tissues

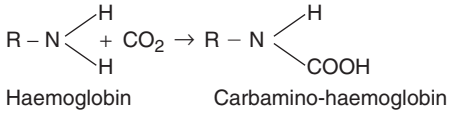
- *In the lungs,* one molecule of O_2 is attached loosely and reversibly at the sixth covalent bond of each iron atom of the Hb to form *oxyhaemoglobin* represented as HbO_2:

Hb	+	O_2	→	HbO_2
Deoxygenated (reduced) haemoglobin				Oxygenated haemoglobin

- Oxygenation of first haem molecule in the Hb increases the affinity of second haem for oxygen which in turn increases the affinity of third haem and so on. In this way, the affinity of Hb for fourth oxygen molecule is many times that for the first molecule.
- The affinity of Hb for oxygen is influenced by pH, temperature and concentration of 2,3-diphosphoglycerate, i.e. 2,3-DPG (a product of metabolism of glucose) in the RBCs.

2. Transport of CO_2 from the tissues to the lungs

Hb also transports CO_2 from the tissues to the lungs. It is important to note that the CO_2 from the tissues is transported by combining with amino acids of the globin part as shown below and not in combination with Fe^{2+} atom like O_2.



Deoxygenated Hb forms carbamino-haemoglobin more readily than oxygenated Hb. That is why venous blood

becomes more suitable for the transport of CO_2 from the tissues to the lungs.

3. Control pH of the blood

The Hb constitutes the most important acid–base buffer system of blood. Hb has six times the buffering capacity as compared to the plasma proteins.

VARIETIES OF HAEMOGLOBIN

Various varieties of Hb can be grouped as under:

- Physiological varieties of Hb and
- Haemoglobinopathies.

Physiological varieties of haemoglobin

Adult haemoglobin or haemoglobin A [HbA ($\alpha_2\beta_2$)]. (see page 109). Adult Hb is of two types.

(i) Haemoglobin A [HbA ($\alpha_2\beta_2$)]. It is the main form of normal adult Hb. As described on page 109, its globin part consists of two α and two β polypeptide chains. It is a spheroidal molecule with a molecular weight of 68,000.

(ii) Haemoglobin A_2 [HbA_2 ($\alpha_2\delta_2$)]. It is a minor component (about 2.5% of the total Hb) in normal adults. Its globin part consists of two α and two δ polypeptide chains. δ chains have slightly different amino acid composition (out of 146, 10 amino acids are different) as compared to β chains.

Fetal haemoglobin or haemoglobin F [HbF ($\alpha_2\gamma_2$)] as the name indicates refers to the Hb present in the fetal RBCs and gradually disappears 2–3 months after birth.

Amount of HbF and HbA present at various stages is as given in Table 3.2-3.

Structure of HbF is similar to that of HbA, except that its globin part consists of two α and two γ polypeptide chains (in place of β chains). γ chains also have 146 amino acids but its 37 amino acids are different than that of β chains.

Special features of HbF are given below:

- *Affinity for oxygen* in case of HbF is more than that of HbA, i.e., it can take more oxygen than HbA at low oxygen pressure. It is owing to poor binding of 2,3-DPG by the γ polypeptide chain. Because of this, movement of oxygen from maternal to fetal circulation is facilitated.
- *Resistance to action of alkalies* is more in HbF than HbA. This property is used in a photoelectric calorimetric method to estimate HbF in the presence of HbA.
- *Life span* of HbF is much less (1–2 week) as compared to that of HbA (120 days).

Table 3.2-3 Amount of HbF and HbA at various stages in human beings

Stage	HbF (%)	HbA (%)
• At 20 weeks of intrauterine	94	6
• At birth	80	20
• At 2 months after birth	50	50
• At 4 months after birth	10	90
• At more than 1 yr after birth	<1	>99

Haemoglobinopathies

Haemoglobinopathies, i.e. abnormal formation of haemoglobin occurs due to the disorders of globin synthesis; haem synthesis being normal. Disorders of the globin synthesis are of two main types:

- *Formation of abnormal polypeptide chains* due to substitution of an abnormal amino acid chain in the HbA. Example of such a disorder is haemoglobin S.
- *Suppression* of synthesis of polypeptide chain of globin as seen in thalassaemia.

1. ***Sickle cell haemoglobin*** or haemoglobin S (HbS) is the most important haemoglobinopathy

- It occurs in 10–20% of Negroes. Sickle cell gene has originated in the black population in Africa.
- HbS is formed due to substitution of valine for glutamic acid at position 6 in the β chain of HbA.When HbS is reduced (e.g. in low O_2 tension or when pH at tissue level is low), it becomes much less soluble and precipitates into crystals within the RBCs. The crystals elongate producing changes in shape of the cells from biconcave to sickle-shaped cells (sickling) (Fig. 3.2-10).
- The cells containing HbS are less flexible as compared to the RBCs containing HbA, hence leading to a blockade of microcirculation.
- Sickle-shaped cells greatly increase blood viscosity thereby decreasing the blood flow to tissues.
- Sickle-shaped cells are more fragile and are very liable to undergo haemolysis producing the so-called sickle cell anaemia. Sickle trait is inherited as Mendelian dominant but the full blown disease is autosomally recessive. Heterozygous individual with sickle cell trait rarely has

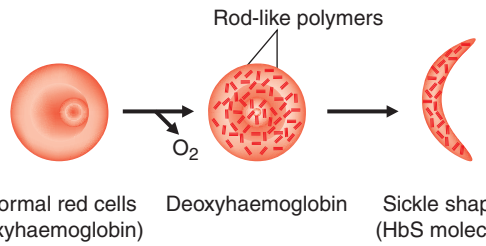


Fig. 3.2-10 Mechanism of sickling of red blood cell containing haemoglobin S (HbS).

Table 3.2-4 Main differentiating features of β thalassaemia—major and minor

S. No.	β thalassaemia major	β thalassaemia minor
1.	Thalassaemia major is also called as Mediterranean anaemia or Cooley's anaemia and is less common.	Thalassaemia minor is more common.
2.	It is inherited as a homozygous transmission (i.e. abnormal genes are inherited from both the parents) therefore: • There is complete absence of β chain synthesis. • Absence of β chain synthesis results in moderate to severe anaemia. • HbF level is markedly increased.	It is inherited as a heterozygous transmission (i.e. abnormal gene is inherited from one parent), therefore: • The synthesis of β chain is not completely absent (partial). • Anaemia is of mild type. • HbF level is either normal or slightly elevated.
3.	The individual suffering from thalassaemia major has short life span, i.e. (dies young 17–18 years).	The individuals suffering from thalassaemia minor comparatively survive longer (up to adult) and transmit abnormal gene to their offsprings.

severe symptoms but homozygous develop full blown disease.

- The individual with sickle cell trait has resistance to one type of malaria.

2. ***Thalassaemia*** (Mediterranean anaemia) is a haemoglobinopathy characterized by following features:

- *Cause.* Thalassaemia results due to defect in the synthesis of polypeptide chain α and β of HbA.
- *Types.* Depending upon whether α or β chains are not synthesized, α thalassaemia or β thalassaemia may occur, respectively. β Thalassaemia is more common and is further of two types: thalassaemia major and thalassaemia minor.

Differentiating features of thalassaemia major and minor are depicted in Table 3.2-4.

DERIVATIVES OF HAEMOGLOBIN (REACTIONS OF HAEMOGLOBIN)

Haemoglobin has the property to readily react with any gas, other substance to form the so-called derivatives of haemoglobin. These include:

1. ***Oxyhaemoglobin.*** Haemoglobin reacts readily with oxygen to form oxyhaemoglobin which is an unstable and reversible compound, i.e. oxygen can be released from this compound. In this compound iron remains in the ferrous state.
2. ***Reduced haemoglobin*** or deoxygenated haemoglobin is formed when oxygen is released from the oxyhaemoglobin.

$$\underset{\text{(Oxyhaemoglobin)}}{HbO_2} \rightarrow \underset{\text{(Reduced haemoglobin)}}{Hb} + O_2$$

3. ***Carbamino-haemoglobin*** is a compound of Hb with carbon dioxide

$$HbNH_2 + CO_2 \rightarrow HbNHCOOH$$

4. ***Carboxyhaemoglobin*** or carbon monoxyhaemoglobin is a compound of Hb with carbon monoxide (CO)

$$Hb + CO \rightarrow COHb$$

The affinity of Hb for CO is much more (200–250 times) than its affinity for oxygen. Because of this, the CO displaces oxygen from Hb, thereby reducing the oxygen carrying capacity of the blood.

5. ***Methaemoglobin.*** When reduced or oxygenated Hb is treated with an oxidizing agent, e.g. potassium ferricyanide, the ferrous Fe^{2+} is oxidized to ferric (Fe^{3+}); the sixth bond is attached to OH to form the compound methaemoglobin. Methaemoglobin is represented as HbOH.
Disadvantages of methaemoglobin are:
   - It cannot unite reversibly with gaseous oxygen; the O_2 of the attached OH is not given off in a vacuum.
6. ***Glycosylated haemoglobin*** is a derivative of haemoglobin A present in very small amount, e.g. haemoglobin A_{1C} (HbA_{1C}), in which glucose is attached to terminal valine in the β chains. The level of glycosylated haemoglobin in the blood increases in poorly controlled patients of diabetes mellitus.

SYNTHESIS OF HAEMOGLOBIN

Haemoglobin is synthesized in the cytoplasm of intermediate normoblasts.

Synthesis of haem

Haem is synthesized in the mitochondria. Steps of synthesis are (Fig. 3.2-11):

- Succinyl-CoA (derived from the citric acid cycle in mitochondria) and glycine are the starting substances in the synthesis of haem. These condense to form α-amino-β-ketoadipic acid. The condensation requires pyridoxal phosphate for activation of glycine.

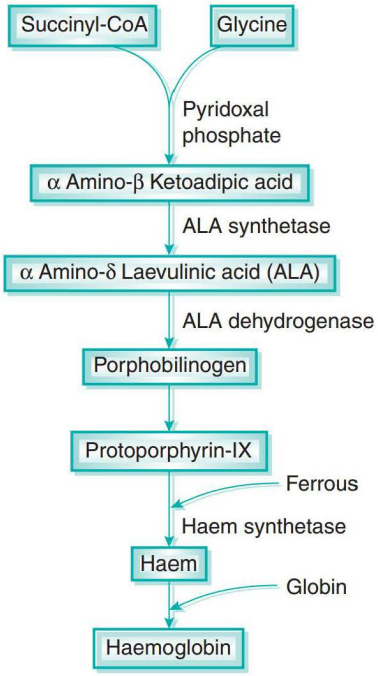


Fig. 3.2-11 Steps of synthesis of haemoglobin.

- The *protoporphyrin IX* is then formed after a series of reactions promoted by other enzymes.
- Finally, ferrous ion is introduced into the protoporphyrin IX molecule to form *haem* in a reaction catalyzed by the enzyme haem synthetase.

Synthesis of globin

Globin, the protein part of the Hb, is synthesized in the ribosomes.

Factors controlling haemoglobin formation

1. Role of proteins. First class proteins provide amino acids required for the synthesis of globin part of the Hb. A low protein intake retards Hb regeneration even in the presence of excess iron; the limiting factor being lack of globin.

2. Role of iron. Iron is necessary for formation of the haem part of haemoglobin. In addition to dietary iron, the iron released by degradation of RBCs is also reused for the synthesis of Hb.

3. Role of other metals like:

- *Copper* is essential for the Hb synthesis, as it promotes the absorption, mobilization and utilization of iron.
- *Cobalt* increases the production of erythropoietin which in turn stimulates RBC formation.
- *Calcium* reported to help indirectly by conserving iron and its subsequent utilization.

4. Role of vitamins. Vitamin B_{12}, folic acid, and vitamin C help in synthesis of nucleic acid which in turn is required for the development of RBCs. Vitamin C also helps in absorption of iron from the gut.

5. Role of bile salts. Presence of bile salts in the intestine is necessary for proper absorption of metals like copper and nickel which in turn are essential factors for synthesis of Hb.

RED CELL FRAGILITY

Red cell fragility refers to the susceptibility of red cell membrane to get broken or bursted. The process of breaking of RBCs and release of Hb into the plasma is called haemolysis. Red cell fragility depending upon the underlying mechanism is of two types:

- Osmotic red cell fragility and
- Mechanical red cell fragility.

OSMOTIC RED CELL FRAGILITY

Osmotic red cell fragility refers to the susceptibility of red cell membrane to get lysed due to changes in the osmotic pressure of the solution in which they are suspended (see page 20).

Normal values (index of fragility)

- *Onset of haemolysis* (fragility) in normal RBCs occurs in 0.48% NaCl and
- *Completion of haemolysis* (ending of fragility) occurs in 0.35% NaCl.

Abnormal osmotic fragility

Increase in osmotic fragility index occurs in following conditions:

- *Congential spherocytosis*, i.e. when the RBCs are spherical. In this condition onset of haemolysis occurs at 0.7% NaCl and it is completed at 0.45% NaCl solution.
- *Autoimmune haemolytic anaemia* in which the autoantibodies damage the structure proteins and render the red cells more fragile.
- *Deficiency of glucose 6-phosphate dehydrogenase (G6PD)* increases the tendency of red cells to get haemolysed by antimalarial drugs and other agents.
- *Venom of cobra and other insects* contains lecithinase which dissolves lecithin from the red cell membranes making them more fragile.

Decrease in osmotic fragility index occurs when the RBCs become slender, e.g. in iron deficiency anaemia. In this condition the onset of haemolysis occurs at 0.36% NaCl and is completed at 0.24% NaCl solution.

MECHANICAL RED CELL FRAGILITY

The red cells are subjected to a mechanical stress and trauma as they pass through the capillaries and trabeculae of spleen

some 300,000 times during their life span of 120 days. They are made more brittle due to unusual mechanical stress. The red cells can become more rigid as a result of the pathological changes in the membrane or in the cell contents caused by a number of red cell disorders. The cells thus become mechanically more fragile, i.e. less liable to tolerate deforming stresses than is the normal healthy red cell.

LIFE SPAN AND FATE OF RED BLOOD CELLS

LIFE SPAN OF RBCs

Normally, the average life span of RBCs is 120 days.

Causes of reduction in the life span of RBCs

I. Defects in RBCs (Corpuscular defects)

- Hereditary spherocytosis,
- Sickle cell anaemia,
- Thalassaemias,
- Deficiency of red cell enzymes,
- Glucose 6-phosphate-dehydrogenase deficiency and
- Pyruvate kinase deficiency.

II. Extracorpuscular defects

- Transfusion of mismatched blood,
- Autoimmune haemolytic disorders and
- Hypersplenism.

FATE OF RBCs

The cell membrane of old RBCs (after about 120 days) becomes more fragile due to decreased NADPH activity. The destruction of red cells occurs mostly in the capillaries of spleen because they have very thin lumen. Because of this, spleen is also called the graveyard of RBCs. The haemoglobin released after the haemolysis of red cells is taken up by the tissue macrophages.

The tissue macrophage system (reticuloendothelial system) includes the following phagocytic cells:

- *In the bone marrow* these cells form part of the lining of the blood sinuses (littoral cells),
- *In the liver* they lie at intervals along the vascular capillaries (Kupffer cells),
- *In the spleen* they are found in the pulp and
- *In the lymph nodes* they line the lymphatic paths.

Fate of haemoglobin (Fig. 3.2-12)

- In the macrophages, the haem part of the haemoglobin molecule is altered by oxidation of one of its methine (=CH) bridges. The tetrapyrrole ring structure is thus broken and four pyrrole groups become arranged as a straight chain. As a result of this chemical change, the green iron-containing compound *choleglobin is formed.* As the name implies, the choleglobin molecule still contains the original globin.
- Next, the *choleglobin splits off* into globin, iron and biliverdin (tetrapyrrole straight chain free from globin and iron).
- *Globin* is degraded into amino acids and joins the amino acid pool of plasma and is released.
- *Iron* released into the circulation is:
  - carried into the bone marrow for reutilization
  - in the other tissues it combines with apoferritin to form the ferritin (storage form of iron).
- *Biliverdin* (tetrapyrrole straight chain free from globin and iron) is converted into bilirubin (by the enzyme biliverdin reductase) and is released into the blood.

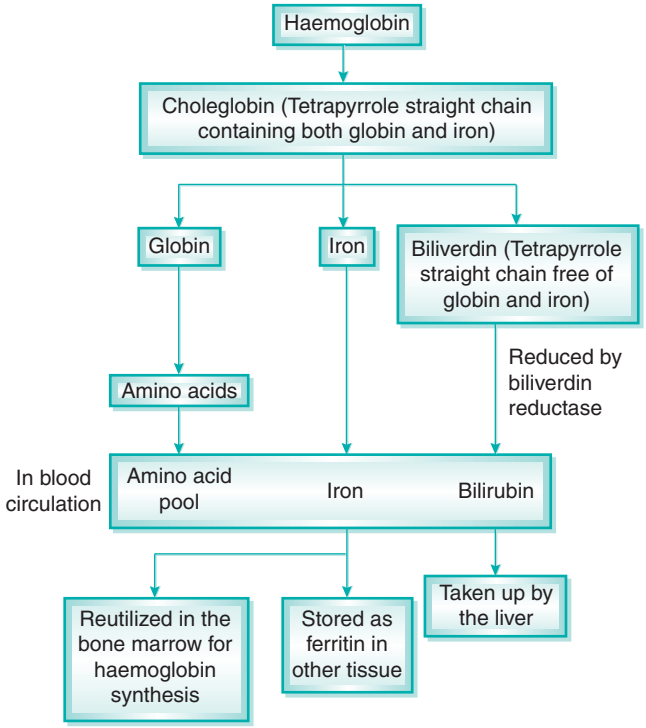


Fig. 3.2-12 Fate of haemoglobin.

BILIRUBIN AND JAUNDICE

BILIRUBIN FORMATION AND ITS FATE

As discussed above, the bilirubin is formed in the macrophages. It undergoes the following changes (Fig. 3.2-13):

1. Uptake of bilirubin. Macrophages release the bilirubin into circulation. This bilirubin is called free or *unconjugated bilirubin.* It is lipid soluble, in the plasma it is bound to the albumin (protein conjugated) which prevents its excretion by the kidneys.

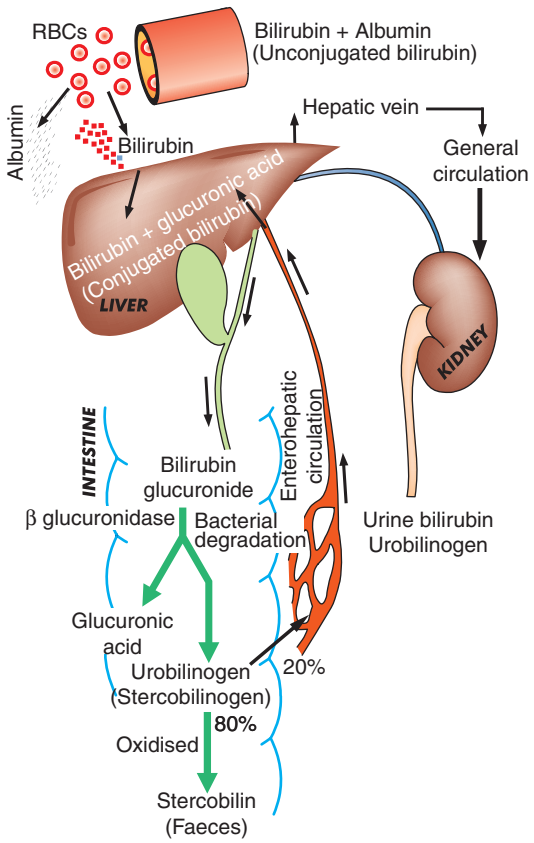


Fig. 3.2-13 State of bilirubin in the body.

2. Conjugation of bilirubin. The unconjugated bilirubin (bound to albumin) from the circulation is taken up by the liver. In the liver the bilirubin is split off from the albumin and enters the hepatic cells. In the hepatic cells, it is conjugated with uridine diphosphate glucuronic acid (UDP-glucuronic acid) making it a water soluble conjugated *bilirubin.* The reaction is catalyzed by the enzyme *glucuronyl transferase* present in the hepatic microsomes (smooth endoplasmic reticulum of liver cells). The reaction occurs in two stages:

- Bilirubin + UDP – glucuronic acid $\xrightarrow{\text{UDP–glucuronyl transferase}}$ Bilirubin monoglucuronide + UDP
- Bilirubin monoglucuronide + UDP – glucuronic acid $\xrightarrow{\text{UDP–glucuronyl transferase}}$ Bilirubin diglucuronide + UDP

3. Excretion of bilirubin. The conjugated bilirubin from the hepatic cells is excreted into the bile and enters the intestine. Some of it escapes into general circulation and is excreted by the kidneys in urine as *urine bilirubin.*

4. Formation and excretion of urobilinogen. The conjugated bilirubin which enters the intestine with the bile is degraded by the intestinal bacteria in the terminal ileum and the large intestine. The bacterial enzyme β-glucuronidase splits off the glucuronide and converts bilirubin into the *urobilinogen* (sterco-bilinogen) which is a colourless compound.

- Some urobilinogen (20%) from the intestine is reabsorbed and goes via the portal system to the liver. From the liver some urobilinogen escapes into general circulation and some are re-excreted into the bile (enterohepatic circulation).
- From general circulation, the urobilinogen is filtered off by the kidney and is excreted in the urine. Remaining 80% of stercobilinogen in the intestine (which is not absorbed) is excreted in the faecal matter (amount varies from 20–250 mg/day). This stercobilinogen is oxidised to stercobilin which imparts brown colour to faeces.

BILIRUBIN

The normal serum bilirubin level ranges from 0.3 to 1.0 mg/dL. The total serum bilirubin includes conjugated as well as unconjugated bilirubin. The Van den Bergh test described below is helpful in determining the type of bilirubin present in the serum.

Van den Bergh test

Van den Bergh test is performed using the diazo reagent (mixture of sulphanilic acid, hydrochloric acid and sodium nitrite). It is of two types:

Direct Van den Bergh reaction. When diazo reagent is added to the serum containing conjugated bilirubin (water soluble) a reddish brown colouration is obtained within 30 s. This is called direct positive Van den Bergh reaction.

Indirect Van den Bergh reaction. When diazo reagent is added to the serum containing mainly unconjugated bilirubin (water insoluble), no colour is obtained. However, if some solvent like alcohol (which dissolves the unconjugated bilirubin) is added, the reddish brown colouration is obtained. This is called indirect positive Van den Bergh reaction.

JAUNDICE

Jaundice (icterus) refers to the yellow appearance of the skin, sclera and mucous membranes resulting from an increased bilirubin concentration *(hyperbilirubinaemia)* in the body fluids. Clinically, jaundice is detectable when the plasma bilirubin exceeds 2–3 mg/dL.

Mechanisms producing jaundice and types

Hyperbilirubinaemia producing jaundice can result from the following mechanisms:

1. *Excessive breakdown (haemolysis) of RBCs* produces the so-called *haemolytic jaundice or pre-hepatic jaundice.*

Table 3.2-5 Characteristic features of three types of jaundice

Haemolytic jaundice (Pre-hepatic jaundice)	Hepatocellular jaundice (Hepatic jaundice)	Cholestatic or obstructive jaundice (Post-hepatic jaundice)
1. *Mechanism of production*		
Excessive breakdown of RBCs producing unconjugated bilirubin in amounts more than the healthy liver can conjugate and excrete.	*Inability of the liver* to efficiently conjugate as well as transport bilirubin into the bile due to the liver cell damage caused by some infective or toxic agent.	*Obstruction to the bile flow* due to any cause from hepatocytes to duodenum.
2. *Types of serum bilirubin accumulated*		
Unconjugated hyperbilirubinaemia occurs since it is being produced in excess of what can be conjugated by the liver.	*Both unconjugated* as well as *conjugated* bilirubin is increased in serum.	*Conjugated hyperbilirubinaemia* results due to impaired flow of bile.
3. *Van den Bergh test*		
Indirect positive reaction (Because unconjugated bilirubin is present in blood).	*Biphasic reaction* (Because conjugated and unconjugated bilirubin are present).	*Direct positive reaction* (Because only conjugated bilirubin is present).
4. *Urine bilirubin*		
Absent (Unconjugated bilirubin is insoluble in water. It is transported in plasma in bound form with albumin. Since albumin is not filtered into urine, unconjugated bilirubin too is not filtered in urine. Because of this haemolytic jaundice is also called *acholuric jaundice*—no bile pigment in urine).	***Present*** (Conjugated bilirubin is water soluble and is present in the plasma in dissolved form. It gets easily filtered in urine. Such a jaundice is also called *choluric jaundice*, i.e. bile pigment present in urine).	***Present*** (Since conjugated bilirubin is filtered in urine).
5. *Urine urobilinogen*		
Increases (Because liver is excreting lot of conjugated bilirubin in the intestine with the bile. So, more urobilinogen is formed. Part of it reabsorbed and goes to general circulation and thus urine urobilinogen is increased).	***Decreases*** (Because, damaged liver cells are producing and excreting less of conjugated bilirubin and thus less urobilinogen is formed).	***Markedly decreased or absent*** (Because of obstruction the conjugated bilirubin is not released into the intestine and thus no urobilinogen is formed).
6. *Faecal stercobilinogen* (Normal 25–250 mg/day)		
• *Markedly increased* (Because of more formation as described above). • So faeces is *dark brown* in colour	• *Reduced* (Because of less formation as described above). • So stools are *pale* in colour	• *Absent* (When obstruction is complete). • So stools are clay coloured.
7. *Faecal fat level*		
Normal, i.e. 5–6% of total intake/day (As bile is present in gut for normal digestion of fats).	*Increased* up to 40–50% (Because of deficiency of bile in the intestine, emulsification and absorption of fat is inadequate. This produces bulky, pale, greasy and foul smelling faeces called steatorrhoea).	*Increased*
8. *Specific blood tests*		
• *Peripheral blood film* shows signs of haemolysis, i.e. anaemia, reticulocytosis and abnormal RBCs.	Normal	Normal
• *Plasma albumin, globulin and A/G ratio* Normal.	Albumin is decreased due to less synthesis by damaged liver globulin increases A/G ratio decreases.	Normal
• *Serum alkaline phosphatase* Normal, i.e., 5–13 KA units/100 mL (Because excreted in bile)	Increased (Because less excretion in bile)	Markedly increased (Because not excreted in bile)
• *Liver function tests* *Normal* (As liver is healthy)	*Impaired* (As liver is damaged)	*Normal* or mildly impaired

2. *Damage to the liver cells* (infective or toxic) produces the so-called *hepatic or hepatocellular jaundice.*
3. *Obstruction to bile ducts* produces the obstructive or *post-hepatic or cholestatic jaundice.*

Characteristic features of jaundice

Characteristic features of these three varieties of jaundice are described in Table 3.2-5.

Physiological jaundice of newborn

Physiological jaundice of newborn is also called neonatal jaundice.

Mechanism of production. A hyperbilirubinaemia producing jaundice may be seen normally in the newborn. It appears within 2–5 days of birth and usually disappears in 2 weeks. Its mechanism of production includes:

- *Excessive destruction of RBCs* occurs in first few days after birth causing increase in the serum bilirubin.
- *Hepatic immaturity* in the first few (7–10) days after birth also contributes to increased serum bilirubin. In the foetus, bilirubin is removed from circulation by the placenta. Immediately after birth, liver has to take up this work which takes 7–10 days to get mature and fully conjugate the bilirubin.

Prevention and treatment

Prevention. Neonatal jaundice can be prevented by the administration of hepatic microsomal enzyme inducers (e.g. phenobarbital) to the pregnant mother or newborn. The microsomal enzyme inducers increase the activity of glucuronyl transferase in liver.

Treatment. Neonatal jaundice can be effectively treated by *phototherapy.* Exposure of the skin to white light causes photoisomerization of bilirubin to water-soluble *lumirubin* which can be rapidly excreted in bile without requiring any conjugation.

ANAEMIAS

DEFINITION AND CLASSIFICATION

DEFINITION

Anaemia is not a single disease but a group of disorders in which Hb concentration of blood is below the normal range for the age and sex of the subject. Therefore anaemia is labelled when the Hb concentration is less than:

- 13 g/dL in adult males,
- 11.5 g/dL in adult females,
- 15 g/dL in newborn, and
- 9.5 g/dL at 3 months of age.

Low RBC count (less than 4 million/μL) is usually, but not always associated with low Hb levels in anaemia.

Grading of anaemia depending upon the level of Hb, has somewhat arbitrarily been made as:

- Mild anaemia – Hb 8–10 g/dL,
- Moderate anaemia – Hb 6–8 g/dL and
- Severe anaemia – Hb below 6 g/dL.

CLASSIFICATION

Aetiological (Whitby's) classification

Types of anaemia depending upon the causative mechanism are:

A. Deficiency anaemias
- Iron deficiency anaemia
- Megaloblastic anaemia (pernicious anaemia) due to deficiency of vitamin B_{12}
- Megaloblastic anaemia due to deficiency of folic acid
- Protein and vitamin C deficiency can also cause anaemia.

B. Blood loss anaemias or haemorrhagic anaemias are commonly known and can be:

- *Acute post-haemorrhagic anaemia* as in accidents and
- *Chronic post-haemorrhagic anaemia.*

C. Haemolytic anaemias. These are relatively uncommon and occur in conditions associated with increased destruction of RBCs. These can be:

1. *Hereditary haemolytic anaemias,* e.g. as seen in:
   - Thalassaemia,
   - Sickle cell anaemia,
   - Hereditary spherocytosis and
   - Glucose 6-phosphate dehydrogenase (G6PD) deficiency.
2. *Acquired haemolytic anaemias* such as Immunohaemolytic anaemia (due to antibodies against RBCs),
   - Haemolytic anaemia due to direct toxic effects (e.g. in malaria, snake venom, toxic effects of drugs and chemicals, etc.),
   - Haemolytic anaemia in splenomegaly and
   - Haemolytic anaemia in paroxysmal nocturnal haemoglobinuria.

D. Aplastic anaemia. It occurs due to the failure of bone marrow to produce RBCs.

E. Anaemia due to chronic diseases. It is seen in tuberculosis, chronic infections, malignancies, chronic lung diseases, etc.

Morphological (Wintrobe's) classification

Based on the mean cell volume (MCV), i.e. cell size and the mean corpuscular haemoglobin concentration (MCHC),

i.e. haemoglobin saturation of RBCs, the anaemias can be classified as:

1. Normocytic normochromic anaemias. These are characterized by normal MCV (78–94 μm^3 or 78–94 μL) and normal MCHC (30–38%). Such a morphological picture is seen in:

- Acute post-haemorrhagic anaemia,
- Haemolytic anaemias and
- Aplastic anaemias.

2. Microcytic hypochromic anaemias. These are characterized by reduced MCV (<78 μm^3) and reduced MCHC (<30%). Examples of such anaemias are:

- Iron deficiency anaemia,
- Chronic post-haemorrhagic anaemia and
- Thalassaemia.

3. Macrocytic normochromic anaemia. It is characterized by increased MCV (>94 μm^3) and normal MCHC (30–38%). Examples are:

- Megaloblastic anaemia (pernicious anaemia) due to deficiency of vitamin B_{12} and
- Megaloblastic anaemia due to deficiency of folic acid.

GENERAL CLINICAL FEATURES OF ANAEMIA

Anaemic hypoxia results due to decreased O_2 carrying capacity of blood in anaemia owing to reduced Hb concentration. The hypoxia brings about several cardiorespiratory compensatory responses (see page 354). So, general clinical features (symptoms and signs) in patients with anaemia are due to those caused by:

- Resulting tissue hypoxia and
- Resulting compensatory mechanisms.

General clinical manifestations of anaemia which occur either due to tissue hypoxia or due to compensatory mechanisms are:

- *Generalized muscular weakness,* tiredness and easy fatiguability occur due to muscle hypoxia.
- *Pallorness of skin and mucous membranes* (buccal and pharyngeal mucous membrane, conjunctiva, lips, ear lobes, palm and nail bed) occurs due to the deficiency of red coloured Hb in the blood.
- *Respiratory symptoms* such as breathlessness with increased rate and force of respiration occur due to compensatory stimulation of respiratory centre.
- *Cardiovascular manifestations,* such as palpitation, tachycardia and cardiac murmurs occur as a result of compensatory mechanisms increasing the cardiac output. In very severe cases of anaemia, features of cardiac failure, angina pectoris may also occur.
- *Central nervous system manifestations* due to cerebral hypoxia include lethargy, headache, faintness, especially on exertion, tinnitus, restlessness, confusion and drowsiness.
- *Ocular manifestations* include visual disturbances and retinal haemorrhages and cotton wool spots.
- *Gastrointestinal system symptoms* include anorexia, flatulence, nausea, constipation. In pernicious anaemia, there occurs atrophy of papillae on tongue.
- *Reproductive system* involvement occurs in females in the form of the menstrual disturbances such as amenorrhoea and menorrhagia and loss of libido.
- *Renal system* involvement may occur in severe anaemia causing disturbances of renal function and albumin urea.
- *Basal metabolic rate* is increased in severe anaemia.

IRON DEFICIENCY ANAEMIA

Iron deficiency anaemia is the commonest nutritional deficiency disorder present throughout the world, but its prevalence is higher in the developing countries. In India, iron deficiency is the commonest cause of anaemia. Iron deficiency anaemia is much more common:

- In women between 20–45 years than in men,
- At periods of active growth in infancy, childhood and adolescence.

Daily requirement and dietary sources of iron

Daily requirement. Only 10% of the dietary intake of iron is absorbed. Therefore, daily requirement in the adult males is 5–10 mg/day and in females is 20 mg/day (to compensate the menstrual loss). Pregnant and lactating women require about 40 mg of iron per day.

Dietary sources. Foodstuffs vary both in their iron content and availability of iron for absorption into the body. The dietary sources of iron are meat, liver, egg, leafy vegetables, whole wheat and jaggery. The iron in foods of animal origin is better absorbed than iron in foods of vegetable origin.

CAUSES OF IRON DEFICIENCY ANAEMIA

Causes of iron deficiency vary with age, sex and country of residence of patient. In general, the causes of iron deficiency anaemia can be grouped as:

1. ***Inadequate dietary intake of iron as in:***
   - Milk fed infants,
   - Poor economic status individuals,
   - Anorexia, e.g. in pregnancy and
   - Elderly individuals due to atrophy and poor dentition.
2. ***Increased loss of iron (as blood loss) from the body,*** e.g.
   - *Uterine bleeding in females* in the form of excessive menstruation, repeated miscarriages, postmenopausal bleeding, etc.

3. ***Increased demand of iron as in:***
   - Infancy, childhood and adolescence,
   - Menstruating females and
   - Pregnant females.
4. ***Decreased absorption of iron, as seen in:***
   - Partial or total gastrectomy,
   - Achlorhydria and
   - Intestinal malabsorption diseases.

CLINICAL FEATURES, LABORATORY FINDINGS AND TREATMENT

Clinical features of anaemia

1. ***General features of anaemia.*** See page 118.
2. ***Characteristic features of iron deficiency anaemia*** are in the form of following epithelial tissue changes:
   - *Nails* become dry, soft and spoon-shaped (koilonychia).
   - *Tongue* becomes angry red (atrophic glossitis).
   - *Mouth* may show angular stomatitis.
   - *Oesophagus* may develop their membranous webs at the postcricoid area leading to dysphagia (Plummer–Vinson syndrome).

Laboratory findings

1. Blood picture and red cell indices

- *Hb* concentration is decreased.
- RBCs are hypochromic (deficient in Hb) and microcytic (smaller in size). They show anisocytosis and poikilocytosis.
- *Red cell indices* like MCV, MCH and MCHC are decreased.

2. Bone marrow findings

- Marrow cellularity: Erythroid hyperplasia,
- Erythropoiesis: normoblastic and
- Marrow iron: Deficient.

3. Biochemical findings

- Serum iron decreases, often under 50 mg% (normal 60–160 mg%).
- Serum ferritin is very low indicating poor tissue iron stores.
- Total iron binding capacity is increased.

Treatment

Treatment of iron deficiency anaemia consists of:

- Oral administration of Fe^{2+} salts and
- Correction of causative factor if possible.

MEGALOBLASTIC ANAEMIA

Megaloblastic anaemias are characterized by the abnormally large cells of erythrocyte series. These are caused by defective DNA synthesis due to deficiency of vitamin B_{12} and/or folic acid (folate).

AETIOLOGICAL TYPES

I. Megaloblastic anaemia due to vitamin B_{12} deficiency

Causes of vitamin B_{12} deficiency are:

1. *Inadequate dietary intake* may occur in:
   - Strict vegetarians and
   - Breast-fed infants.
2. *Malabsorption of vitamin B_{12}* is more often the cause of deficiency and may be due to:
   - *Gastric causes* leading to the deficiency of intrinsic factors such as an autoimmune cause of failure of secretion of intrinsic factor (Addisonian pernicious anaemia), gastrectomy and congenital lack of intrinsic factor.
   - *Intestinal causes* which are associated with decreased vitamin B_{12} absorption are tropical sprue, ileal resection, Crohn's disease, fish tapeworm infestation and intestinal blind loop syndrome.

Addisonian pernicious anaemia

Aetiology. Addisonian pernicious anaemia is the term which is used specifically for the megaloblastic anaemia due to vitamin B_{12} deficiency occurring as a result of failure of secretion of intrinsic factor by the stomach owing to an autoimmune atrophy of gastric mucosa. Thus, pernicious anaemia is an autoimmune disease and in about 50% of patients, antibodies to intrinsic factor can be demonstrated. The disease is rare before the age of 30 years, occurs mainly between 45 and 65 years, and affects females more frequently than males.

Features of pernicious anaemia include:

- *Features of megaloblastic anaemia* (described on page 120) and
- *Specific features* of pernicious anaemia are:
  - Anti-intrinsic factor antibodies in serum (present in 50% cases)
  - Abnormal vitamin B_{12} absorption test corrected by the addition of intrinsic factor (Schilling test).

Treatment of pernicious anaemia consists of regular administration of vitamin B_{12} by intramuscular route:

II. Megaloblastic anaemia due to folate deficiency

Salient features of folic acid and its role in erythropoiesis (see page 108).

Causes of folate deficiency are:

1. *Inadequate dietary intake* due to poor intake of vegetables as seen in poor people, infants and alcoholics.
2. *Malabsorption,* e.g. in coeliac disease, tropical sprue and Crohn's disease.

3. *Increased demand* as occurs in:
   - Physiological conditions, such as pregnancy, lactation and infancy and
   - Pathological conditions of cell proliferation, such as increased haematopoiesis (as in haemolysis) and malignancies.
4. *Effect of drugs,* such as certain anticonvulsants (e.g. phenytoin), contraceptive pills and certain cytotoxic drugs (e.g. methotrexate).
5. *Excess urinary folate loss,* e.g. in active liver disease and congestive heart failure.

Features of folate deficiency anaemia include:

- *Features of megaloblastic anaemia* and
- *Specific features of folate deficiency* are:
  - Low serum folate levels
  - Low red cell folate levels.

CLINICAL FEATURES OF MEGALOBLASTIC ANAEMIA

A. General features of anaemia

(see page 118).

B. Characteristic features of megaloblastic anaemia

1. Blood picture and red cell indices

- *Hb* level is low.
- RBCs are larger in size (macrocytosis) but contain a normal concentration of Hb (normochromia).
- *MCV* increases to 95–160 μm^3 (normal 78–94 μm^3).
- *MCH* increases to 50 pg (normal 28–32 pg).
- *MCHC* usually normal (35 ± 3%) because both MCV and MCH increase. In late stages, MCHC may decrease.
- *Peripheral smear* shows nucleated RBCs with marked anisocytosis and poikilocytosis.
- *Reticulocyte count* increases to more than 5% (normal less than 1%).
- *Life span of RBCs* is decreased.
- *WBCs and platelets* decrease because of encroachment of megaloblastic tissue.

2. Bone marrow picture

- Bone marrow shows *megaloblastic hyperplasia* characterized by presence of:
  - 70% proerythroblasts and early normoblasts (normal 30%) and
  - 30% intermediate and late normoblasts (normal 70%).
- *Marrow iron.* Prussian blue staining for iron in the marrow shows an increase in the number and size of iron granules in the erythroid precursors.

3. Biochemical finding

- *Serum bilirubin* increases more than 1 mg/dL (normal 0.2–0.8 mg/dL) due to excessive destruction of RBCs in spleen, liver and bone marrow.
- *Urine urobilinogen* excretion may increase due to increased serum bilirubin.
- *Serum iron and* ferritin is usually increased because iron is not utilized by the immature RBCs.
- *Serum vitamin B_{12}* levels are decreased (normal 200–900 pg/mL) in patients with megaloblastic anaemia due to vitamin B_{12} deficiency.
- *Serum folate levels* are decreased in the patients with megaloblastic anaemia due to folic acid deficiency.
- *Red cell folate levels are more reliable indicator of tissue stores of folate than serum. In folic acid deficiency, red cell folate levels are decreased.*

Chapter 3.3

White Blood Cells

TYPES OF WHITE BLOOD CELLS AND THEIR COUNTS

TYPES OF WHITE BLOOD CELLS

The white blood cells (WBCs) or leucocytes are so named since they are colourless in contrast to the red colour of RBCs. These are nucleated cells and play an important role in the defence mechanism of the body. The leucocytes of the peripheral blood are of two main varieties, distinguished by the presence or absence of granules. These are granulocytes and agranulocytes (non-granulocytes).

Granulocytes

The white blood cells with granules in their cytoplasm are called granulocytes. Depending upon the colour of granules, granulocytes are further divided into three types:

Neutrophils. They contain granules which take both acidic and basic stain.

Eosinophils. They contain granules which take acidic stain.

Basophils. They contain granules which take basic stain.

Agranulocytes

White blood cells which do not contain granules in their cytoplasm are called agranulocytes. These are of two types:

- Lymphocytes and
- Monocytes.

NORMAL WBC COUNTS

Total leucocyte count

Total leucocyte count (TLC) varies with age as:

Adults: 4000–11,000/μL of blood.
At birth, in full-term infant: 10,000–25,000/μL of blood.
Infants up to 1 year of age: 6000–16,000/μL of blood.

Differential and absolute leucocyte count

Differential leucocyte count (DLC) and absolute count in normal adults is shown in Table 3.3-1.

Clinical significance of differential and absolute counts

The DLC determines if there is an increase or decrease in a particular type of leucocyte, because in different diseases,

Table 3.3-1 Differential leucocyte count and absolute count in normal adults

WBCs	Differential count (%)	Absolute count (per μL)
Granulocytes		
• Neutrophils	40–75	2000–7500
• Eosinophils	1–6	40–440
• Basophils	0–1	0–100
Agranulocytes		
• Lymphocytes	20–40	1500–4000
• Monocytes	2–10	500–800

one or the other type of cells show an increase or decrease in its numbers. The differential count is done in 100 or 200 cells and shows only a relative increase or decrease in particular variety of cells. DLC alone is not of much importance and so never done as an isolated test, but always it is a part of full blood counts including TLC and then calculating absolute count.

VARIATIONS IN WBC COUNT

Leucocytosis

Leucocytosis refers to increase in total WBC count above 11,000/μL.

Physiological causes of leucocytosis are:

1. Age,
2. Exercise,
3. After food intake,
4. Mental stress,
5. Pregnancy and
6. Exposure to low temperature.

Pathological causes of leucocytosis are:

1. Acute bacterial infections especially by the pyogenic organisms,
2. Acute haemorrhage,
3. Burns,
4. Post-operative period,
5. Tuberculosis and
6. Glandular fever.

Leucopenia

Leucopenia refers to decrease in the total WBC count below 4000/μL.

Causes of leucopenia are:

1. Infections by the non-pyogenic bacteria, especially typhoid fever and paratyphoid fever.
2. Viral infections, such as influenza, smallpox, mumps, etc.
3. Protozoal infections.
4. Starvation and malnutrition.
5. Aplasia of bone marrow.
6. Bone marrow depression due to:
   - Drugs, such as chloromycetin and cytotoxic drugs used in malignant diseases.
   - Repeated exposure to X-rays or radiations.
   - Chemical poisons like arsenic, dinitrophenol and antimony.

FORMATION OF WHITE BLOOD CELLS

The process of development and maturation of white blood cells (leucocytes), called *leucopoiesis*, is a part of haemopoiesis (formation of blood cells). All the blood cells develop from the so-called pluripotent *haemopoietic stem cells (PHSCs).* The stem cells after a series of divisions differentiate into *progenitor cells* which are also called colony forming units (CFU) (for details see page 105). The leucopoiesis can be discussed under two headings:

- Formation of granulocytes (granulopoiesis) and monocytes, and
- Formation of lymphocytes (lymphopoiesis).

FORMATION OF GRANULOCYTES AND MONOCYTES

The granulocytes and monocytes are formed in the bone marrow from the colony forming unit called CFU-GM (colony forming unit granulocytes and monocytes): The progenitor cells (CFU-GM) forming different cells are further named as:

- CFU-G are neutrophil forming units,
- CFU-EO refers to eosinophil forming units,
- CFU-Ba are basophil forming units and
- CFU-M refers to monocyte forming units.

The development of granulocytes through various stages is called *myeloid series* and development of monocytes through various stages is called *monocyte–macrophage series.*

MYELOID SERIES

Some facts about granulopoiesis

- The cells of myeloid series include myeloblast (most primitive precursor), promyelocytes, myelocytes, metamyelocytes, band forms and segmented granulocyte (mature form).
- The process of granulopoiesis takes about 12 days.
- Granulocytes are formed and stored in the bone marrow. When need arises they are released in circulation.

Normally the number of granulocytes in bone marrow is about three times as compared to circulating in the peripheral blood.

Features of the cells of myeloid series (Fig. 3.3-1)

The characteristic features of cells of myeloid series are summarized in Table 3.3-2.

MONOCYTE–MACROPHAGE SERIES (FIG. 3.3-1)

These are described separately because of different morphological stages which include: monoblast, promonocytes and monocyte.

1. Monoblast. It is a large cell similar in structure to the myeloblast from which it cannot be distinguished on morphological grounds alone.

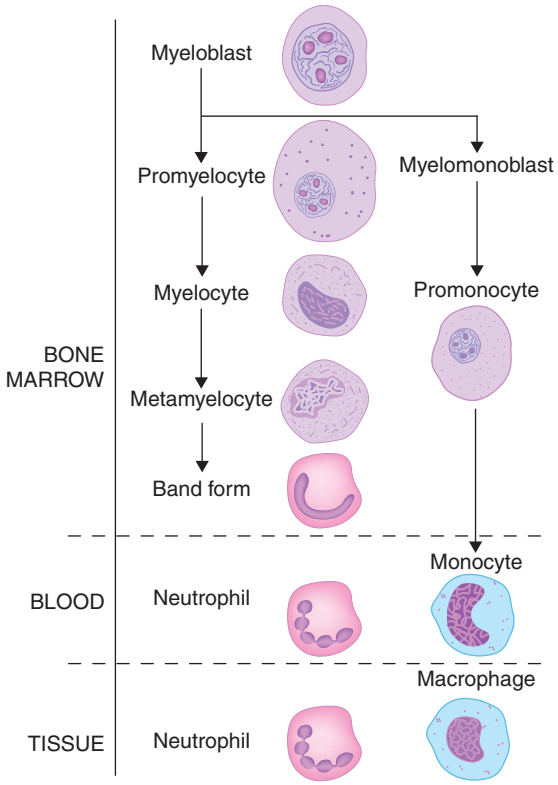


Fig. 3.3-1 Granulopoiesis and features of the cells of myeloid series and monocyte–macrophage series.

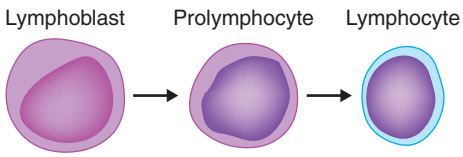


Fig. 3.3-2 Formation of lymphoid series cells.

2. Promonocyte. It is a young monocyte about 20 µm in diameter. Its *nucleus* is large, indented (often kidney shaped) and contains one nucleolus. The nuclear chromatin is arranged in a loose network. The *cytoplasm* is basophilic and contains *no azurophilic granules,* but may have fine granules which are larger than those in the mature monocyte.

3. Monocyte. Morphological features of a mature monocyte are described on page 131.

From the bone marrow, the monocytes migrate into spleen and lymphoid tissues in considerable numbers. The transformed stages of these cells in the various tissues are called *tissue macrophages* and form a part of *tissue–macrophage system*, which was previously known as *reticuloendothelial system*.

FORMATION OF LYMPHOCYTES

The lymphocytes are formed from the *lymphocyte stem cells* which are formed from the PHSCs in the bone marrow. The lymphocyte stem cells migrate into the thymus and the peripheral lymphoid tissues, where they proliferate and mature into lymphocytes. In man, the bone marrow and thymus form the primary lymphopoietic *organs,* where lymphoid stem cells undergo spontaneous division independent of antigenic stimulation. The tissues which actively produce lymphocytes from the germinal centres of lymphoid follicles as a response to antigenic stimulation constitute the so-called *secondary or reactive lymphoid tissue.* It is comprised of the:

- Lymph nodes,
- Spleen and
- Gut associated lymphoid tissue.

Lymphoid series

The maturation stages of lymphoid series are (Fig. 3.3-2):

1. Lymphoblast. It is the earliest recognizable cell of the lymphoid series. It is actively dividing cell and resembles the myeloblast morphologically except for the following minor differences:

- Nuclear chromatin is slightly clumped and stippled as compared to the fine meshwork in myeloblast and
- Nuclear membrane is fairly dense as compared to very fine membrane of the myeloblast.

2. Prolymphocyte. It is the intermediate stage between the lymphoblast and mature lymphocyte. Its features are:

- *Diameter* is 9–18 µm.
- *Nucleus* is round to indented with slightly stippled or coarse chromatin and may have 0–1 nucleoli.
- *Cytoplasm* is scanty and non-granular.

Table 3.3-2 Characteristic features of cells of myeloid series

Cells	Size (μm)	Nucleus	Cytoplasm	Mitosis
1. ***Myeloblast***	16–20	• Large (nearly filling the cells), round to oval • Fine chromatin and contains 2–5 well-defined pale nucleoli	Basophilic, present as a thin rim around the nucleus Granules absent	Marked (+++)
2. ***Promyelocytes***	14–18	• Round or oval, slightly smaller than the nucleus of myeloblast • Chromatin fine and condensed • The nucleoli are present but are less prominent and fewer than those in the myeloblast	Amount increases and is characterized by the presence of *azurophil granules*. These granules are also called *primary non-specific granules* and these give a positive reaction with the peroxidase stain	Marked (+++)
3. ***Myelocyte***	12–16	Eccentric, round to oval Coarse nuclear chromatin and no nucleoli	*Specific (secondary) granules present* and accordingly the cell can be identified at this stage as: • Neutrophil myelocyte • Eosinophil myelocyte • Basophil myelocyte The primary granules are also present at this stage, but their formation is stopped	Continues up to this stage. Multiplication of these cells is maximum.
4. ***Metamyelocyte***	10–14	• Decreases in size, becomes indented and lobed (horse-shoe shape) • The nuclear chromatin is dense and clumped. Nucleoli are absent	Amount increases and becomes more liquid Both primary and secondary granules are present. Depending upon the features of secondary granules the metamyelocytes are distinguished as: • Neutrophil metamyelocyte • Eosinophil metamyelocyte • Basophil metamyelocyte	Stops at this stage.
5. ***Band or stab form*** (juvenile granulocytes)	*Size* is slightly smaller than metamyelocyte	• Further condensation of chromatin, nucleus becomes band shaped (deeply indented V shaped)	Pink and contains fine evenly distributed granules	Absent
6. ***Mature granulocytes*** (neutrophils, eosinophils and basophils)				

Morphological features of different granulocytes are described at page 125.

3. Lymphocytes. Prolymphocytes mature successively into a large lymphocyte and a small lymphocyte, both of which are found in circulation.

- Then some lymphocytes enter thymus where they are processed and come out as T lymphocytes. In thymus, a factor called *thymosin* plays an important role in the processing.
- Some lymphocytes are processed in liver (in fetal life) and bone marrow (after birth). These come out as B *lymphocytes.* The word B comes from 'bursa of Fabricius' which is the site of B cell processing in birds.

Morphological features of large and small lymphocytes are described on page 130.

REGULATION OF LEUCOPOIESIS

The constancy of leucocyte count suggests an efficient feedback mechanism to control their production and release (Fig. 3.3-3). During tissue injury and inflammation, bacterial toxins, products of injury, etc. cause a great increase in the rate of production and release of leucocytes. Thus, unlike erythropoiesis, the products of dead and dying

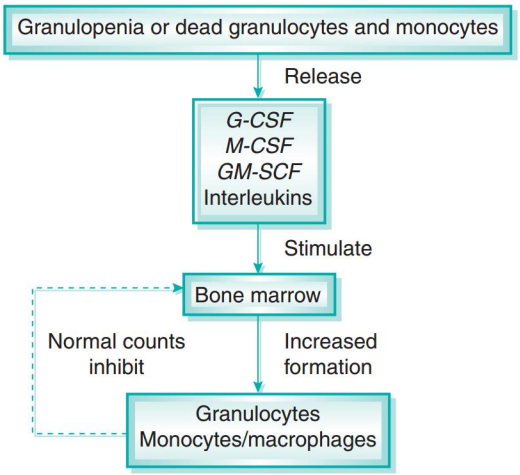


Fig. 3.3-3 Schematic diagram of regulation of leucopoiesis.

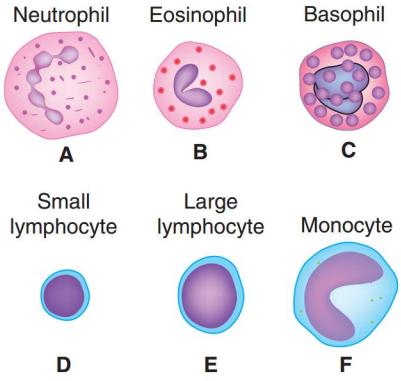


Fig. 3.3-4 Morphological features of white blood cells: A, neutrophil; B, eosinophil; C, basophil; D, small lymphocyte; E, large lymphocyte and F, monocyte.

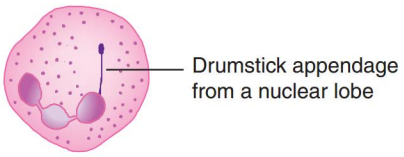


Fig. 3.3-5 Neutrophil showing female sex chromatin.

white cells themselves control leucopoiesis. The substances which stimulate or inhibit the process of leucopoiesis are complex and include the following:

Role of cytokines

The cytokines which control the formation of different types of granulocytes are called *colony stimulating factors (CSFs).* The CSFs are glycoproteins formed by monocytes and T lymphocytes and include:

- *G-CSF* which stimulates granulocyte precursors,
- *M-CSF* which stimulates monocytic precursors and
- *GM-CSF* which stimulates both the granulocyte and monocytic precursors,
- The cytokines that control lymphocyte formation are called *interleukins,* e.g. IL-I, IL-3, etc. The interleukins are formed by monocytes, macrophages and endothelial cells.

Role of prostaglandins

Prostaglandins formed by monocytes, *lactoferrin* and possibly some other agents also play role in control of leucopoiesis.

MORPHOLOGY, LIFE SPAN, FUNCTIONS AND VARIATIONS IN COUNTS OF WBCs

Morphological features of various types of WBCs as studied under microscope with Leishman's staining and haematoxylin–eosin stain (Fig. 3.3-4) are summarized below along with their functions and variations in their counts.

NEUTROPHILS

Morphological features

The polymorphonuclear neutrophils commonly called polymorphs or neutrophils have following morphological features (Fig. 3.3-4A):

Diameter. Diameter of a neutrophil varies from 10 to 14 μm.

Nucleus. A young neutrophil has a single horse-shoe shaped nucleus which becomes lobed as the cell grows. Nucleus of a mature neutrophil is purple in colour and multilobed (2–6 lobes), that is why a neutrophil is also called polymorphonuclear leucocytes. The lobes of the nucleus are connected by the chromatin filaments, seen clearly through the cytoplasm.

Note. Sex chromatin is a normal finding seen in 2–3% of neutrophils in females. It consists of a drumstick appendage of chromatin about 1 μm across, and attached to one of the nuclear lobes by a thin chromatin strand (Fig. 3.3-5). Their presence is indicative of 2X chromosomes.

Cytoplasm. Cytoplasm of neutrophil is pale bluish in colour and full of fine (pinpoint) granules.

- Granules take both acidic and basic stain and look violet-pink in colour.
- Granules are *lysosomal* in origin and contain variety of enzymes which include glycosidases, sulphatases,

phosphatases, nucleases and proteolytic enzymes. They can thus lyse any type of substance.
- In addition, the granules liberate histamine and peroxidase enzyme which help in killing the ingested bacteria.

KINETICS, LIFE SPAN AND FATE OF NEUTROPHILS

- The neutrophils released from the bone marrow enter the circulation. In the blood, they exist in two equal populations:
  - The *circulating pool* comprises 50% cells which are circulating in the blood at any instant and
  - The *marginal pool* is constituted by the rest of 50% of cells which remain marginated or sidelined, i.e. sticking to the endothelial cells of closed capillaries, venules, small veins and sinusoids.
  - There is a rapid exchange between the two pools.
- For every neutrophil in the blood, there are about 100 mature neutrophils held in the bone marrow as a reserve. These are released in the circulation when required. The stimulus for their release comes from the dead leucocytes which release a granulocyte inducing factor and also by the hormone cortisol.
- Following their release from the bone marrow, granulocytes remain in the circulating blood for 8–10 h and then they enter the tissues. After migration into tissues, they never return to the blood stream. In the tissues, they are either destroyed during phagocytosis or die due to senescence after 4–5 days. The dead neutrophils are taken up by the macrophages.

FUNCTIONS

The neutrophils along with the monocytes constitute the first line of defence against the micro-organisms, viruses and other injurious agents that enter the body. Neutrophils subserve this role by the following mechanisms:

1. Phagocytosis. The neutrophils engulf the foreign particles or bacteria and digest them and ultimately may kill them by a process called phagocytosis. Various steps of phagocytosis are described below in detail.

2. Reaction of inflammation. The neutrophils also release leukotrienes, prostaglandins, thromboxanes, etc. that bring about the reactions of inflammation like vasodilatation and oedema.

3. Febrile response. The neutrophils contain a fever-producing substance called *endogenous pyrogen* which is an important mediator of febrile response to the bacterial pyrogens.

Phagocytosis

Phagocytosis (cell eating) refers to the process of engulfment and destruction of solid particulate material by the cells. The process of phagocytosis involves following steps:

1. Margination. In the area of infection, the neutrophils gets *marginated,* i.e. get attached towards the capillary endothelium and start rolling along its surface. This process is called *margination* or *pavementing*. The margination is caused by binding of *selectins* (cell adhesion molecules) present on the endothelial cells with the carbohydrate molecules present on the surface of neutrophils. The endothelial selectins are markedly increased in areas where there is inflammation (Fig. 3.3-6A).

2. Emigration and diapedesis. The marginated neutrophils are emigrated in large number from the blood to the site of *infection* by diapedesis into the tissues by passing through the junction between endothelial cells of the blood vessels (Fig. 3.3-6B).

3. Chemotaxis. Chemotaxis refers to the process by which the neutrophils are attracted towards bacteria at the site of inflammation (Fig. 3.3-6C). The process of chemotaxis is mediated by the chemotactic agents called *chemokines* which are released at the infected area. There are various types of chemokines and include:

- Leukotriene B_4 (LTB_4),
- Components of complement system (C_5) and cytokines (polypeptides from lymphocytes and monocytes).

The chemoattractants increase the adhesive nature of neutrophils which form clumps surrounding the infected area (Fig. 3.3-6C).

4. Opsonization (attachment stage). Opsonization refers to the process of coating of bacteria by the *opsonins* by which bacteria become tasty to the phagocytes (Fig. 3.3-6D). The principal opsonins are the naturally acting factors in the serum and include IgG opsonin and opsonin fragment of the complement protein.

5. Engulfment stage. The neutrophils project *pseudopodia* in all directions around the opsonized particle which is bound to the surface of neutrophil (Fig. 3.3-6E). Pseudopodia meet each other on opposite side and fuse. This creates an enclosed chamber with the engulfed material. It breaks away from the membrane forming a *phagocytic vesicle.* Then the lysosomes of the cell fuse with the phagocytic vesicle to form *phagolysosome* or phagosome (Fig. 3.3-6F).

6. Secretion (degranulation) stage. Once the bacterium is engulfed, the lysosomes pour their enzymes into the vesicle and also in interstitial space (Fig. 3.3-6G). This process is called degranulation. There are large numbers of proteolytic enzymes especially geared up for digesting bacteria. In addition, lysosomes of macrophages also contain lipases

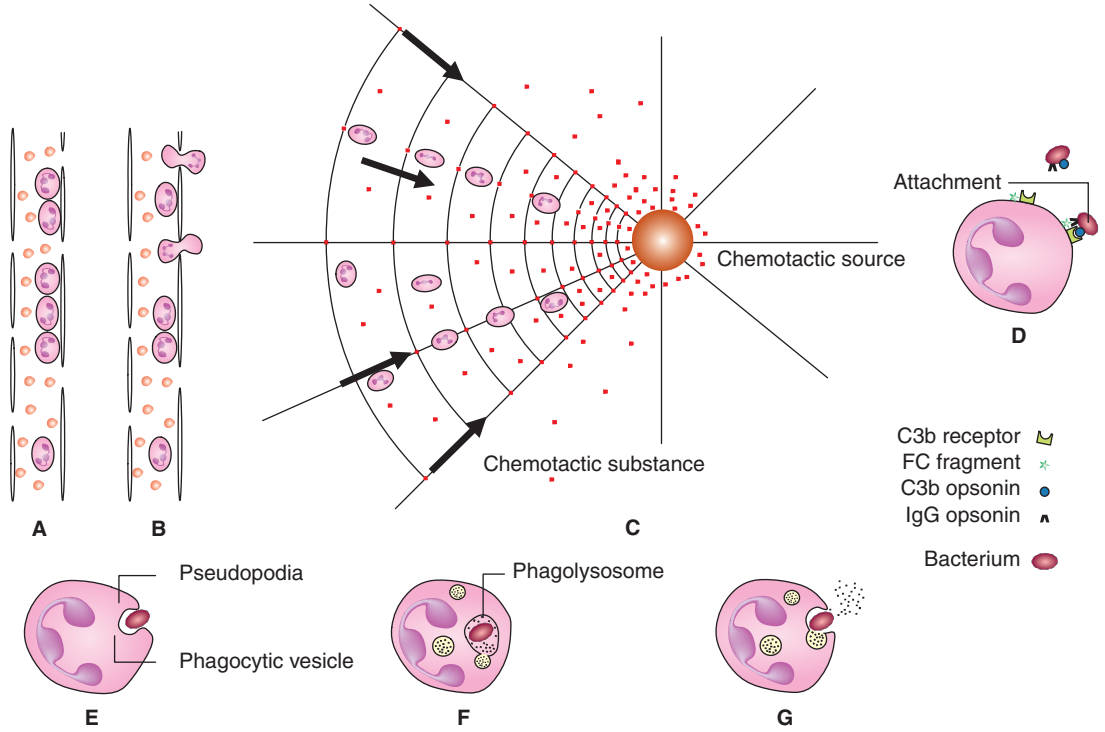


Fig. 3.3-6 Stages of phagocytosis: A, margination; B, diapedesis; C, chemotaxis; D, opsonization; E, engulfment; F, formation of phagolysosome and G, degranulation.

which can digest the thick lipid membranes possessed by certain bacteria.

7. Killing or degradation stage. The neutrophils and macrophages contain *bactericidal agents* (defensins α and β) which can kill most of the bacteria. The bactericidal substance accomplishes the killing process by the following mechanisms:

- *Oxygen-dependent bactericidal mechanism* which is mediated by oxidizing agents (superoxides, H_2O, etc.) formed by the membrane bound enzyme NADPH oxidase which leads to sharp increase in O_2 intake (respiratory burst) and O_2^- is generated by the following reaction:

  $$NADPH + H^+ + 2O_2 \rightarrow NADP\ 2H^+ + 2O_2^-$$

  O_2^- (free radical) is formed by the addition of one electron to O_2. Two O_2^- react with two H^+ to form H_2O_2. This reaction is catalyzed by enzyme superoxide dismutase present in the cytoplasm.

  $$O_2^- + O_2^- + H^+ + H^+ \rightarrow H_2O_2 + O_2$$

  O_2^- and H_2O_2 (oxidants) both are the bactericidal agents.
- Neutrophils also discharge another enzyme known as myeloperoxidase which converts Cl^-, Br^-, I^- and CN^- to their corresponding acids (HOCl, HOBr, HOI, etc.). These acids are also potent oxidants.
- *Oxygen-independent bactericidal mechanism* works through lysosomal hydrolases, defensins and cationic protein.

IMPORTANT NOTE

- Phagocytosis is completed after the stage of killing is over.
- A neutrophil can usually phagocytose 5–20 bacteria before it itself become inactivated or dead.
- Neutrophils are not capable of phagocytosing particles much larger than bacteria.
- Neutrophils killed by the toxins released from the bacteria are collected in the centre of infected area. These are called pus cells and together with plasma leaked from the blood vessels, liquefied tissue cells and red blood cells escaped from the damaged capillaries constitute the pus.

VARIATIONS IN COUNTS

Neutrophilia

Neutrophilia refers to increase in the circulating neutrophil counts (absolute count >10,000/μL). It is the commonest cause of leucocytosis.

Causes

Physiological causes of neutrophilia are:

- Newborn babies,
- After exercise,
- After meals,
- Pregnancy, menstruation, parturition and lactation,
- Mental stress and emotional stress, and
- After injection of epinephrine.

Pathological causes of neutrophilia are:

- Acute pyogenic bacterial infections,
- Non-infective inflammatory conditions like gout, acute rheumatic fever,
- Acute tissue destruction as in:
  - Burns,
  - Post-operatively and
  - Myocardial infarction.

Neutropenia

Decrease in the neutrophil count is known as neutropenia (absolute count < 2500/μL).

Causes of neutropenia are:

- Typhoid and paratyphoid fever,
- Malaria,
- Aplasia of bone marrow,
- Bone marrow depression due to:
  - Drugs, such as chloromycetin and cytotoxic drugs used in the malignant diseases
  - Repeated exposure to X-rays and radiations
  - Chemical poisons like arsenic.

Arneth count

Counting the number of neutrophils with different nuclear lobes and expressing the count as percentage of cells with different number of nuclear lobes is called Arneth or Cooke's Arneth count. Different stages with a normal count are depicted in Table 3.3-3.

Clinical significance. The Arneth count is useful in judging the rate of formation of neutrophils. The three-lobed cells are fully mature and functionally the most efficient. Thus, presence of younger cells (shift to the left) and more mature cells (shift to the right) in the blood can provide important information about the rate of formation and release of neutrophils from the bone marrow.

Table 3.3-3 Cooke's Arneth count

Stage	Nuclear lobes	Normal count (%)
• Stage I (N_1)	One (the nucleus is C-shaped)	5–10
• Stage II (N_2)	Two lobes are connected by a filament	20–30
• Stage III (N_3)	Three lobes connected by a chromatin filament	40–50
• Stage IV (N_4)	Four lobes connected by a chromatin filament	10–15
• Stage V (N_5)	Five lobes or more	3–5

- In *left shift* (more younger cells) $N_1 + N_2 + N_3$ is more than 80%. It indicates hyperactive bone marrow (high rate of formation).
- In *right shift* (more mature cells) $N_4 + N_5$ is more than 20%. It indicates hypoactive bone marrow (slow rate of formation).

EOSINOPHILS

MORPHOLOGICAL FEATURES

Morphological features of eosinophils (Fig. 3.3-4B) are:

Diameter. Diameter of eosinophils is similar to neutrophils, i.e. 10–14 μm.

Nucleus. Nucleus is purple in colour and is bilobed in 85% of the cells. The two lobes are connected by the chromatin strands and thus look *spectacle shaped.* The remaining 15% of the eosinophils have a trilobed nucleus.

Cytoplasm. Cytoplasm is acidophilic and appears bright pink in colour.

- It contains coarse, deep red staining granules which do not cover the nucleus.
- The granules contain histamine, lysosomal enzymes and eosinophil chemotactic factor of anaphylaxis (ECF-A).
- The granules in eosinophils contain basic protein and stain more intensely for peroxidase than granules in the neutrophils.

FUNCTIONS

1. Mild phagocytosis. Eosinophils are not very motile and thus have a very mild phagocytic activity.

2. Role in parasitic infestations. They play an important role in the defence mechanism of body especially *in parasitic infestations.* Eosinophils act through the following lethal substances present in their granules:

- *Major basic protein.* It is highly larvicidal polypeptide. Because of this, eosinophils are able to damage the parasitic larvae which are large to be engulfed by phagocytosis.
- *Eosinophil cationic protein* is a potent bactericidal and major destroyer of helminths.
- *Eosinophil peroxidase* is capable of destroying helminths, bacteria and tumour cells.

3. Role in allergic reaction. The eosinophils increase in number in allergic conditions like bronchial asthma and hay fever.

- They are capable of *detoxifying* inflammation inducing substances (released by mast cells and basophils) like histamine and bradykinin.

- They inhibit mast cell degranulation.
- They phagocytose and destroy antigen–antibody complexes and thus prevent spread of local inflammatory process.
- *Arylsulphatase B*, present in the fine granules of eosinophils, has the ability of inactivating sulphur-containing leukotrienes that tissue mast cells liberate in immediate hypersensitivity reactions.
- *Lysophospholipase* is an unusual membrane-bound protein present in the eosinophils which forms crystals called *Charcot-Leyden crystals* in the pulmonary secretions of patients with bronchial asthma.

4. Role in immunity. The eosinophils are present in abundance in the mucosa of respiratory tract, gastrointestinal tract and urinary tract, where they probably provide mucosal immunity.

VARIATIONS IN COUNTS

Eosinophilia

Eosinophilia refers to increase in the eosinophil count (absolute count > 500/μL).

Causes of eosinophilia are:

- Allergic conditions like bronchial asthma and hay fever,
- Parasitic infestation, e.g. intestinal worms like hookworm, roundworm and tapeworm,
- Skin diseases like urticaria and
- Scarlet fever.

Eosinopenia

Eosinopenia is the decrease in the eosinophil count (absolute count < 50/μL).

Causes of eosinopenia are:

- Adrenocorticotrophic hormone (ACTH) and steroid therapy,
- Stressful conditions and
- Acute pyogenic infections.

BASOPHILS

MORPHOLOGICAL FEATURES (FIG. 3.3-4C)

Diameter. Diameter of a basophil is similar to neutrophil and eosinophil, i.e. 10–14 μm.

Nucleus. Nucleus of basophils is irregular, may be bi-lobed or tri-lobed, and its boundary is not clearly defined because of overcrowding with the coarse granules.

Cytoplasm. Cytoplasm of basophils is slightly basophilic and appears blue. It is full of granules.

- Granules of basophils are very coarse and stain deep purple or blue with basic (methylene) dye.
- Granules are in plenty and completely fill the cell and overload the nucleus.
- The granules of basophils contain heparin, histamine and 5-HT
- The granules also contain eosinophil chemotactic factor-A (ECF-A).

FUNCTIONS

1. Mild phagocytosis. Basophils have very mild phagocytic function.

2. Role in allergic reaction. Basophils release histamine, bradykinin, slow reacting substances of anaphylaxis (SRS-A) and serotonin (5HT). These substances, in turn, cause local vascular and tissue reactions that cause many allergic manifestations.

3. Role in preventing spread of allergic inflammatory process. Basophils also release eosinophil chemotactic factor that causes eosinophils to migrate toward the inflamed allergic tissue. Eosinophils then phagocytose and destroy antigen–antibody complexes and prevent spread of local inflammatory process.

4. Release of heparin. Basophils release heparin in the blood which:

- prevents clotting of the blood and
- activates the enzyme lipoprotein lipase which removes fat particles from the blood after a fatty meal.

VARIATIONS IN COUNTS

Basophilia

Basophilia refers to increase in the basophil count (absolute count > 100/μL).

Causes of basophilia are:

- Viral infections, e.g. influenza, small pox and chicken pox,
- Allergic diseases and
- Chronic myeloid leukaemia.

Basopenia

Decrease in the basophil count is called basopenia.

Causes of basopenia are:

- Corticosteroid therapy,
- Drug-induced reactions and
- Acute pyogenic infections.

MAST CELLS

Mast cells are large tissue cells resembling the basophils. These are present in bone marrow and immediately outside

the capillaries in the skin. These do not enter the blood circulation.

Functions. Mast cells play role in the allergic reactions similar to the basophils.

LYMPHOCYTES

MORPHOLOGICAL FEATURES

There are two types of lymphocytes, large and small having almost similar structure. Morphological features of lymphocytes are:

Diameter. Diameter of large lymphocytes varies from 12 to 16 µm and that of small lymphocytes from 7 to 10 µm (Fig. 3.3-4D and E).

Nucleus. Lymphocytes have a large round, single nucleus which almost completely fills the cell. It stains blue very deeply giving *ink-spot appearance.* Nuclear chromatin is coarsely clumped and shapeless.

Cytoplasm. The cytoplasm is scanty, i.e. its amount is always less than that of the nucleus. It is seen as a crescent of clear light blue colour around the nucleus. Cytoplasm does not contain visible granules.

Functional subtypes

Based on their developmental background, life span and functions, the small lymphocytes have been broadly classified into three subtypes (Fig. 3.3-7):

1. ***B lymphocytes*** which are processed in the bone marrow and concerned with the humoral immunity.
2. ***T lymphocytes*** which are processed in the thymus and concerned with cellular immunity.
3. ***Natural killer (NK) cells*** are lymphocyte-like cells that non-specifically kill any cell that is coated with immunoglobulin IgG. This phenomenon is called *antigen-dependent cell-mediated cytotoxicity (ADCC).* Thus NK cells provide *innate immunity.* The NK cells lack identifying surface markers.

FUNCTIONS

Lymphocytes play an important role in immunity. B lymphocytes as well as their derivatives, the plasma cells are responsible for the development of *humoral immunity* also called antibody-mediated immunity (see page 139). T lymphocytes are responsible for the development of *cellular immunity*, also called cell-mediated immunity or T cell immunity (see page 142).

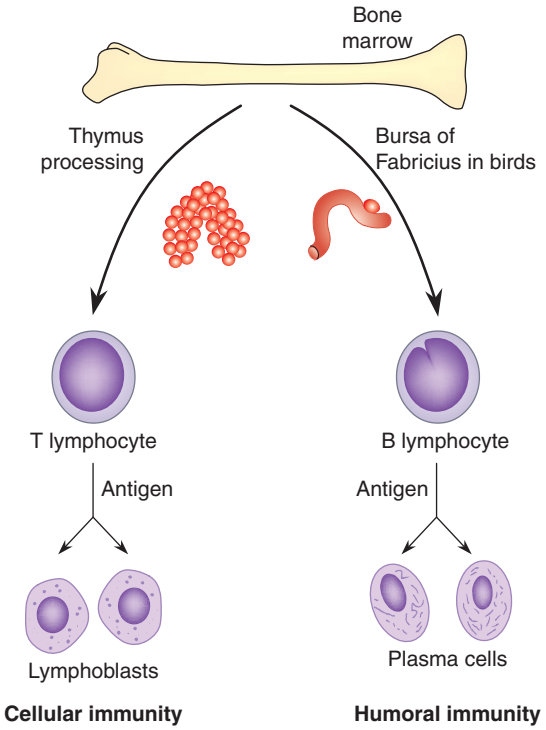


Fig. 3.3-7 Processing of stem cells in thymus and bone marrow to become immunocompetent T and B lymphocytes.

VARIATIONS IN COUNTS

Lymphocytosis

Lymphocytosis refers to increase in the lymphocyte count (absolute count > 4000/µL).

Causes

Physiological causes of lymphocytosis are:

- *In healthy infants and young children,* the lymphocytes count is usually high (about 60% in DLC) while the TLC is normal (relative lymphocytosis).
- *In females,* during menstruation lymphocytes are increased.

Pathological causes of lymphocytosis are:

- Chronic infections like tuberculosis, hepatitis and whooping cough,
- Viral infections like chicken pox,
- Autoimmune diseases like thyrotoxicosis,
- Infectious mononucleosis and
- Lymphatic leukaemia (most common cause of lymphocytes > 10,000/µL).

Lymphopenia

Lymphopenia or lymphocytopenia refers to decrease in lymphocyte count (absolute count below 1500/µL).

Causes of lymphopenia are:

- Patients on corticosteroid and immunosuppressive therapy,
- Hypoplastic bone marrow,
- Widespread irradiation and
- Acquired immunodeficiency syndrome (AIDS, see page 148).

MONOCYTES

MORPHOLOGICAL FEATURES (FIG. 3.3-4F)

Diameter. The monocyte is the largest mature leucocyte in the peripheral blood measuring some 12–20 μm in diameter.

Nucleus. The nucleus of a monocyte is large, single and eccentric in position, i.e. present on one side of the cell. It may be notched, or indented, i.e. horseshoe or kidney shaped.

Cytoplasm. The cytoplasm is abundant, pale blue and usually clear (no granules); sometimes, it may contain fine purple, dust-like granules called *azure granules* which may be few or numerous.

KINETICS, LIFE SPAN AND FATE OF MONOCYTES

- The kinetics of monocytes is less well understood than that of granulocytes.
- After release from the bone marrow, the monocytes remain in circulation for 10–20 to over 40 h and then they leave the blood to enter the extravascular tissues.
- In the tissues the monocytes get converted to macrophages and form the part of so-called *tissue macrophage system* (reticuloendothelial system). In the tissues, they can live for months or even years unless destroyed while performing the phagocytic function.

FUNCTIONS

1. Role in defence mechanism. Monocytes along with the neutrophils play an important role in the body's defence mechanism. Their main function *is phagocytosis.* These are more powerful phagocytes than neutrophils and are capable of phagocytosing as many as 100 bacteria. They also have ability to engulf large particles such as red blood cells and malarial parasites.

The process of phagocytosis by monocytes is similar to that described in the neutrophils (see page 126).

2. Role in tumour immunity. Monocytes may also kill tumour cells after sensitization by the lymphocytes.

3. Synthesis of biological substances. Monocytes synthesize complement and other biologically important substances.

VARIATIONS IN COUNTS

Monocytosis

A rise in the blood monocytes above 800/μL is termed monocytosis.

Causes of monocytosis are:

1. *Certain bacterial infections*, such as tuberculosis, syphilis and subacute bacterial endocarditis.
2. *Infectious mononucleosis* or the so-called glandular fever.
3. *Viral infections.*
4. *Protozoal* and rickettsial infections, e.g. malaria and kala-azar.

Monocytopenia

Monocytopenia refers to decrease in the monocyte count.

Causes. Monocytopenia is rare. It may be seen in the hypoplastic bone marrow.

LEUKAEMIAS

Leukaemias constitute a group of malignant diseases of the blood in which there occurs an increase in the total WBC count associated with presence of immature WBCs in the peripheral blood. The total WBC count is usually above 50,000/μL and may be as high as 100,000–300,000/μL.

- The proliferation of leukaemic cells takes place primarily in the bone marrow and in certain form in the lymphoid tissues.
- There are associated features of:
  - Bone marrow failure (e.g. anaemia, thrombocytopenia), and
  - Involvement of other organs (e.g. liver, spleen, lymph nodes, meninges, brain, skin, etc.)

Types of leukaemias

Leukaemias account for 4% of all cancer deaths. Leukaemias are classified on the basis of cell types pre-dominantly into *myeloid* (involving cells derived from the myeloid stem cells) and *lymphoid* (involving the cells derived from lymphoid stem cells) on the basis of natural history of disease each variety can be divided into *acute* and *chronic* types. In this way, the main types of leukaemias are:

1. Acute myeloblastic leukaemia,
2. Acute lymphoblastic leukaemia,
3. Chronic myeloid leukaemia and
4. Chronic lymphoid leukaemia.

Chapter 3.4

Immune Mechanisms

ARCHITECTURE OF IMMUNE SYSTEM
- **Mononuclear–phagocytic system**
- **Lymphoid organs**
  - Central lymphoid organs
  - Peripheral lymphoid organs

IMMUNITY
- **Innate immunity**
- **Acquired immunity**
  - Active immunity
  - Passive immunity

ANTIGENS
- Definition
- Some facts about antigenicity
- Histocompatibility antigens

ANTIBODIES
- Structure of antibody
- Functions of immunoglobulins

DEVELOPMENT OF IMMUNE RESPONSE
- **Development of humoral immunity**
  - Role of humoral immunity
  - Types of humoral immune response
  - Stages of humoral immune response
- **Development of cellular immune response**
  - Role of cellular immunity
  - Types of cellular immune response
  - Stages of cellular immune response
- **Cytokines**

OTHER IMMUNE MECHANISM RELATED ASPECTS
- Immune tolerance
- Autoimmunity
- Hypersensitivity
- Immunodeficiency diseases

ARCHITECTURE OF IMMUNE SYSTEM

The immune system which constitutes the body's defence system consists of immunological cells distributed in two main components:

1. Mononuclear–phagocytic system and
2. Lymphoid organs.

MONONUCLEAR–PHAGOCYTIC SYSTEM

Mononuclear–phagocytic system (MPS) also known as *tissue–macrophage system* is the new name given to the system previously called as *reticuloendothelial system.*

Formation of mononuclear–phagocytic system

The monocytes enter the blood from the bone marrow and circulate for about 3 days. From the blood, the monocytes migrate into the tissue where they attain maturity and they acquire the ability to phagocytose and thus get converted to macrophages. These tissue macrophages scattered in different parts of the body combinedly constitute the *tissue macrophage system* or the so-called *mononuclear–phagocytic system.* The tissue macrophage system includes the macrophages present at the following sites in the body:

- Macrophages lining the sinusoids of liver (Kupffer cells),
- Spleen,
- Bone marrow (littoral cells),
- Lymph nodes,
- Lungs (pulmonary alveolar macrophages or PAM also called dust cells),
- Connective tissue (histiocytes),
- Pleura and peritoneum,
- Subcutaneous tissue,
- Bones (osteoclasts) and
- Central nervous system (microglial cells).

Constituent cells of mononuclear–phagocytic system

The term mononuclear–phagocytic system was coined in 1960 to include the following constituents:

- Precursor cells of the monocyte series from bone marrow,
- Promonocytes from the bone marrow,
- Monocytes from the bone marrow and blood, and

- Tissue macrophages present in the above cited sites in the body.

Functions of mononuclear–phagocytic system

The MPS plays the following roles in the body:

1. Role in inflammation and healing. The cells of MPS ingest cell debris, broken down RBCs, fibrin and bacteria from the inflamed area and promote healing process.

2. Role in defence against the bacteria invading the body tissues. These cells ingest the invading bacteria and are thus concerned with the defence of the body against the bacteria. During infection, these cells rapidly increase in number resulting in the enlargement of the organs which are rich in these cells, e.g. spleen, lymph nodes, etc.

3. Role in the immune response played by MPS:
- The cells of MPS ingest and process the antigen entering the body. Processing of antigen is essential before an antigen can evoke cell-mediated immunity (CMI) or stimulate antibody formation in the plasma cells.
- The cells of MPS have receptors for immunoglobulins and complements, so these are very efficient in phagocytosing the antigen–antibody complement complexes.

4. Role in removal of old RBCs. The aged RBCs or those damaged by the action of antibodies are removed by the cells of MPS in different organs like spleen, bone marrow, liver, etc.

5. Role in removal of old WBCs and platelets. Like RBCs, WBCs and platelets are also removed by the cells of MPS system.

6. Storage function. The cells of MPS store excess lipids and mucoprotein and become swollen.

LYMPHOID ORGANS

The lymphoid component of the immune system consists of a network of lymphoid organs, tissues and cells and the product of these cells. Lymphoid organs can be classified into:

A. *Central or primary lymphoid organs,* which include:
   I. Thymus and
   II. Bursa equivalent (fetal liver and bone marrow)

B. *Peripheral lymphoid organs* which include:
   I. Lymph nodes,
   II. Spleen and
   III. Mucosa associated lymphoid tissue.

A. PRIMARY (CENTRAL) LYMPHOID ORGANS

I. Thymus

The thymus gland is located in mediastinum just above the heart. It consists of many lobules. Histologically, each lobule consists of outer *cortex* and inner *medulla.* Both cortex and medulla contain two types of cells:

- *Epithelial cells.* These cells form a network in which thymocytes and macrophages are found.
- *Thymocytes* refer to the immature lymphocytes predominantly present in the cortex and mature lymphocytes mainly present in the medulla. Lymphocytes produced in the thymus are called thymus-derived lymphocytes, T lymphocytes or simply T cells.

Role of thymus in the immune system

- The main function of the thymus is development of cell-mediated immunity.
- The thymus confers immunological competence on the lymphocytes during their stay in the organ. In the thymus, T lymphocytes are educated so that they become capable of mounting cell-mediated immune response against appropriate antigen.
- The *immunologically competent lymphocytes* migrate from the thymus into peripheral lymphoid organs as mature and these are selectively seeded into the paracortical areas of the peripheral lymph nodes and into the white pulp of the spleen around the central arterioles. These regions are known as *thymus-dependent* areas.

Note. Thymectomy in neonates results in lymphopenia and atrophy of peripheral lymphoid tissue, and there is marked susceptibility to infections.

II. Bursa equivalent

In the human being, the fetal liver and bone marrow appear to be the equivalent of *avian bursa of Fabricius*. The immunocompetent lymphocytes produced in the bursa equivalent are called B lymphocytes or B cells. The mature B cells migrate into outer or superficial cortex of the germinal follicles and medullary cords of lymph nodes and lymphoid follicles of spleen. These sites are known as *bursa-dependent* or *thymus-independent* areas. Following appropriate antigenic stimulation, B lymphocytes transform into plasma cells and secrete antibodies which constitute the *humoral immunity* or *antibody-mediated immunity (AMI).*

B. PERIPHERAL LYMPHOID ORGANS

I. Lymph nodes

The lymph nodes are small bean-shaped or oval structures which form part of the lymphatic network distributed throughout the body.

Structural characteristics of lymph node (Fig. 3.4-1)

- *Capsule* of connective tissue covers each lymph node. From the capsule, trabeculae penetrate into the lymph node.

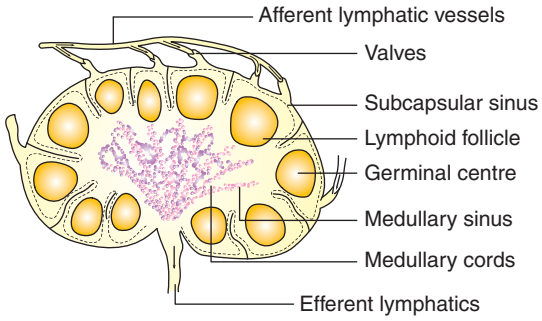


Fig. 3.4-1 Structure of a lymph node.

- *Afferent lymphatics* enter into each lymph node at its convex surface and drain into the peripheral subcapsular sinus.
- *Efferent lymphatics* leave the lymph node at the concavity (hilum) as a single large lymph vessel.
- *Microscopically,* the lymph node consists of two parts—peripheral cortex and central medulla.
- *Cortex* of the lymph node consists of several rounded aggregates of lymphocytes called *lymphoid follicles,* representing *B-cell area* of the node.
- *Paracortex* is deeper part of cortex, i.e. the zone between the peripheral cortex and the inner medulla and represents the *T-cell area* (the *bursa-independent area).*
- *Medulla* is predominantly composed of cords of plasma cells and some lymphocytes *(medullary cords).* The medullary cords contain B lymphocytes and along with the lymphoid follicles constitute the *bursa-dependent areas* of the lymph node.

Functions of lymph nodes

1. ***To mount immune response in the body.*** The bulk of antigens are processed and antibody production occurs in the lymphoid follicles (for details see page 139).
2. ***The lymph nodes constitute a series of inline filters.*** Lymph must pass through at least one lymph node before mixing with the blood stream. Over 99% of the lymph passes through the lymph sinuses and only 1% penetrates the lymphoid follicles.

II. Spleen

The spleen is the largest lymphoid organ of the body. Under normal conditions, the average weight of the spleen is about 150 g.

Structural characteristics (Fig. 3.4-2)

Capsule of the connective tissue surrounds the spleen. From the capsule, the connective tissue *trabeculae* extend into the pulp of the organ and serve as a supportive network.

Grossly on cut section, the spleen consists of homogeneous, soft, dark red mass called *red pulp.* In the red pulp are seen scattered white nodules called *white pulp* (malpighian bodies).

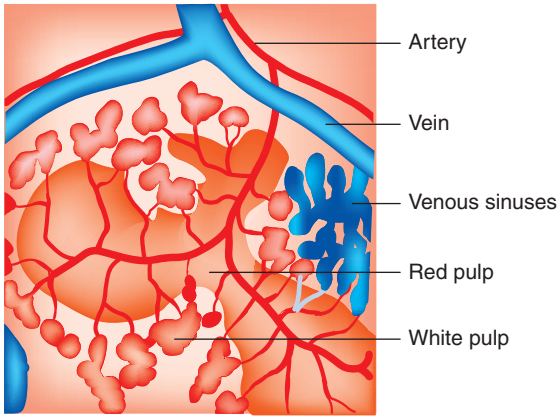


Fig. 3.4-2 Structure of spleen.

Microscopically, the structural characteristics are:

- *Red pulp* consists of the thin walled *blood sinuses* with *splenic cords* between them. The splenic cords consist of collection of lymphocytes and macrophages arranged in cords in the fine network of reticular cells and fibres. This finer network forms a filtering bed that filters out red cells, white cells and platelets passing through the red pulp.
- *White pulp* consists of lymphocytes surrounding an eccentrically placed central artery. These periarteriolar lymphocytes are mainly *T lymphocytes.* In addition, the white pulp also contains the lymphoid follicles composed principally of *B lymphocytes.*

Functions of spleen

1. Role in immune response. The spleen is an active site for production of T and B lymphocytes and antibodies.

2. Role in removal of old RBCs, WBCs and platelets. Tissue macrophages present in the spleen like other components of the MPS play an important role in the removal of old RBCs, WBCs and platelets.

3. Role in haematopoiesis. During fourth and fifth month of fetal life erythropoiesis occurs in the spleen.

4. Role in iron metabolism. Spleen macrophages have a special ability to recycle the iron, liberated from the phagocytosed RBCs, for synthesis of fresh haemoglobin in the bone marrow.

5. Role as a reservoir. Spleen serves as a reservoir for the mobilization of RBCs in some animals like cat and dog.

6. Role in regulating portal blood flow. The vasculature of spleen also plays a role in regulating the portal blood flow.

APPLIED ASPECTS

In conditions like congenital spherocytosis, autoimmune haemolytic states and hypersplenism, splenectomy is of therapeutic value because spleen is the major site of RBC destruction.

III. Mucosa associated lymphoid tissue

Mucosa associated lymphoid tissue including: tonsils, adenoids and Peyer's patches of small intestine are known as gut associated lymphoid tissue. Peyer's patches are small patches of organized lymphoid tissue along the intestine containing B lymphocytes (in germinal centre) and T lymphocytes. They play a primary role in defence against infectious organisms entering via the gastrointestinal tract.

IMMUNITY

Immunity refers to resistance of the body to pathogens and their toxic products. It can be classified as:

I. *Innate immunity*
  1. Non-specific
  2. Specific innate immunity
II. *Acquired immunity*
  1. Active acquired immunity
  2. Passive acquired immunity

INNATE IMMUNITY

- *Innate or natural immunity* is the inborn capacity of the body to offer resistance to pathogens and their toxic products. It is due to genetic and constitutional make up of an individual.
- It may be *specific* (against a particular organism) or *non-specific.*

Mechanisms of innate immunity

1. ***Mechanical barrier*** against invading microorganism is provided by the intact skin and mucosa in the body.
2. ***Surface secretions*** constitute one of the important mechanisms of innate immunity. These include:
   - *Secretions* from the sebaceous glands of skin contain both saturated and unsaturated fatty acids that kill many bacteria and fungi.
   - *Saliva,* constantly produced in the mouth cavity, has an inhibitory effect on many micro-organisms.
   - *Gastric juice* and highly acidic environment of stomach may hydrolyze microbial invaders.
   - *Tears* poured in the conjunctival sac mechanically wash away the particles and a hydrolytic enzyme, lysozyme present in the tears can destroy most of the micro-organisms.
3. ***Humoral defence mechanisms*** provide innate immunity by the non-specific microbicidal substances present in the body fluids. A few examples are:
   - *Lysozyme* is found in high concentration in most tissue fluids except cerebrospinal fluid, sweat and urine. It is a mucolytic enzyme which kills micro-organisms by splitting sugars of the structural mucopeptide of their cell wall.
   - *Basic polypeptides* containing non-specific microbicidal activity include leukins, arginine and lysine containing proteins protamine and histone.
   - *Complements* have lytic and several other effects on the foreign substances (see page 141).
   - *Interferons* are antiviral substances produced by the cells stimulated by live or killed viruses. The α and β interferons are part of the innate immunity.
4. ***Cellular mechanisms of defence,*** which provide non-specific innate immunity are:
   - *Phagocytes,* i.e. neutrophils and the monocyte–macrophage system cells constitute the most important non-specific cellular defence against the invading micro-organisms.
   - *Natural killer (NK) cells* refer to a subpopulation of lymphocytes which provide non-specific cellular defence against viruses, tumour cells and other infected cells.
   - *Eosinophil* granules contain enzymes and toxic molecules that act against larvae of helminths.

ACQUIRED IMMUNITY

The resistance that an individual acquires during his lifetime is known as acquired immunity. It is antigen-specific and may be antibody-mediated or cell-mediated. It is of two types: active and passive.

1. ACTIVE IMMUNITY

Active immunity is acquired by the synthesis of antibodies (humoral immunity) and production of immunocompetent cells (cell-mediated immunity) by the individual's own immune system in response to an antigenic stimulation.

Natural and artificial active immunity

Active immunity can be induced naturally or artificially.

(i) Natural active immunity. Natural active immunity results either from a subclinical or a clinical infection.

(ii) Artificial active immunity. Artificially, the active immunity is induced by introducing antigens in the body in the form of vaccines and this process is called active immunisations. The vaccines are preparations of live or killed micro-organisms or their products. Examples of vaccines are:

- Bacterial vaccines
  – Live: BCG vaccine for tuberculosis
  – Killed: TAB vaccine for typhoid
- Bacterial product vaccines
  – Tetanus toxoid
  – Diphtheria toxoid

- Viral vaccines
  - Live: Sabin vaccine for poliomyelitis, MMR vaccine for measles, mumps and rubella.
  - Killed: Salk vaccine for poliomyelitis, neural and non-neural vaccines for rabies.

2. PASSIVE IMMUNITY

Passive immunity refers to the immunity that is transferred to a recipient in a ready-made form. Here the individual's immune system does not play an active role.

Natural and artificial passive immunity

(i) Natural passive immunity. It is transfer of ready-made antibodies from the mother as:

- *In fetus,* the IgG antibodies are transferred from the mother through the placenta.
- *After birth,* immunoglobulins are passed to the newborn through the breast milk. Human colostrum is rich in IgA antibodies which are resistant to digestion in stomach and small intestine.

Passively transferred antibodies are generally against all common infectious diseases in the locality. These confer immunity on the neonate up to three months of age. Therefore, most paediatric infections are more common after the age of three months when maternal immunoglobulins disappear. By active immunization of mother during pregnancy the immune status of the neonate can be improved. Therefore, immunization of pregnant women with tetanus toxoid is recommended in countries where neonatal tetanus is common.

(ii) Artificial passive immunity. Artificially, passive immunity can be transferred to the recipients by injecting ready-made antibodies. This is done by the administration of hyperimmune sera.

Examples of artificial passive immunity include injection of:

- Antitetanus serum,
- Antidiphtheritic serum and
- Antigas gangrene serum.

The passively administered antibodies are removed by metabolism. Therefore, immunity conferred is short lived.

DIFFERENCES BETWEEN ACTIVE AND PASSIVE IMMUNITY

Differences between active and passive immunity are summarized in Table 3.4-1.

Table 3.4-1 Differences between active and passive immunity

Active immunity	Passive immunity
1. ***Production*** Antibodies are produced by the body's own immune system in response to antigens introduced naturally or artificially in the body.	Received passively by the host. No participation of host's immune system. It is conferred by administration of ready-made antibodies naturally or artificially in the body.
2. ***Negative phase*** Negative phase is present in the development of active immunity during which the immunity is transiently lowered. This is due to antigen combining with the pre-existing antibodies and lowering their level.	There is no negative phase in passive immunity as antigens are not injected.
3. ***Latent period*** Active immunity develops after a latent period varying from 4 days to 4 weeks. This is the time required for generation of antibodies and immunocompetent cells.	There is no latent period. Passive immunity is effective immediately.
4. ***Secondary response*** Due to immunological memory, the secondary response, i.e. response to antigen introduced. Second time is more enhanced.	There is no immunological memory. Rather subsequent administration of antibodies is less effective due to immune elimination.
5. ***Duration*** Active immunity is long lasting.	Passive immunity is short lasting.
6. ***Effectivity*** Active immunity is more effective and confers better protection.	Passive immunity is less effective and provides inferior immunity.
7. ***Applicability in immunodeficient individuals*** Active immunity is not applicable in the immunodeficient individuals.	Passive immunity is applicable in the immunodeficient hosts.

ANTIGENS

DEFINITION

Antigen. Antigens are substances that can stimulate an immune response in the body. Most antigens are proteins, but some are carbohydrates, lipids and nucleic acids. The specificity of an antigen is due to specific areas of its molecule called *determinant sites* or *epitopes* (Fig. 3.4-3).

SOME FACTS ABOUT ANTIGENICITY

Immunogenicity, i.e. ability of an antigen to stimulate an immune response.

Antigen specificity is determined by chemical grouping and acid radicals.

Species specificity. Tissues of all individuals in a species contain species-specific antigens. However, some degree of cross-reactivity is seen between antigens from related species.

Isospecificity. Isoantigens are the antigens which are found in some but not all members of a species. The best example of isoantigens is human blood group antigens on the basis of which all humans can be divided into blood groups A, B, AB and O. Each of these groups may be further divided into Rh-positive and Rh-negative. This carries clinical importance in blood transfusion, isoimmunization during pregnancy and disputed paternity.

HISTOCOMPATIBILITY ANTIGENS

Histocompatibility antigens refer to the antigens present on the plasma membrane of cells of each individual of a species. These antigens are encoded by genes known as histocompatibility genes, which collectively constitute major histocompatibility complex (MHC). These are located on the short arm of chromosome 6. MHC present on the surface of leucocytes is known as human leucocyte associated antigens (HLA). These have been studied extensively in organ transplantation. No two persons except identical twins have the same MHC proteins.

There are three subclasses of MHC genes: class I, II and III.

MHC class I. Molecules are found on the surface of virtually all the cells of the body excluding red blood cells (Fig. 3.4-4).

MHC class II. In man, MHC class II antigens are only found on immunologically reactive cells, such as B lymphocytes, macrophages, monocytes and activated T lymphocytes.

MHC class III. The genes coding for the complement components of the classical (C_2 and C_4) and the alternative (properdin factor B) pathway also reside in the MHC genes complex located between MHC class I and class II regions.

HLA tissue typing

Histocompatibility typing or the so-called HLA tissue typing refers to the detection of the MHC class I and MHC class II antigens. HLA typing is used to determine HLA compatibility prior to organ/tissue transplantation from one individual to another within a species.

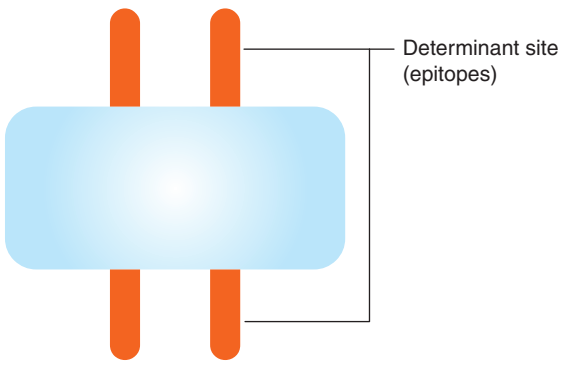


Fig. 3.4-3 Structure of an antigen.

ANTIBODIES

Antibodies or immunoglobulins (Igs) are γ globulins which are produced in response to an antigenic stimulation. These react specifically with the antigens which stimulated their production. All antibodies are immunoglobulins, but all immunoglobulins are not antibodies. Igs have been divided into five distinct classes or isotypes, namely IgG, IgA, IgM, IgD and IgE.

Characteristic features of various immunoglobulins are summarized in Table 3.4-2.

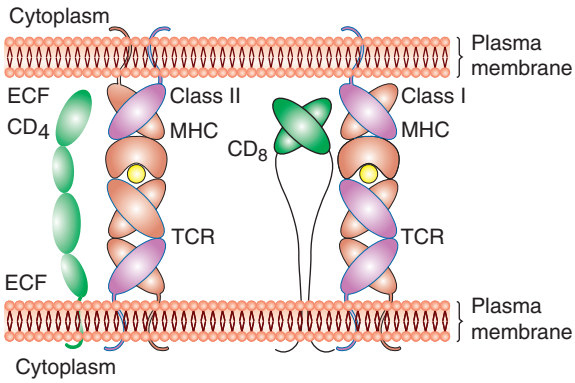


Fig. 3.4-4 Structure of MHC proteins and their relation to CD_4 and CD_8.

Table 3.4-2 Characteristic differentiating features of various immunoglobulins

Feature	IgG	IgA	IgM	IgD	IgE
• ***Structural characteristics***					
– Structural unit	Monomer	Monomer SIgA is dimer	Pentamer	Monomer	Monomer
– Heavy chain class	γ1, γ2, γ3, γ4	α1, α2	μ	δ	ε
– Light chain class	κ or λ	κ or λ	κ or λ	κ or λ	
• ***Additional chain***	–	J, SC	J	–	–
• ***Molecular weight (KDa)***	150	160–385	900	180	190
• ***Carbohydrate content (%)***	3	8	12	13	12
• ***Serum concentration (ng/mL)***	12	2	1.2	0.03	0.00004
• ***Half-life (days)***	21	6	5	3	2
• ***Secretion from serous membranes***	No	Yes (SIgA)	No	No	Yes
• ***Placental transfer***	Yes	No	No	No	No
• ***Heat stability (56°C)***	Yes	Yes	Yes	Yes	No
• ***Complement fixation***	Classical pathway	Alternative pathway	Classical pathway	None	None
• ***Binding to tissue***	Heterogeneous	None	None	None	Heterogeneous
• ***Role in the body***	Protects the body fluids	Protects the body surfaces	Protects the blood stream	Role not known	Mediate-type I hypersensitivity

STRUCTURE OF ANTIBODY

IgG has been studied extensively and serves as a model of basic structural unit of all Igs. An Ig is a Y-shaped molecule made of four polypeptide chains: two heavy (H) and two light (L). These are held together by the disulphide bonds (Fig. 3.4-5).

Heavy chains

- Heavy (H) chains have a molecular weight of 50,000 Da.
- H chains are structurally and antigenically distinct for each class of Ig and are named as:
  - α (alpha) in IgA
  - δ (delta) in IgD
  - ε (epsilon) in IgE
  - γ (gamma) in IgG
  - μ (mu) in IgM
- The NH_2 terminal half of each chain has a variable sequence of amino acids and is called the *variable region.* In heavy chain, it is designated as VH.
- The COOH terminal of each chain has a relatively constant sequence and is called the *constant region.* In heavy chain, it is designated as CH.
- Two H chains are always identical in a given molecule. In IgG, each H chain contains 440 amino acids.
- Intrachain sulphide bonds fold each chain into incomplete loops. In the heavy chains, there are 4 loops of 110 amino acids and each loop forms a globular domain. One domain designated as VH lies in the variable region and CH_1, CH_2 and CH_3 lie in the constant region (Fig. 3.4-5).

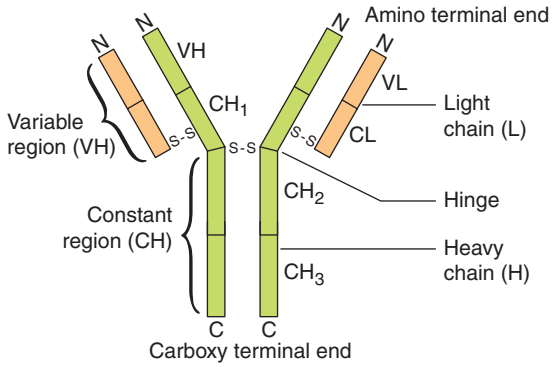


Fig. 3.4-5 Basic structure of an antibody showing the arrangement of heavy and light chains, and its variable and constant domains.

Light chains

- Molecular weight of light (L) chains is 25,000 Da.
- L chains are of two types: κ (Kappa) and λ (lambda). A molecule of Ig may have either κ or λ chains, but never both together. κ and λ chains occur in a ratio of about 2:1 in human serum.
- Similar to H chains, the variable region of L chains is designated as VL and constant region is designated as CL.
- Like H chains, the L chains are also always identical in a given molecule. In IgG, each L chain contains 220 amino acids.
- In the light chain, there are two loops each containing 110 amino acids. Each loop forms a globular domain. One

domain in variable region is designated as VL and the other in the constant region is designated as CL (Fig. 3.4-5).

FUNCTIONS OF IMMUNOGLOBULINS

From the available information, the specific functions of various immunoglobulins appear as:

- ***IgG*** protects the body fluids,
- ***IgA*** protects the body surfaces,
- ***IgM*** protects the blood stream,
- ***IgE*** mediates type I hypersensitivity and
- ***IgD's*** role is not clearly known.

DEVELOPMENT OF IMMUNE RESPONSE

Development of immune response implies *development of acquired active immunity* in the body. The immune system of the body responds to an antigen by two ways:

1. Humoral or antibody-mediated immunity and
2. Cell-mediated immunity.

DEVELOPMENT OF HUMORAL IMMUNITY

The humoral immunity is mediated by antibodies and so is also called antibody-mediated immunity. The antibodies are produced by the plasma cells which in turn are produced by the B lymphocytes.

ROLE OF HUMORAL IMMUNITY

1. The humoral immunity provides defence against most extracellular bacterial pathogens and viruses that infect through the respiratory and intestinal tract.
2. It participates in immediate hypersensitivity reactions of type I, II and III.
3. Humoral immunity is also associated with certain autoimmune diseases.

TYPES OF HUMORAL IMMUNE RESPONSE

The antibody response to stimulation by an antigen is of two types (Fig. 3.4-6):

- Primary humoral response and
- Secondary humoral response.

Primary response refers to the response of the body's immune system to an antigen which is introduced into the body for the first time. Always there is a latent period varying from 4 days to 4 weeks before the primary response in the form of a rise in the serum antibodies titre can be detected.

Secondary response refers to the response of the body's immune system to an antigen which is introduced into the body on a second occasion. Such a response occurs more quickly and more abundantly. This is because of the fact that the immune system is liable to retain the memory of a prior antigenic exposure for long periods (immunological memory) and produce enhanced response when encountered with the same antigen for the second time.

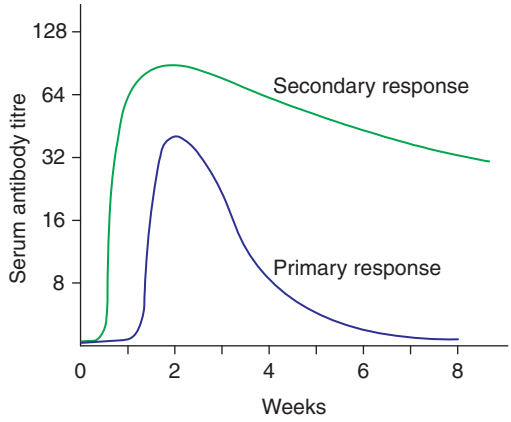


Fig. 3.4-6 Response of body immune system to an antigen.

STAGES OF HUMORAL IMMUNE RESPONSE (FIG. 3.4-7)

1. Antigen processing and presentation

Once the antigen enters the body, it is phagocytosed by the macrophages (non-specific response). Phagocytosed material is broken down into polypeptide fragments. The antigen polypeptide fragments then combine with the MHC II present in the macrophages and move to the cell surface. This is called *processing of antigen.* The processed antigen is then presented to immunocompetent lymphocytes by the macrophages. So, the macrophages are also called *antigen presenting cells.*

2. Recognition of antigen by lymphocytes

The lymphocytes possess the antigen recognition receptors. These include the membrane-bound (surface) immunoglobulins (mIgs or sIgs) in B lymphocytes and T cell receptors in the T lymphocytes. These receptors serve as specific surface receptors, recognizing and interacting with only single antigenic determinant on the antigen presented to the lymphocytes. The process of binding of processed antigen to specific receptors on the surface of lymphocytes is called *recognition of antigen by lymphocytes.* Thus many million different T and B lymphocytes, each with the ability to respond to particular antigen are present in the body.

3. Lymphocyte activation

The lymphocytes that have combined with antigen are activated, i.e. the lymphocytes become larger and look like a lymphoblast. This is known as *blast transformation.* Activated B lymphocytes and helper T cells (CD_4 cells) play a major role in

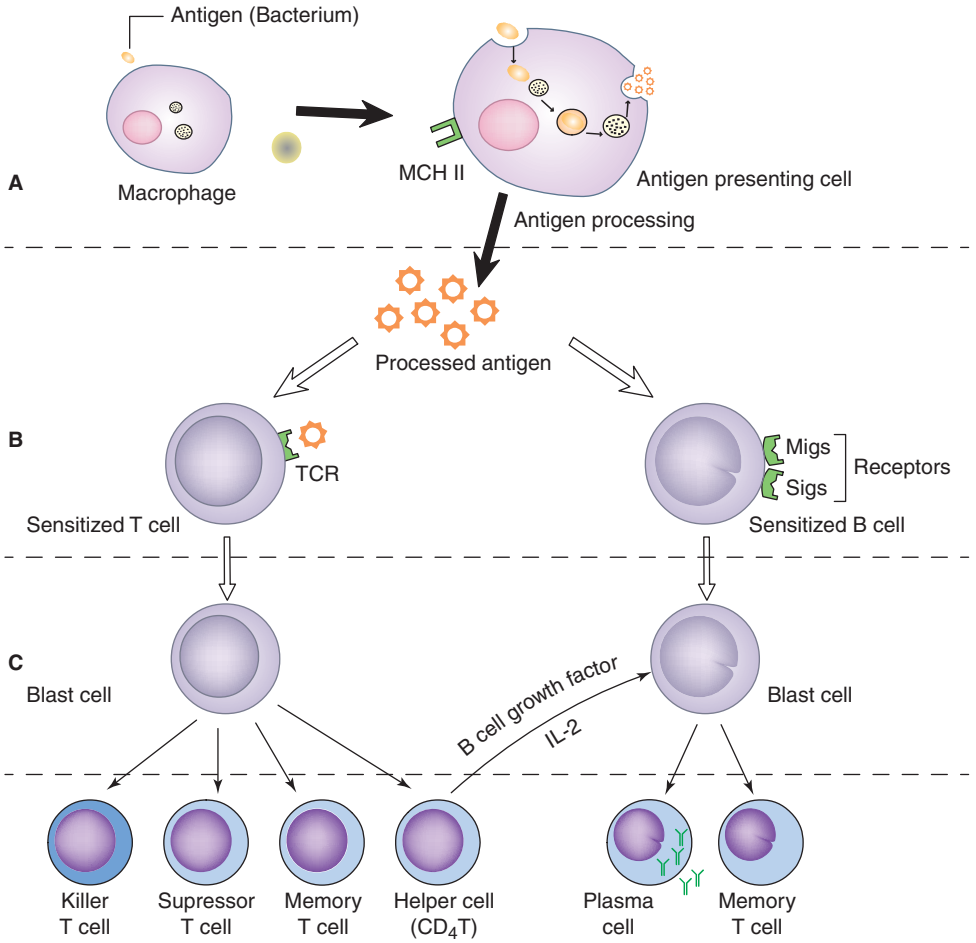


Fig. 3.4-7 Broad outline of development of humoral immune response: A, antigen processing and presentation to immunocompetent cell; B, recognition of antigen by the lymphocytes and C, activation of lymphocytes (blast transformation).

humoral immunity. The macrophages liberate IL-1 and cause further activation of B lymphocytes and helper (CD_4T) cells.

Activation of T lymphocytes

The activated helper T cells (CD_4) secrete interleukins 2 (IL2) and B cell growth factor which further promote proliferation of B lymphocytes and their transformation into the plasma cells. This phenomenon is called *T-B Co-operation.*

Activation of B lymphocytes

After receiving co-operation from the T helper cells (T-B co-operation), the B lymphocytes proliferate and transform into:

- Plasma cells and
- Memory B cells.

Role of plasma cells. When B lymphocyte is converted to plasma cell, its cytoplasm expands. It is filled with a granular endoplasmic reticulum. The plasma cells secrete antibodies.

Role of memory B cells. A small portion of the activated B lymphocytes only proliferate and transform into small sized memory B cells, which occupy the lymphoid tissue throughout the body. Memory B cells have a long lifespan and remain inactive. When the body is exposed to the same antigen for the second time, they are able to recognize it and become active, i.e. they are responsible for secondary response of antibodies.

4. Production of antibodies

The plasma cells so formed secrete antibodies which are also called Igs. Each plasma cell produces about 2000 molecules of antibodies per second. A plasma cell secretes an antibody of a single specificity of a single antibody class and a single light chain type. However, in primary antibody response, plasma cell produces IgM initially and later it may switch onto IgG production.

The antibody produced in response to an antigen is highly specific, i.e. it reacts only with the antigen which has evoked its production. How this is possible is not known exactly. Various theories have been put forward to explain this immune specificity, but ***Clonal selection theory*** put forward by Burne in 1957 is more widely accepted.

- According to this theory, during immunological development a large number of B lymphocytes capable of

reacting with different antigens are formed. Each specific B cell multiplies and establishes a population of genetically and immunologically identical B cells called a *clone.* An antigen only has to select and stimulate its specific clone. That is why, this theory is named clonal selection theory.

- This theory also states that cells with immunological reactivity with self-antigens are eliminated during the embryonic life. Such clones are known as the forbidden clones. Their persistence or development in the later life leads to autoimmunity.

5. Inactivation of antigen or attack phase or effector phase of immune response

This is the last phase of immune response and involves the inactivation of an antigen by the antibodies. Antibodies act on the invading antigen in two ways:

(i) Direct attack on the invading agents

Antibodies can inactivate the invading agent by the following reactions:

Agglutination. By this reaction, large number of particles (bacteria or red cells) with antigens on their surface are bound together to form a clump. Clumping increases the susceptibility to phagocytosis.

Precipitation. In this reaction, antigen–antibody complex forms insoluble precipitate.

Neutralization. Antibodies cover the toxic sites of antigen and neutralize them.

Cytolysis. Antibodies attach the membranes of cellular agents thereby causing rupture of cells.

(ii) Attack on the antigen through complement system

The complement system includes 11 enzymatic proteins which are named as C1–C9, B and D. All these are present in the blood as plasma proteins.

These are also present in the tissue fluid. The complement system acts in two ways (Fig. 3.4-8):

(a) Classical pathway. This is activated by the antigen–antibody reaction. When an antibody binds with an antigen, a specific reactive site on the constant portion of the antibody becomes uncovered where the protein C1 binds and thus gets activated. The activated C1 in turn activates the other complements in a series of cascade reactions.

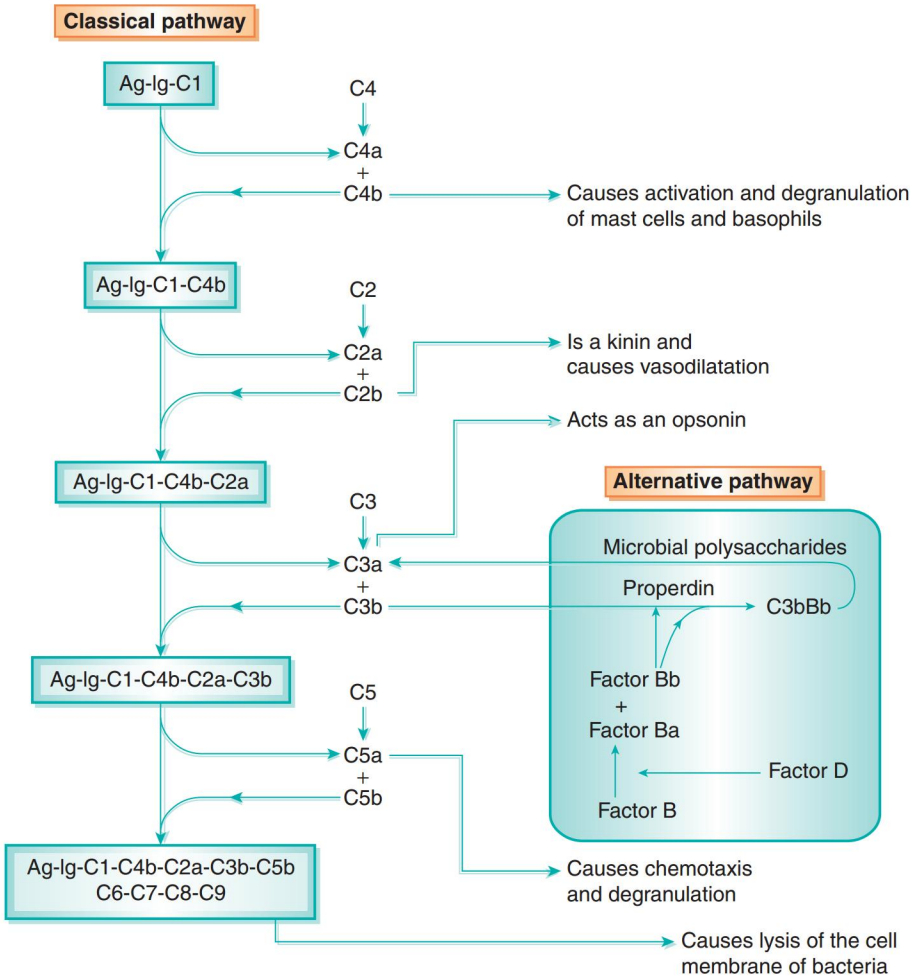


Fig. 3.4-8 Broad outline of classical and alternative pathway of complement system.

The products of complement activation cause following effects:

Opsonization. The activated C3a product acts as an opsonin.

Lysis, i.e. destruction of bacteria by rupturing the cell membrane. Membrane-attack complex is formed by C5b–C6–C7–C8–C9.

Agglutination, i.e. clumping of bacteria and RBCs.

Chemotaxis, i.e. attraction of leucocytes to the site of antigen–antibody reaction. Chemotaxis is enhanced by C5b–C6–C7 complex.

Neutralization, i.e. covering the toxic sites of the antigenic products.

Activation and degranulation of mast cells and basophils is caused by C4b. This releases factors producing vasodilatation and chemotaxis. Vasodilatation increases capillary permeability, therefore, plasma proteins enter the tissues and the antigenic products are inactivated.

(b) Alternative pathway or properdin pathway. In the alternative pathway, the complement system is activated by binding of the protein in circulation (properdin) with a polysaccharide present in the cell wall of invading organism, i.e. bacteria (endotoxin) and yeast cell wall (zymogen). This binding triggers reactions that activate C3 and C5, which ultimately attack the antigenic products of invading organisms.

DEVELOPMENT OF CELLULAR IMMUNE RESPONSE

The cellular immunity refers to the specific acquired immunity which is accomplished by the effector T cells and macrophages (Figs 3.4-9 and 3.4-10). It is also called cell mediated immunity (CMI).

ROLE OF CELLULAR IMMUNITY

1. Cellular immunity protects the host against fungi, most of the viruses and intracellular bacterial pathogens like *Mycobacterium tuberculosis*, *M. leprae* and Brucella.
2. It participates in the allograft rejection and graft versus host reaction.
3. CMI participates in delayed hypersensitivity reaction.
4. CMI is also associated with certain autoimmune diseases.
5. It provides an immunological surveillance and immunity against cancer (tumour immunity).

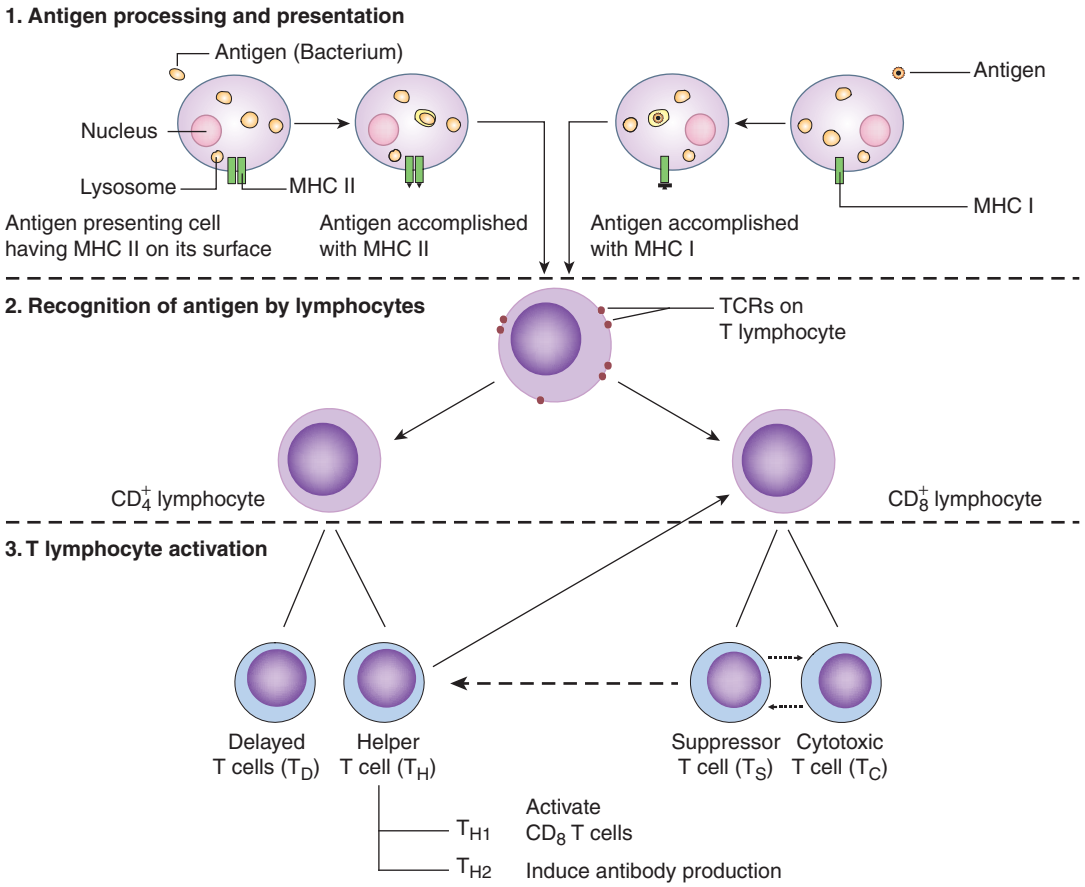


Fig. 3.4-9 Development of cellular immune response.

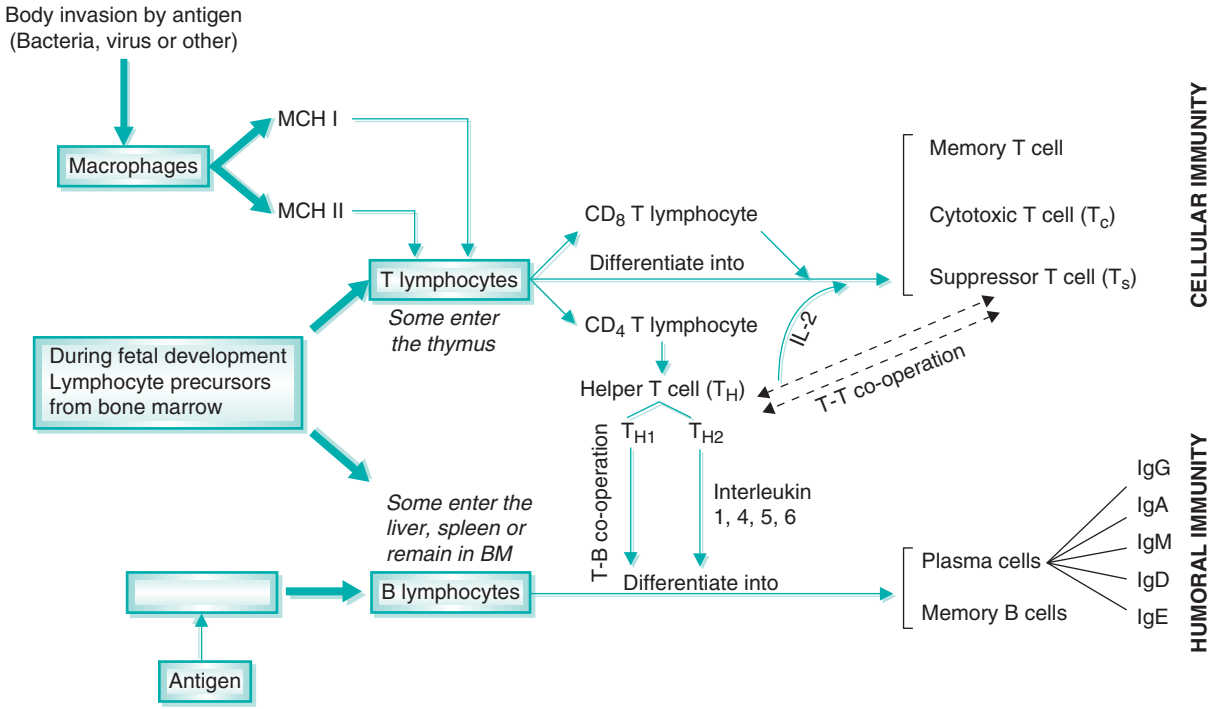


Fig. 3.4-10 Summary of immune system.

TYPES OF CELLULAR IMMUNE RESPONSE

Like humoral immune response, the cellular immune response is also of two types:

Primary cellular response, which is produced by an initial contact with a foreign antigen and

Secondary cellular response, which is produced when the host is subsequently exposed to the same antigen. The secondary cell-mediated immune response is usually more pronounced and occurs more rapidly.

STAGES OF CELLULAR IMMUNE RESPONSE

1. Antigen processing and presentation

Antigen entering the host body is phagocytosed and degraded into polypeptide fragments by the antigen processing cells (APCs) which include macrophages and dendritic cells present in the peripheral lymphoid tissue. The antigen polypeptide fragments then become associated with the MHC antigen and are expressed on the surface of APC.

2. Recognition of antigen by lymphocytes

T lymphocytes possess the antigenic recognition receptors known as *T cell receptors*. These receptors serve as a specific surface receptors recognizing and interacting with only single antigenic determinant on the antigen presented to lymphocytes. Further, the mature T lymphocytes can be differentiated into two antigenic subtypes depending on the ensemble of their surface antigens: CD_4^+ cells and CD_8^+ cells.

- CD_8^+ cells recognize the combination of foreign antigen and class I MHC antigen, and
- CD_4^+ cells recognize the combination of foreign antigen and class II MHC antigen.

3. T lymphocyte differentiation (activation)

The CD_8^+ type of T lymphocytes after combining with foreign antigen–MHC-I complex are activated and differentiated into:

- Cytotoxic T cells (T_C cells) and
- Suppressor T cells (T_S cells).

CD_4^+ type of T lymphocytes after combining with foreign antigen–MHC-II complex are activated and differentiate into:

- Helper T cells (T_H cells) and
- Delayed-type hypersensitivity T cells (T_D cells).

T-T co-operation. The differentiation of T lymphocytes into T_H, T_C, T_S and T_D cells is interdependent. This interdependence is called T-T co-operation.

Release of differentiated T cells. The differentiated T lymphocytes so formed are released into the lymph and then enter the blood through which they are distributed throughout the body.

T lymphocyte memory cells are also formed. These spread throughout the lymphoid tissues of entire body. Therefore, on subsequent exposure to the same antigen, release of T cells occurs far more rapidly and much more powerfully than in first response (secondary response).

4. Attack phase of cell-mediated immunity

Role of cytotoxic T cells

Cytotoxic T cells (T_C cells) and NK cells are responsible for the attack phase of cell-mediated immunity.

T_C cells have some receptor protein on their outer membrane and bind antigen bearing cells (target cells) tightly and destroy them by the following mechanisms:

(i) ***Perforin-mediated killing.*** The T_C cells after binding with the target cell secrete hole-forming protein called *perforin*. The perforins literally punch round holes in the membrane of target cells in the presence of extracellular calcium (calcium dependent lysis). The pores so formed cause cell death by disrupting cell homeostasis.

(ii) ***Lysis through cytotoxic substances***. After binding with target cells the T_C cells enlarge and release cytotoxic substances.

(iii) ***Induction of apoptosis.*** T_C cells secrete tumour necrosis factor β (TNF-β) which increases the Ca^{2+} permeability of antigen-bearing cell. The increased intracellular calcium activates enzymes that cause degradation of nucleus producing apoptosis.

Role of helper T cells (T_H cells)

Helper T cells are of two types: T_{H1} and T_{H2}.

(i) ***Helper T_1*** (T_{H1}) *cells* play their roles by secreting three cytokines:
- Interleukin-2 (IL-2)
- Gamma interferon (IFN-γ)
- Tumour necrosis factor-beta (TNF-β)

(ii) ***Helper T_2*** (T_{H2}) *cells* secrete interleukins 4, 5, 6, 10 and 13 and are primarily with activation of B lymphocytes to produce antibodies (T-B co-operation).

Role of suppressor T lymphocytes (T_S cells)

- T_S cells regulate the activity of cytotoxic T cells.
- T_S cells also play an important role in preventing the cytotoxic T cells from destroying the body's own tissue along with the invading organism.
- T_S cells also suppress the activities of helper T cells.

CYTOKINES

Cytokines are small protein molecules, which act like hormones to regulate immune response.

The cytokines are secreted not only by lymphocytes and macrophages, but also by endothelial cells, neuroglial cells and other types of cells. Broadly, cytokines can be grouped as:

Interleukins (IL). These are the principal cytokines and include IL-1 to IL-13.

Other cytokines include:

- Chemokines,
- Growth factors,
- Colony stimulating factors,
- Tumour necrosis factors (TNF-α and TNF-β) and
- Interferons (IFN).

The characteristics of interleukins and cytokines are summarized in Table 3.4-3.

Chemokines are the substances that attract the neutrophils and other white blood cells to the area of immune response or inflammation. About 40 chemokines have been identified. The receptors of the chemokines are serpentine and act via G proteins. Chemokines also play role in cell growth and angiogenesis.

OTHER IMMUNE MECHANISM RELATED ASPECTS

IMMUNE TOLERANCE

Types of immune tolerance

Immune tolerance may be defined as a state of unresponsiveness to an antigen. It occurs in two forms: natural and acquired.

1. ***Natural tolerance*** refers to the non-responsiveness to a self-antigen. During embryonic development, when immune system is immature, any antigen which comes in contact with the immature immune system is recognized as a self-antigen. Therefore, it does not evoke any response in later life when body is exposed to the same antigen.
2. ***Acquired tolerance*** means unresponsiveness to a potential antigen. It results due to impairment of immune system, hence there is lack of responsiveness to the potential antigens.

Mechanism of tolerance

Immunotolerance can arise by three possible mechanisms:

- Clonal deletion,
- Clonal anergy and
- Suppression.

1. Clonal deletion. During embryonic life, clones of B and T cells are formed. These B and T cells possess receptors, which recognize the antigens and are selectively deleted or eliminated and therefore, not available to respond on the subsequent exposure to that antigen in later life.

2. Clonal anergy. Clones of B and T cells receptors which recognize self-antigen might remain, but cannot be activated. This is to be referred as clonal anergy.

Table 3.4-3 Main characteristics of human interleukins and other immunoregulatory cytokines

Type of cytokine	Cell source	Effects
I. Interleukins (IL)		
IL-1 α and β	Macrophages and other antigen processing cells (APCs)	B cell proliferation. Igs production. Stimulation of T cells. Inflammation fever.
IL-2	Activated helper cells (T_{H2}), cytotoxic cells (T_c) and Natural killer cells (NK)	Proliferation of activated T cell. B cell proliferation, Igs expression.
IL-3	T lymphocytes	Growth of early progenitor cell.
IL-4	T_{H2} and mast cells	Eosinophil growth and its function. B cell proliferation. Igs expression. MHC class II expression. Proliferation of T_{H2} and T_c cells. Inhibition of production of inflammatory cytokines.
IL-5	T_{H2} and mast cells	Growth of eosinophils and its function.
IL-6	Activated T_{H2} cells APC and other somatic cells	Act with IL-I and TNF to stimulate T cells. Proliferation of B cells and Igs production. Stimulates thrombopoiesis.
IL-7	Thymic and bone marrow stromal cells	Lymphopoiesis (T and B cells).
IL-8	Macrophages	Stimulates neutrophil activity and promote their accumulation.
IL-9	From cultured T cells	Stimulates haematopoietic and thymopoietic factors.
IL-10	Activated Helper cells (T_{H2}), TCD_8^+, B Lymphocyte and macrophages	Inhibition of cytokine production. Stimulates B cells and antibodies production and its functions. Suppresses cell-mediated immunity. Causes growth of mast cells.
IL-11	Stroma cells	Stimulates haematopoiesis and thrombopoiesis.
IL-12	Macrophages and B cells	Stimulates proliferation of cytotoxic T cells and killer cells (T_c and NK).
IL-13	T_{H2} cells	Promote cell-mediated immune response. B cell proliferation, IgG expression, class II MHC expression. Proliferation of T_{H2} and T_c cells and their function. Inhibition of production of inflammatory cytokines.
II. Other cytokines	Activated macrophages	IL-I type effects.
TNF-α		Cause vascular thrombosis and necrosis of tumour cells.
TNF-β	Activated T_{H1} cells	Vascular thrombosis and tumour cell necrosis.
Interferon (IFN-α & IFN-β)	Macrophages, neutrophils and other somatic cells	Antiviral effects. Stimulate class II MHC cells. Activation of macrophages and NK cells.
IFN-γ	Activated T_{H1} and NK cells	Antiviral effect. Activation of class I MHC cells and class II MHC cells. Promote cell-mediated immunity.
TGF-β	Activated T lymphocytes, platelets, macrophages and somatic cells	Anti-inflammatory effect by suppressing cytokine production and MHC-II cells. Inhibit proliferation of macrophages and lymphocytes. Proliferation of B cells. Healing (by stimulating fibroblast cells).

3. Suppression. Clones of B and T cells expressing receptors that recognize self-antigen are preserved and capable for recognition of antigen when activated. However, immune response might be inhibited through active suppression.

Tolerance to fetus

Fetus is genetically different from the mother and thus it should evoke an immune response in the mother. However, it usually never happens and it is considered to be the best example of immune tolerance. Various factors which prevent immunological response in a mother against its fetus are:

1. *Placenta.* Placenta plays an important role by different ways:
   - Immediately after implantation, the trophoblast cells loose their immunogenic capacity due to decrease in MHC antigen density.
   - There is formation of mucoprotein coating on the cell surface and
   - Anti-MHC antibodies which are produced in the mother get absorbed into the placenta and their entry into fetal circulation is prevented. Thereby placenta acts as a shield against the immunological response.
2. α *fetoprotein.* During embryonic development, α fetoprotein (AFP) is produced which acts as an immunosuppressive agent.
3. *Progesterone.* During pregnancy, high levels of progesterone have got immunosuppression effect.
4. *Fetal T cells* get activated when fetus is exposed to mother T cells through placenta and suppress mother's T cells.

AUTOIMMUNITY

During fetal life, when many antigens are presented to immune system, they are recognized as self-antigens and antibodies and cytotoxic T cells are not produced. Therefore, tolerance to self-antigen is produced. However, sometimes body starts producing antibodies or T cells against self-antigen (own cells or tissue) leading to an autoimmune disease. Therefore, autoimmunity may be defined as immune response to self-antigen.

Mechanism of autoimmunity

The possible mechanisms involved in the development of autoimmunity are:

1. Forbidden clones. According to the clonal selection theory, antibody forming lymphocytes are formed against different antigens. In fetal life, lymphocytes are also formed against self-antigens, but get depleted. The clones of these cells are called forbidden clones and hence immune response does not occur against self-antigen. However, persistence of these clones or their development in later life by some mutations leads to autoimmunity.

2. Hidden antigen or sequestrated antigen. Certain self-antigens are present in the close system and never exposed to the immune system during fetal life. These are known as hidden antigens or sequestrated antigens, e.g. *lens protein* being enclosed by its capsule does not come in contact with blood, therefore, immunological tolerance against such antigens does not develop. When such antigens in later life somehow exposed to the immune system (accidental leak of lens protein during cataract surgery) leads to an immune response and damage the other eye also. Another example of a hidden antigen is *sperm antigen*. Injury to testes or viral infections (mumps) lead to leakage of sperm proteins into the circulation and thus evoke an immune response against own testes and orchitis occurs.

3. Neoantigen or altered antigen. Certain cells of the body undergo alterations due to the exposure to irradiations, drugs and sunlight, etc. and start producing immune response.

4. Cross-reacting antigen. Although antibodies are highly specific for a particular antigen, but in some cases they cross-react with other cells or body tissue. This phenomenon is called as *molecular mimicry* and these antigens are called cross-reacting antigens. For example, in rheumatic heart disease, heart is damaged by antibodies formed against streptococci.

5. Mutations. The body immune system becomes competent for self-antigen by certain mutations.

6. Unbalanced activity of T_H and T_S cells. It has been observed that the optimum antibody response always depends upon the balance activity of T_H and T_S cells. If somehow the activity of these cells is altered, i.e. overactivity of T_H cells and underactivity of T_s cells may result in autoimmunity.

Tissue transplant (graft)

Transplantation of tissue or organs from a donor to a recipient is known as grafts. The grafts are of following types:

- *Homograft (allograft)* refers to grafting of tissue from one person to another.
- *Autograft* refers to grafting of tissue from one part of body to another site in the same individual.
- *Xenograft (heterograft)* refers to the transplantation of tissue from one animal of species to another animal of different species.

The grafts are not usually taken up (graft ejection) except in the identical twins, i.e. autografts.

Graft rejection

Graft rejection occurs due to immune response to transplanted tissue due to histocompatibility antigen (HLA) present on the plasma membrane of the cell. For detail see page 137.

Prevention of rejection of grafts may be possible by an immune suppression. Immune suppression refers to reduction in an immunological response. Various types of immunosuppressive methods inhibit immune response of macrophages and B- and T-cells either by lowering phagocytosing capacity of macrophages or by production of antibodies and lymphokines.

Methods of immunosuppression. These have been grouped as physical, chemical and biological agents.

A. Physical immunosuppressive agents (methods):

1. *Irradiations.* This is the most common method for prolonged survival of transplants. Irradiations cause breakage in the nucleic acid chains of replicating cells.
2. *Surgical procedures* like thymectomy, splenectomy and thoracic duct drainage.

B. *Chemical methods* are non-specific suppressants and have limited effectiveness. This group includes following drugs:

1. *Corticosteroids* suppress the immune response by the following ways:
   - They impair the maturation of activated cells.
   - They suppress the production of antibodies.
   - Have anti-inflammatory effect and diminish the responsiveness of B and T lymphocytes.
   - They also inhibit production of IL-1 and IL-2.

   Corticoids though commonly used but has limited effectiveness due to their side effects, as prolonged use leads to hypertension, bone necrosis, cataract and mental disturbances.
2. *Cyclosporin or Tacrolimus (FK-506)* has been widely used as an immunosuppressive drug in organ transplants. This acts by inhibiting the production of IL-2. It also has adverse effects on liver and kidney.
3. *Cytotoxic drugs* such as *azathioprine* and *cyclophosphamide* act on various stages of nucleic acid synthesis and thus prevent replication of lymphocyte.

   Methotrexate, an antagonist of folic acid, produces competitive inhibition of an enzyme reductase (essential for synthesis of DNA). This drug is known as anticancer drug.

C. *Biological methods* include:

1. *Antigen-induced suppression.* This method is used for desensitization against an allergen. In this method, if the body is exposed to small doses of antigen for long time, then it can develop resistance to that antigen.
2. *Antilymphocytic serum* is used for depletion of T cell population. In this method, antilymphocytic serum is prepared from horse by injecting human lymphocytes. The antibodies present in the horse serum destroy body T cell pool, but antibodies production remains normal. The main drawback of this method is that ability to fight against viral infection is tremendously decreased.

Autoimmune diseases

Common autoimmune diseases include:

1. *Autoimmune anaemia.* For example:
   - Haemolytic anaemia: Antibodies react with its own RBCs.
   - Pernicious anaemia: Antibodies react against gastric mucosa.
2. *Thrombocytopenic purpura.* Autoantibodies react with self-platelets.
3. *Graves' disease.* Autoantibodies bind to the thyroid cells and stimulate them.
4. *Hashimoto's disease.* T cells react against with antigen on the thyroid cells.
5. *Insulin-dependent diabetes mellitus.* Antibodies damage the β cells (insulin producing cells) of the pancreas.
6. *Rheumatoid arthritis.* Antibodies damage the joints.
7. *Rheumatic fever.* Antibodies cross-react with valves of the heart.

HYPERSENSITIVITY

Hypersensitivity is an abnormal response which produces physiological or histopathological damage in the host. There are four types of hypersensitivity reactions:

- Type I (Anaphylaxis or IgE mediated),
- Type II (Antibody-mediated cytotoxicity),
- Type III (Immune complex-mediated disorders) and
- Type IV (Delayed type or T cell-mediated hypersensitivity).

The characteristic features of each hypersensitivity reaction are given in Table 3.4-4.

IMMUNE MODULATION

Immune modulation refers to modification of the immunological response. It can be either enhanced or suppressed (see page 147).

Immune enhancement

Immune enhancement means there is increase in the response in terms of rate, intensity, duration and even induction of response to substances which were earlier non-immunogenic.

Immunological response can be potentiated by the use of certain substances referred to as adjuvants.

IMMUNODEFICIENCY DISEASES

Immunodeficiency diseases occur when the body defence mechanisms are impaired.

Immunodeficiency diseases may be classified as primary or secondary.

Table 3.4-4 Characteristics of hypersensitivity reactions

Characteristics	Type I	Type II	Type III	Type IV
1. Time of onset of reaction	1/2–8 h	5–12 h (peak 48–72 h)	3–8 h	24–48 h
2. Reaction mediators	IgE, histamine, serotonin SRS-A, etc.	IgG, IgM and complement	IgG, IgM neutrophils eosinophils, lysosomal enzymes	T lymphocytes and macrophages lymphokines.
3. Response to intradermal injection of antigen (allergen)	Wheal and flare	—	Erythema and oedema	Erythema and induration
4. Passive transfer with	Serum	Serum	Serum	T cells
5. Examples	• Anaphylaxis, • Asthma • Hay fever • Allergic with food and insect bite	• Transfusion reactions (incompatibility reaction) • Haemolytic disease of newborn • Drug induced allergies	• Arthus reaction • Serum sickness	Tuberculin test • Contact dermatitis • Graft rejection

Primary immunodeficiency

Primary immunodeficiency occurs due to defect in the development of the immune system.

X-Linked agammaglobulinaemia is the first immune deficiency disorder recognized by Bruton in 1952, hence also known as Bruton's disease.

Secondary immunodeficiency disease

Acquired deficiencies of immunological response mechanisms can occur secondarily to number of diseases. Secondary immunodeficiency is more common than the primary immunodeficiency. Acquired immune deficiency syndrome (AIDS) is the most important.

Acquired immune deficiency syndrome

AIDS, i.e. acquired immune deficiency syndrome is characterized by reduction in the number of T_H cells because of infection by human immunodeficiency virus (HIV) (Fig. 3.4-11). AIDS was first of all detected in USA in 1981.

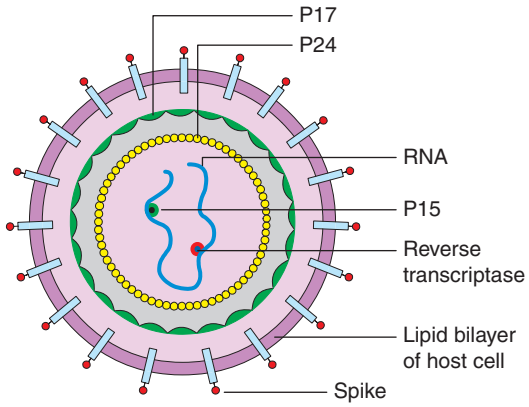


Fig. 3.4-11 Schematic diagram of structure of HIV.

Spread of disease. AIDS is a major worldwide life-threatening disease spreading rapidly. Daily about 8500 persons get infected with HIV. The high-risk groups include: sex workers, drug addicts, homosexual males, persons with extramarital relations and recipients of unscreened blood transfusion.

Chapter 3.5

Platelets, Haemostasis and Blood Coagulation

PLATELETS
- Structure and composition
- Properties and functions
- Normal count and variations
- Formation of platelets

HAEMOSTASIS
- Vasoconstriction
- Formation of temporary haemostatic plug
- Formation of definitive haemostatic plug

BLOOD COAGULATION
- **Clotting factors**
- **Mechanism of coagulation**
  - Blood clot retraction
    - Role of calcium in blood coagulation
    - Role of vitamin K, liver and vascular wall in haemostasis and coagulation
  - Why circulating blood does not clot?
  - Thrombosis

ANTIHAEMOSTATIC MECHANISMS
- Factors preventing platelet aggregation
- Circulatory anticoagulants
- Fibrinolytic mechanism
- Anticoagulants

BLEEDING DISORDERS
- Purpura
- Haemophilia
- Disseminated intravascular coagulation
- Laboratory tests in bleeding disorders

PLATELETS

STRUCTURE AND COMPOSITION

Platelets (small plates), also known as thrombocytes, (thrombo = clot; cytes = cells) have following features:

- *Size.* Platelets are the smallest blood cells varying in diameter from 2 to 4 µm with an average volume of 5.8 μm^3.
- *Shape and colour.* Platelets are colourless, spherical or oval discoid structures.
- *Leishman staining* shows a platelet to consist of faint bluish cytoplasm containing reddish purple granules.
- *Nucleus* is absent in the platelets and therefore these cannot reproduce.

Electron microscopic structure

Under electron microscope, a platelet shows following structural and compositional characteristics (Fig. 3.5-1):

1. Cell membrane. Each platelet is enclosed in a 6 nm thick trilaminar membrane identical with the plasma membrane of tissue cells. It consists of lipids (phospholipids, cholesterol and glycolipids), carbohydrates, proteins and glycoproteins. Its salient features are:

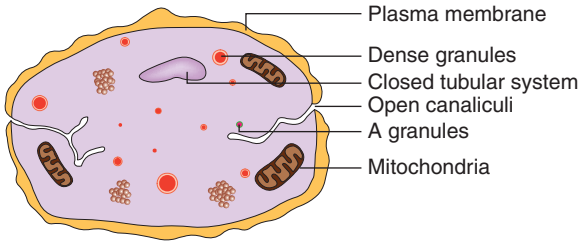


Fig. 3.5-1 Ultra structure of a platelet.

- *Glycoproteins* forming the surface coat of the platelet membrane prevent adherence of platelets to normal endothelium but accelerate the adherence of platelets to collagen and damaged endothelium in injured blood vessels.
- *Phospholipids* of the platelet membrane contain platelet factor-3, which plays an activating role at several points in the blood clotting process.
- *Invagination* of the surface membrane forms the so-called *canalicular system* or the *surface connecting system.*
- *Receptors* present on the platelet membrane are meant for combining with specific substances like collagen and fibrinogen.

- *Precursors* of various substances like thromboxane A_2, prostaglandins, leukotrienes and platelet factors 3 and 4 are also present in the platelet membrane.

2. Microtubules. Microtubules are made up of polymerized proteins called tubulins. These form a compact bundle which is present immediately beneath the platelet membrane and encircles the whole cytoplasm. These are responsible for maintenance of discoid shape of the circulating platelets.

3. Cytoplasm. Cytoplasm of the platelets contains:

- *Endoplasmic reticulum and Golgi apparatus.* These structures synthesize various enzymes and store large quantities of calcium.
- *Mitochondria.* These are capable of forming ATP and ADP.
- *Contractile proteins* include actin, myosin and thrombosthenin. Contractile proteins can cause the platelet to contract and are thus responsible for the clot retraction.
- *Other proteins* present in the cytoplasm are:
  - Fibrin stabilizing factor
  - Platelet-derived growth factor
  - Von Willebrand factor
- *Granules* present in the cytoplasm of platelets, clotting factors and platelet-derived growth factor (PDGF).
- *Enzymes* present in the cytoplasm of platelets include adenosine triphosphatase and the enzyme necessary for the synthesis of prostaglandins.

PROPERTIES AND FUNCTIONS

Properties of platelets (Fig. 3.5-2)

1. Adhesiveness. Platelets possess the property of adhesiveness, i.e. when they come in contact with any wet surface or rough surface, these are activated and stick to the surface. Factors responsible for adhesiveness are collagen, thrombin, ADP, thromboxane A_2, calcium ions and von Willebrand factor.

2. Aggregation. Platelets have the property to aggregate, i.e. they stick to each other. This is due to ADP and thromboxane A_2.

3. Agglutination. Clumping together of platelets is called agglutination. This occurs due to the actions of some platelet agglutinins.

Functions of platelets

When activated, platelets perform the following functions:

1. Role in haemostasis. Haemostasis refers to the spontaneous arrest of bleeding from an injured blood vessel (see page 151).

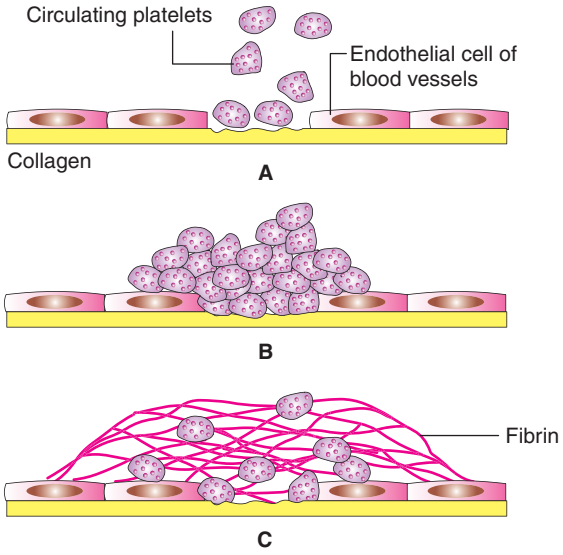


Fig. 3.5-2 Properties of platelets: A, adhesiveness; B, aggregation and C, formation of haemostatic plug.

2. Role in clot formation. Platelets play an important role in the formation of the intrinsic prothrombin activator which is responsible for the onset of blood clotting.

3. Role in clot retraction. Contraction of contractile proteins (actin, myosin and thrombosthenin) presents in the platelets play an important role in clot retraction.

4. Role in repair of injured blood vessels. The PDGF present in the cytoplasm of platelets plays an important role in the repair of endothelium and other structures of the injured/damaged blood vessels.

5. Role in defence mechanism. Platelets, due to their property of agglutination, are capable of phagocytosis. These are particularly helpful in phagocytosis of carbon particles, viruses and immune complexes.

6. Transport and storage function. The 5HT is stored in the platelets and transported to the site of injury where it is released.

NORMAL COUNT AND VARIATIONS

Normal count

Normal platelet count ranges from 150,000 to 450,000/μL with an average count of 2.5 lac/μL.

Pathological variations

A. Thrombocytosis. An increase in the number of platelets more than 4.5 lac/μL is called thrombocytosis.

Causes of thrombocytosis. Platelet count is increased:

1. After splenectomy
2. After:
   - Haemorrhage,

- Severe injury,
- Major surgical operation and
- Parturition.

3. In myeloproliferative disorders such as:
   - Chronic myeloid leukaemia,
   - Polycythaemia vera and
   - Myelofibrosis.

B. Thrombocytopenia. Decrease in the number of platelets below 1.5 lac/μL is called thrombocytopenia.

Causes of thrombocytopenia are:

1. Idiopathic thrombocytopenic purpura
2. Bone marrow depression due to:
   - Effects of various cytotoxic drugs,
   - Whole body irradiation,
   - Hypoplastic and aplastic anaemia.
3. Acute leukaemia or secondary deposits of malignancy in the bone marrow.
4. In infections like smallpox, chickenpox, typhoid and dengue fever.
5. In hypersplenism.
6. In toxaemia, septicaemia and uraemia.

FORMATION OF PLATELETS

Formation or development of platelets is called thrombopoiesis. The platelets are produced in the bone marrow. The pluripotent stem cell destined to form platelets is converted into colony forming units called Meg-CFU, which develop into platelets after passing through various stages.

Stages in platelet production (Fig. 3.5-3)

1. Megakaryoblast. The earliest recognizable precursor of platelets in the bone marrow is megakaryoblast. It arises from the Meg-CFU by a process of differentiation.

- Diameter of megakaryoblast is about 20–30 mm,
- *Cytoplasm* is small, blue and non-granular, and
- *Nucleus* is large, oval or kidney shaped with several nucleoli.

2. Promegakaryocyte. Promegakaryocyte is formed from the megakaryoblast. A megakaryoblast undergoes endoreduplication of nuclear chromatin, i.e. nuclear chromatin replicates repeatedly in multiples of two without division of the cell. Ultimately, a large cell containing up to 32 times the normal diploid content of nuclear DNA is formed when further nuclear replication ceases and cytoplasm becomes granular. The granules are intensely basophilic.

3. Megakaryocyte. A promegakaryocyte matures into a megakaryocyte with the following features:

- *Diameter.* Mature megakaryocyte is large cell, 30–90 μm in diameter.

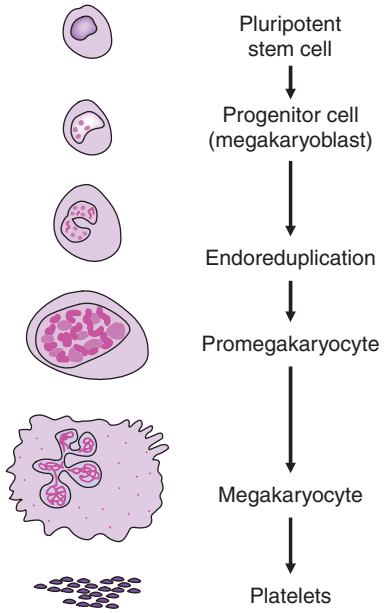


Fig. 3.5-3 Stages of thrombopoiesis.

- *Nucleus.* Megakaryocyte has single multilobed (4–16 lobes) nucleus with coarsely clumped chromatin.
- *Cytoplasm* is abundant, light blue in colour and contains red-purple granules.
- *Cell margin* is irregular and shows many pseudopodia. Platelets are formed from pseudopodia of megakaryocyte cytoplasm which get detached into the blood stream. Each megakaryocyte may form up to 4000 platelets. The formation of platelets from the stem cell takes about 10 days.

Control of thrombopoiesis

Thrombopoiesis seems to be regulated by thrombopoietin, megakaryocyte—colony stimulating activity (Meg-CSA).

Life span and fate of platelets

Life span of platelets varies from 8 to 12 days with an average of 10 days. Platelets are destroyed by tissue macrophage system in spleen. Therefore:

- *Splenomegaly* causes reduction in the platelet count
- *Splenectomy* is followed by an increase in the platelet count.

HAEMOSTASIS

Haemostasis refers to the spontaneous arrest or prevention of bleeding from the injured/damaged vessels by the physiological process. It involves three main steps (Fig. 3.5-4):

- Vasoconstriction,
- Formation of temporary haemostatic plug and
- Formation of the definitive haemostatic clot.

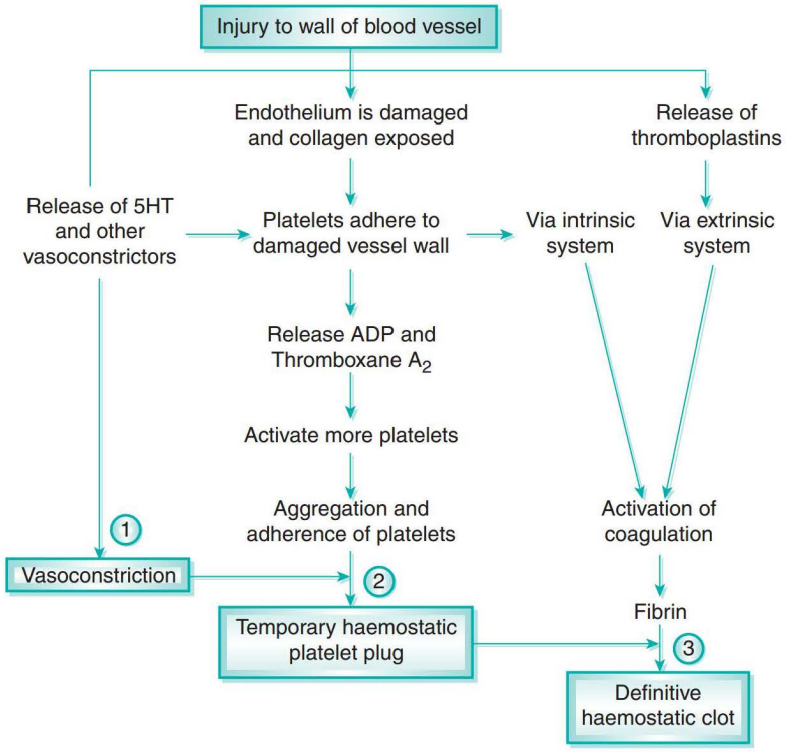


Fig. 3.5-4 Steps of haemostasis.

1. Vasoconstriction

Initial vasoconstriction is caused by direct effect of injury on the vascular smooth muscles. The initial vasoconstriction is transient but is maintained for several minutes or even hours by humoral facilitation due to release 5HT and other vasoconstrictors.

2. Formation of temporary haemostatic plug

Formation of a temporary haemostatic plug by the platelets at the site of injury involves following steps:

Platelet adhesion. Following injury, platelets come in contact with the damaged collagen fibres and endothelial cells of the vessel wall and change their characteristics. That is, they begin to swell and assume irregular forms with large number of pseudopodia protruding from the surface. The contractile proteins of the platelets contract forcibly and cause release of granules that contain multiple factors. They become sticky and therefore adhere to the collagen of damaged cell wall and to the damaged endothelium.

Platelets activation. The platelets secrete large quantities of ADP and thromboxane A_2, which act on the nearby platelets and cause their activation. Stickiness of these additional platelets causes them to adhere to originally activated platelets. In this way, a vicious cycle is initiated which leads to activation and adherence of large number of platelets.

Platelets aggregation. The large numbers of activated sticky platelets stick to each other forming platelets aggregation. Platelets aggregation is also increased by *platelet activating factor*, a cytokine secreted by neutrophils, monocytes and platelet cell membrane lipids.

Platelet aggregation initiates a series of reactions which result in formation of thromboxane A_2 and prostacyclin from the platelet membrane phospholipids.

Note. Aspirin prevents platelet aggregation by inhibiting formation of thromboxane A_2. Therefore, aspirin in low doses is of value in preventing myocardial infarction.

Formation of temporary haemostatic plug. The platelets adherence and aggregation ultimately lead to the formation of platelet plug. At first, it is a fairly loose plug but is successful in blocking the blood loss if the vascular opening is small.

Inhibition of further plug formation. Prostacyclin formed from the membrane phospholipids inhibits thromboxane formation and thus curtail the process of further plug formation. This reaction keeps the platelets plug localized, i.e. prevents intravascular spread of plug.

3. Formation of definitive haemostatic plug

The temporary platelet plug is converted into the definitive haemostatic plug by the process of clot formation (blood coagulation) which involves a complex series of events (see page 153). Platelets play an important role in the formation

of the intrinsic prothrombin activator which is responsible for initiating the process of clot formation.

Note. The blood clot formed at the site of injury results in a tight unyielding seal or the so-called definitive haemostatic plug.

BLOOD COAGULATION

Blood remains in fluid condition within the blood vessels throughout life. But, when the blood is shed from the blood vessels or collected in a container, it looses its fluidity within a few minutes and gets converted into a jelly-like mass which is called clot. This phenomenon is called coagulation or clotting of blood.

The process of blood coagulation consists of a complex cascade of reactions. Before discussing the mechanism of blood coagulation in detail, it will be worthwhile to study the essential features of the various clotting factors involved in this process.

CLOTTING FACTORS

The process of coagulation essentially involves a stepwise activation of certain substances mostly proteins present in the blood and/or tissue fluids. These substances are called clotting factors and have been given Roman numerals:

- Factor I (Fibrinogen),
- Factor II (Prothrombin),
- Factor III (Thromboplastin),
- Factor IV (Calcium),
- Factor V (Labile factor or proaccelerin or accelerator globulin),
- Factor VI (non-existent),
- Factor VII (Stable factor or proconvertin),
- Factor VIII (Antihaemophilic factor A (AHF) or antihaemophilic globulin (AHG),
- Factor IX (Christmas factor or plasma thromboplastic component (PTC or antihaemophilic factor B),
- Factor X (Stuart-Prower factor),
- Factor XI (Plasma thromboplastin antecedent, i.e. PTA or antihaemophilic factor C,
- Factor XII (Hageman factor or glass factor or contact factor) and Factor XIII (Fibrin stabilizing factor or fibrinase or Laki-Lorand factor),
- HMW – K (High molecular weight kininogen or Fitzgerald factor),
- Pre-Ka (Prekallikrein or Fletcher factor),
- Ka – Kallikrein and
- PL – Platelet phospholipid.

Their characteristic features and role of clotting factors are summarized in Table 3.5-1.

MECHANISM OF COAGULATION

Normally, blood circulates in the blood vessels and does not clot spontaneously. Clot formation is initiated under the following situations:

- Trauma to the vascular wall and adjacent tissues,
- Trauma to blood and
- Contact of blood with damaged endothelial cells or collagen or other tissue elements outside the vessel.

The process of coagulation involves a *cascade* of reactions in which activation of one factor leads to activation of next clotting factor (Fig. 3.5-5). This enzyme cascade reaction is also called *water fall sequence* by R.G. Macfarlane in 1967. The process of coagulation can be divided into three main steps:

A. Formation of prothrombin activator,
B. Conversion of prothrombin to thrombin and
C. Conversion of fibrinogen into fibrin.

A. FORMATION OF PROTHROMBIN ACTIVATOR

Two different mechanisms involved in the formation of prothrombin activator are:

1. Extrinsic pathway and
2. Intrinsic pathway.

1. Extrinsic pathway

The extrinsic pathway of formation of prothrombin activator begins with trauma to the vascular wall or the tissues outside the blood vessel. It includes following three basic steps (Fig. 3.5-5):

Release of tissue thromboplastins. The traumatized tissues release several substances which are together known as *tissue thromboplastin (factor III).*

Activation of factor X to form activated factor X. Tissue thromboplastin combines with factor VII (stable factor) to form the tissue thromboplastin–factor VII complex which in the presence of Ca^{2+} activates factor X to form activated factor X (Xa).

Effect of activated factor X to form prothrombin activator. The activated factor X along with the tissue phospholipids or phospholipids released from platelets, factor V (Labile factor) and Ca^{2+} forms a complex which is called prothrombin activator.

2. Intrinsic pathway

The intrinsic pathway of formation of prothrombin activator begins in the blood itself following trauma to blood itself or exposure of blood to collagen in a traumatized

Table 3.5-1 Characteristics of clotting factors

International nomenclature	Name	Description
Factor I	Fibrinogen	It is a soluble plasma protein globulin in nature. Its molecular weight is 500,000 Da. It is synthesized in liver. It has six polypeptide chains. Its plasma concentration is about 0.3 g/dL. It is converted into fibrin in the presence of enzyme thrombin.
Factor II	Prothrombin	Prothrombin is a plasma protein (an α_2 globulin) with the following features: • It is the inactive precursor of the enzyme thrombin (which is not present normally in the circulating blood). • Its molecular weight is about 69,000 Da. • It is synthesized in liver in the presence of vitamin K. • Its concentration in plasma of an adult is 40 mg/dL which falls in liver diseases: In newborn baby plasma concentration of prothrombin is lower.
Factor III	Thromboplastin	It is also called tissue factor or tissue thromboplastins. It is released in the extrinsic pathway of formation of prothrombin activator.
Factor IV	Calcium	Ionic calcium is essential for blood coagulation. Its role in coagulation is described on page 156.
Factor V	Labile factor, proaccelerin	It is also called proaccelerin. It is a protein and as the name indicates it is labile or unstable factor of the plasma. It is required for the formation of prothrombin activator and thus conversion of prothrombin to thrombin in both, extrinsic as well as intrinsic mechanisms of blood coagulation. Factor V is consumed during clotting and is therefore absent from serum.
Factor VII	Stable factor or autoprothrombin I or Proconvertin, SPCA	It is a stable protein synthesized in the liver in the presence of vitamin K. It is required for the activation of factor X in the extrinsic pathway. It is not consumed during clotting and therefore is present in serum as well as plasma.
Factor VIII	Antihaemophilic globulin (AHG), Antihaemophilic factor-A	It is a protein of β_2 globulin type synthesized in the liver. It is required for the activation of factor X and thus formation of prothrombin activator in intrinsic pathway. It is consumed during clotting and is therefore absent from the serum. Its congenital deficiency causes classical haemophilia (haemophilia A), which is an inherited disease in which the clotting time is prolonged.
Factor IX	Christmas factor, plasma thromboplastin component (PTC), Antihaemophilic factor-B	It is also called plasma thromboplastic component (PTC) or autoprothrombin II. It is a protein synthesized in liver independent of vitamin K. It is activated by the active factor XI in the presence of Ca^{2+}. It is essential for the formation of prothrombin activator in the intrinsic pathway. Its absence or deficiency causes haemophilia B, which is an inherited disease and is similar to haemophilia A.
Factor X	Stuart-Prower factor	It is a protein present in plasma and is synthesized in the liver. It is activated by an active factor IX in the presence of factor VIII, Ca^{2+} and phospholipids. Activated factor X along with an active factor V, Ca^{2+} and phospholipids forms a complex which is called prothrombin activator, both in extrinsic as well as intrinsic pathways.
Factor XI	Plasma thromboplastin antecedent	It is activated by an active factor XII. It is required for the activation of factor IX in the presence of Ca^{2+} in intrinsic pathway.
Factor XII	Hageman factor, glass factor	Factor XII is activated to XIIa when it comes in contact with a negatively charged surface, foreign substances or rough surface. Its activation in the blood initiates intrinsic pathway by activating factor XI (PTA) to XIa.
Factor XIII	Fibrin stabilizing factor, Laki-Lorand factor	This is a plasma protein which is required for stabilization of fibrin polymers in the presence of Ca^{2+}.
HMW-K	High molecular weight kininogen	It is responsible for attracting prekallikrein and factor XI to the site of reaction with factor XII. This is possible because HMW-K, like factor XII, is attracted towards the negatively charged surfaces which provide the site of reactions.

Pre-Ka, Ka	Prekallikrein, Fletcher factor and kallikrein	Prekallikrein is activated to kallikrein by XIIa, which in turn activates XII to XIIa. This phenomenon is called feedback activation of XII and is shown as: XII → XIIa; Kallikrein → Prekallikrein
PL	Platelet phospholipids	Platelets contain phospholipids (PPL) which are essential for clotting in the absence of tissue extract, i.e. in intrinsic pathway of coagulation. Arabic numerals are sometimes used for platelet activities affecting blood coagulation. For example the term: • Platelet factor 3 (PF-3) is used for the platelet phospholipid procoagulant activity. • Platelet factor 4 (PF-4) is used for heparin neutralizing activity of platelets.

Note. Factor VI is not a separate entity and has been dropped.

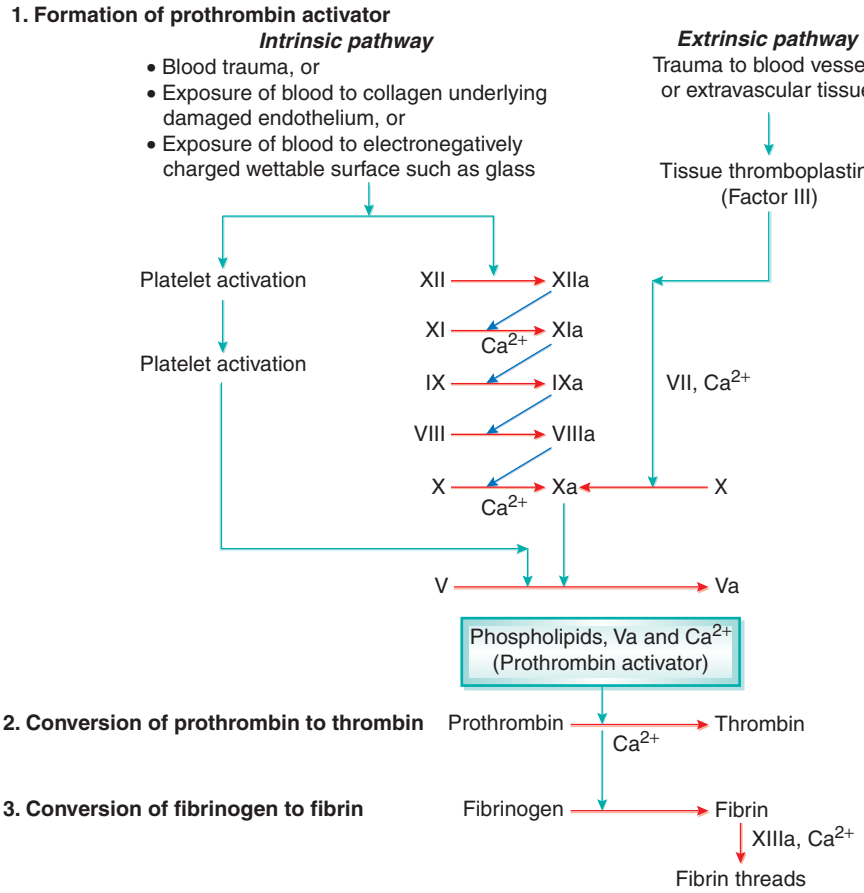


Fig. 3.5-5 Mechanism of blood coagulation.

vascular wall. The steps of intrinsic pathway are summarized in Fig. 3.5-5:

Activation of factor XII. Trauma to blood or exposure to collagen fibres underlying damaged vascular endothelium (or electronegatively charged wettable surface such as glass, in vitro) activates plasma factor XII to form XIIa and initiates the intrinsic pathway. Platelets are also activated.

Activation of factor XI to form XIa is caused by the activated factor XII.

Activation of factor IX to form IXa is in turn caused by the activated factor XI in the presence of Ca^{2+}.

Activation of factor X. Factor IXa in the presence of the activated factor VIII, Ca^{2+} and phospholipids (released by activated platelets) activates factor X to form Xa.

Formation of prothrombin activator. The activated factor X along with the phospholipids released by the activated platelets, activated factor V and Ca^{2+} forms a complex which is called prothrombin activator.

Note. Factor Va acts as a co-factor. Phospholipids (released from platelets) provide a surface where clot formation starts.

B. CONVERSION OF PROTHROMBIN TO THROMBIN

Conversion of prothrombin to thrombin is caused by the prothrombin activator in the presence of Ca^{2+}. This occurs at the surface of platelets which form the platelet plug at the site of injury.

Thrombin so formed acts as a proteolytic enzyme. It has been estimated that the amount of thrombin produced during clotting of only 1 mL of blood is sufficient to coagulate 3 L of blood.

Roles played by thrombin are:

- *Conversion of fibrinogen to fibrin* (discussed below).
- *Positive feedback role of thrombin.* It accelerates the rate of formation of prothrombin activator by the activating factors VIII, V and XIII. In this way, thrombin itself can cause further conversion of prothrombin into thrombin (amplification effect).
- It also activates protein-C (which is an anticoagulant).

C. CONVERSION OF FIBRINOGEN TO FIBRIN

Conversion of fibrinogen into fibrin involves three reactions (Fig. 3.5-6):

1. Proteolysis. Thrombin acting as a proteolytic enzyme removes four low molecular weight peptide chains from each molecule of fibrinogen to convert it into fibrin monomer.

2. Polymerization. Fibrin monomer polymerizes with another monomer to form *long fibrin* threads, which form reticulum of the clot. Initially, the clot is weak because the fibrin threads are not cross-linked with each other.

3. Stabilization of fibrin polymers. Fibrin stabilizing factor (factor XIII) which is activated by the thrombin to form XIIIa but XIIIa in the presence of Ca^{2+} causes formation of covalent cross-linkages between fibrin threads, thus adding tremendous strength to the fibrin meshwork. The fibrin meshwork traps the remaining components of plasma and blood cells to form a solid mass called clot.

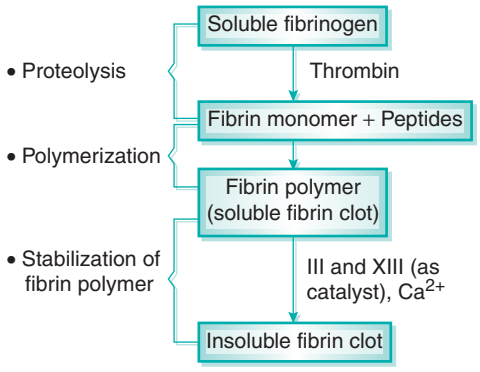


Fig. 3.5-6 Types of reactions involved in conversion of soluble fibrinogen into insoluble fibrin clot.

BLOOD CLOT RETRACTION

The blood clot formed at the end of coagulation process is composed of a meshwork of fibrin threads running in all directions along with the entrapped blood cells, platelets and plasma. The fibrin threads adhere to the damaged surface of blood vessels.

At this juncture, it is important to note that *coagulation is the property of plasma alone.* The RBCs and WBCs do not take part in it. They only become caught up in the meshwork of the clot.

Within a few minutes after a clot is formed, it begins to contract and usually squeeze out most of the fluid called *serum* (plasma without fibrinogen and other clotting factors) within 30–60 min.

Platelets are essential for the clot retraction. The contractile proteins (platelet thrombosthenin, actin and myosin) present in the cytoplasm of platelets cause strong contraction of platelet spicules attached to fibrin fibres.

- If a blood clot is kept for several hours, the clot retracts to about 40% of its original volume.
- Clot retraction is impaired if blood platelets have been removed.

ROLE OF CALCIUM IN BLOOD COAGULATION

From the study of mechanism of blood coagulation, it is quite clear that except for the first two steps in the intrinsic pathway, calcium ions are required for the promotion of all the reactions. Therefore in the absence of calcium ions, blood clotting will not occur. Thus, coagulation of blood can be prevented in vitro (e.g. for storage in the blood bank or for separation of plasma) by removal of calcium ions. The use of oxalates and citrates as in vitro anticoagulants is based on this principle. However, in vivo the degree of hypocalcaemia (e.g. due to deficiency of vitamin D or hypoparathyroidism) is compatible with life and does not cause bleeding disorder.

ROLE OF VITAMIN K, LIVER AND VASCULAR WALL IN HAEMOSTASIS AND COAGULATION

Role of vitamin K

Vitamin K is a complex naphthoquinone derivative. Vitamin K is obtained from the food as well as synthesized by bacterial flora in the gut.

In the liver, synthesis of following factors is dependent upon vitamin K:

- Coagulant like prothrombin,
- Factors VII, IX and X, and
- Circulatory anticoagulant protein.

Vitamin K deficiency. In the deficiency of vitamin K, prothrombin time and blood clotting time is prolonged and serious haemorrhages may occur.

Role of liver

Liver plays following significant role in the coagulation mechanism:

1. *Synthesis of procoagulants.* It is the site of synthesis of factors V, VII, IX, X, prothrombin and fibrinogen.
2. *Removal of activated procoagulants.* Liver also removes the activated procoagulants from the blood.
3. *Synthesis of anticoagulants.* Liver also synthesizes anticoagulants like heparin, antithrombin III and protein C.

Liver failure can cause both:

- *Bleeding disorders* due to hypocoagulability of the blood and
- *Uncontrolled extensive clotting* inside the blood vessels where clotting is not only unwanted but dangerous as well.

Role of blood vessels

Role played by endothelium, subendothelial tissue and smooth muscles of the media of the blood vessels in coagulation and haemostasis mechanisms is summarized.

Endothelium

Endothelium plays both anticoagulatory as well as coagulatory roles.

Anticoagulatory roles played by endothelium are:

- Smoothness of uninjured endothelial cells prevent platelet aggregation.
- Endothelial cells produce PGI_2 (a prostaglandin), which opposes platelet aggregation.

Role in clotting mechanism

- *Endothelium secretes von Willebrand's factor (VWF).* The plasma VWF initiates platelet aggregation and haemostasis.
- *Tissue factor (TF)* is released by the endothelial cells following trauma initiates the process of extrinsic pathway of clotting mechanism.
- *Plasminogen activator* which activates plasminogen to plasmin is also released by the endothelial cells.

Subendothelial tissue

Subendothelial tissue which chiefly consists of collagen fibres plays following roles in coagulation:

- *Platelet aggregation* is initiated when blood comes in contact with the subendothelial collagenous tissue.
- *Intrinsic coagulation pathway* is initiated when factor XII is activated following contact of blood with subendothelial collagenous tissue.

Vascular smooth muscle

Smooth muscles of vascular wall play role in haemostasis by causing vasoconstriction initiated by a mechanical injury to muscles.

WHY CIRCULATING BLOOD DOES NOT CLOT?

We all know that blood circulating in the blood vessels does not clot and that fluidity of the blood is essential for life. By now we have discussed most of the factors responsible for fluidity of the blood. They are summarized below.

1. Velocity of circulation. Blood is pumped into the vessels and circulated at a constant velocity which contributes to its fluidity. That is why, decrease in circulation velocity in certain conditions is associated with the intravascular clotting.

2. Surface effects of endothelium

- Smoothness of the endothelial lining inhibits platelet adhesion and thus prevents initiation of intrinsic clotting mechanism.
- A layer of glycocalyx (mucopolysaccharide) adsorbed to the inner surface of endothelium being negatively charged repels clotting factors (anion proteins) and platelets and thereby prevents clotting.
- Intact endothelium acts as a barrier between the thrombogenic subendothelial collagenous tissue and the blood.

3. Circulatory anticoagulants or the so-called natural anticoagulants present in the blood which prevent clotting are:

- Heparin,
- Antithrombin III,
- α_2 macroglobulin and
- Protein C (for details see page 159).

4. Fibrinolytic mechanism. Protein C is a naturally occurring anticoagulant which inactivates factors V and VIII and also inactivates an inhibitor of tissue plasminogen activator increasing the formation of plasmin which acts as fibrinolytic system.

- Further, whenever there is trauma, along with activation of clotting mechanism the fibrinolytic system is also activated which prevents spread of intravascular clotting.

5. Removal of activated clotting factors. Liver plays a role in preventing the intravascular clotting by removing activated clotting factors in the event of onset of spontaneous clot formation.

THROMBOSIS

We have studied that physiologically under normal conditions, the circulating blood does not clot and that clotting of blood occurs only extravascularly when a vessel has been injured and bleeding has occurred. However, under certain pathological conditions the intravascular clotting may occur. The intravascular clotting is called *thrombosis* and the clot so formed is called thrombus.

Predisposing factors

Virchow described three primary events which predispose to the thrombus formation *(Virchow's triad).* These are:

1. Endothelial injury. We have studied how an intact endothelium prevents coagulation (page 157). Endothelial injury may occur in many conditions. A few important ones are ulcerated plaques in advanced atherosclerosis, haemodynamic stress in hypertension, arterial disease, diabetes mellitus and hypercholesterolaemia.

2. Alterations in flow of blood. Both in turbulence as well as stasis of blood, normal axial flow of blood is disturbed and platelets come in contact with the endothelium initiating thrombus formation. Stasis of blood is commonly associated with venous thrombosis especially in the leg veins after major operations on the abdomen (postoperative thrombosis) or otherwise bedridden patients in which muscular contraction in legs and trunk (responsible for normal venous blood flow) is decreased.

3. Hypercoagulability of blood which predisposes to thrombosis may occur due to:

- Increase in coagulation factors such as fibrinogen, prothrombin, factors VIa, VIIa and Xa.
- Increase in the platelet count and their adhesiveness and
- Decreased levels of coagulation inhibitors such as antithrombin III and fibrinogen degradation products.

Effects of thrombi

Intravascular thrombi may cause variable effects (may be even life-threatening) depending upon their size and site. Thrombi cause harmful effects by one of the following mechanisms:

1. ***Ischaemia and infarction.*** Thrombi may decrease or stop the blood supply to part of an organ and cause ischaemia, which may subsequently result in infarction. For example, thrombus formation in coronary arteries may cause myocardial ischaemia and infarction.
2. ***Thromboembolism.*** The thrombus or its part may get dislodged and be carried along in the blood stream as *embolus* to lodge in a distant vessel. Examples of emboli formation are:
   - *Pulmonary embolism* and
   - *Cerebral embolism.*

Prevention of thrombi

Formation and/or extension of a thrombus can be prevented by the administration of:

- *Drugs which decrease platelet adhesiveness,* such as aspirin, dextran or dipyridamole,
- *Anticoagulants,* such as low doses of heparin and dicoumarol and
- *Intermittent compression* or electrical stimulation of the calf muscles is necessary in addition to above drugs for preventing post-operative venous thrombosis.

ANTIHAEMOSTATIC MECHANISMS

The factors which balance the tendency of the blood to clot in vivo constitute the *antihaemostatic factors.* These can be grouped as:

- Factors preventing platelet aggregation,
- Factors preventing coagulation (circulatory anticoagulants) and
- Factors causing fibrinolysis (fibrinolytic mechanism).

A. FACTORS PREVENTING PLATELET AGGREGATION

Prostacyclin

Prostacyclin is an endogenous factor which prevents platelet aggregation by inhibiting the thromboxane A_2 formation (which promotes platelet aggregation).

Note. The drug *aspirin* also inhibits the formation of thromboxane and thus when used can prevent platelet plug formation. This makes aspirin a valuable drug for the prevention of thrombosis in patients prone to myocardial infarction and stroke.

B. CIRCULATORY ANTICOAGULANTS

The natural anticoagulants circulating in the blood constitute the anticoagulant mechanism of the body. These include:

- Heparin,
- Antithrombin III or heparin co-factor II and
- Protein C.

1. Heparin

Heparin is a powerful naturally acting anticoagulant since it was first isolated from the liver so it is named heparin (hepar=liver). However, it is also present in many other organs. It is a polysaccharide containing many sulphate groups. Its molecular weight is 15,000–18,000 Da.

Secretion. Heparin is secreted by the basophils and mast *cells* (present in various tissues such as liver, lungs and tissues rich is connective tissue).

Mechanism of action. Heparin is present on the luminal surface of vascular endothelium. It acts as an anticoagulant by the following mechanisms:

- Prevents activation of prothrombin to thrombin,
- Inhibits the action of thrombin on fibrinogen and
- Facilitates the action of antithrombin III, i.e. acts as its co-factor and thereby inhibits the active forms of clotting factors IX, X, XI and XII.

2. Antithrombin III

Antithrombin III is present in plasma as well as vascular endothelium. It inactivates a number of coagulation factors including thrombin.

3. Protein C

Protein C is a plasma protein synthesized in liver. It, along with *thrombomodulin* and protein S, constitutes an important negative feedback pathway that keeps the coagulatory process under control. The steps of protein C pathway are given in Fig. 3.5-7.

Thrombomodulin is a thrombin binding protein produced by all endothelial cells except those in the cerebral microcirculation. It converts thrombin into *protein* C activator.

Protein C is activated by the protein C activator. For the formation of activated protein C (APC), a co-factor protein-S is required.

Activated protein C (APC) inactivates factors VIIIa and Va and thus inhibits the clotting mechanism. It also increases fibrinolysis by promoting plasmin formation.

C. FIBRINOLYTIC MECHANISM

Fibrinolysis refers to the process that brings about the dissolution of fibrin. The important component of the fibrinolytic system is *plasmin* or fibrinolysin which is present in the blood in an inactive form called *plasminogen* or profibrinolysin.

Plasmin

Plasmin is a powerful protease formed from its precursor, the plasminogen. It lyses fibrin and fibrinogen into fragments known as fibrin degradation products (FDP) that inhibit thrombin. Thus, there is a negative feedback which controls plasmin generation. There are two plasminogen activator systems in the body: intrinsic and extrinsic.

Extrinsic plasminogen activator system

The extrinsic activator system (Fig. 3.5-8) operates through the following agents:

1. Tissue plasminogen activator (TPA) also called vascular plasminogen activator is released from the vascular endothelium. In violent deaths (e.g. of a soldier in the battlefield), the blood is a fluid and incoagulable due to fibrinolysis. This is due to large amount of adrenaline released into blood before death.

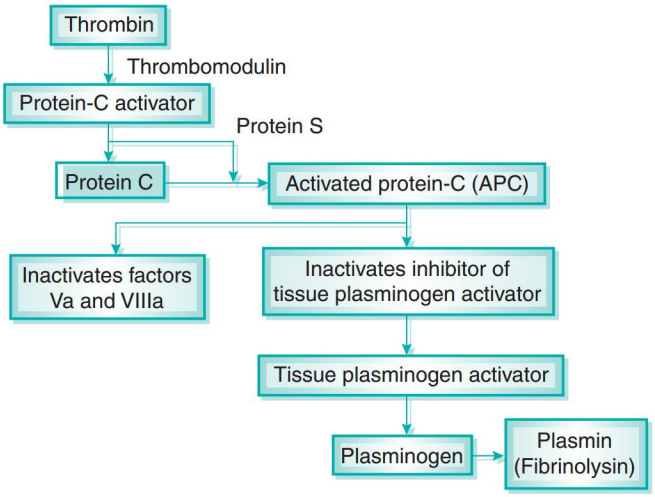


Fig. 3.5-7 Steps of protein C pathway in inhibiting coagulation and promoting fibrinolysis.

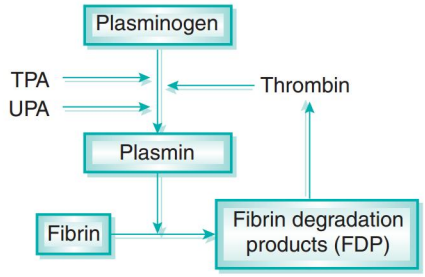


Fig. 3.5-8 Fibrinolytic mechanism operating through extrinsic plasminogen activator system.

Note. It is important to note that TPA is now produced by recombinant DNA techniques and is available for clinical use. It lyses clots in the coronary arteries if given to patients soon after the onset of myocardial infarction.

2. Urokinase-type plasminogen activator (UPA). The UPA is found in number of tissues including endothelial cells, renal cells and tumour cells.

Note. Streptokinase and staphylokinase are bacterial enzymes known to produce activation of plasmin like TPA and UPA. Therefore, they are being used in the treatment of early myocardial infarction.

Intrinsic plasminogen activator system

Contact factors (factor XIIa and kallikrein) that initiate clotting mechanism also stimulate the dissolution of clot by activating plasminogen and constitute the intrinsic plasminogen activator system (Fig. 3.5-9).

Factors affecting fibrinolytic system

The rate of fibrinolysis is influenced by the promotors (i.e. plasminogen activators) and inhibitors. Fibrinolysis

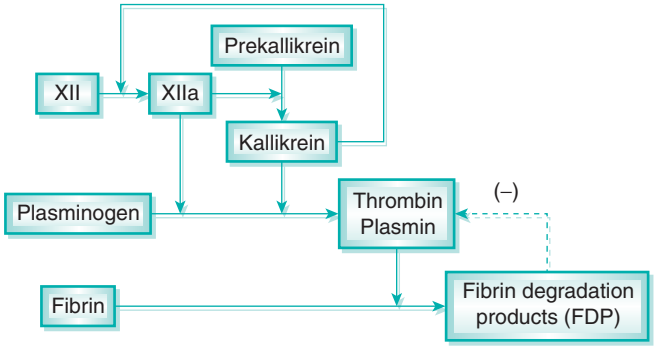


Fig. 3.5-9 Fibrinolytic mechanism operating through intrinsic plasminogen activator system.

inhibitors are present in plasma, blood cells, tissues and extracellular matrix. These can inhibit plasmin (antiplasmin) or prevent the activation of plasminogen. The various inhibitors are:

- *Antiplasmins* such as α_2-antiplasmin,
- *Drugs* like aprotinin (a trypsin inhibitor) and epsilon–amino caproic acid inhibit fibrinolysis.

Note. Fibrinolysis is promoted by stress (physical or mental).

Physiological role of fibrinolysis system

Plasmin of the fibrinolysis system plays the following physiological roles:

1. Cleaning the minute clots of tiny vessels. The fibrinolytic system is constantly in action to prevent excessive fibrin formation.

2. Promote normal healing process. Lysis of clot formed as a result of tissue injury helps to promote normal healing process.

3. Liquefaction of menstrual clot in the vagina is carried out by the fibrinolytic system.

4. Liquefaction of sperms in the epididymis when seminal ejaculation does not occur is caused by the fibrinolysin system.

5. Role in inflammatory response. In addition to its fibrinolytic activity, plasmin can form plasma kinins (bradykinins, kallidin) and thus contribute to the vascular and sensory features (pain) of the inflammatory response to injury.

ANTICOAGULANTS

Anticoagulants refer to the substances which delay or prevent the process of coagulation of blood.

Types. Anticoagulants may be divided into endogenous and exogenous anticoagulants.

A. Endogenous anticoagulants

Endogenous anticoagulants are those which are present inside the blood naturally:

- Heparin,
- Antithrombin III and
- Protein C.

For details, see page 158.

B. Exogenous anticoagulants

Exogenous anticoagulants are administered from outside or are used in vitro. These include:

- Heparin,
- Calcium sequesters,
- Vitamin K antagonist and
- Defibrination substances.

1. Heparin. Heparin, a naturally acting anticoagulant can also be synthesized. It inhibits blood coagulation both *in vivo* and *in vitro*. For details see page 158.

2. Calcium sequesters or decalcifying agents. In vitro, blood clotting can be prevented by substances which sequester (remove) calcium from the blood. These include two types of agents:

- Substances which form insoluble salts with calcium, such as sodium citrate and sodium oxalate and
- Calcium chelators which bind calcium, such as ethylene diamine tetraacetic acid.

3. Vitamin K antagonists. These are used orally and thus can prevent coagulation in vivo effectively. These include Coumarin derivatives, e.g. *dicoumarol* and *Warfarin*.

Mechanism of action. These agents occupy vitamin K receptor sites in the liver and prevent vitamin K to carry out its normal physiological function, hence the name vitamin K antagonists. Thus, these substances inhibit synthesis of vitamin K-dependent factors, i.e. factors VII, IX and X.

4. Defibrination substances. Defibrination substances are those which cause destruction of fibrinogen. These include:

- *Malaysian pit viper venom.* It is a type of snake venom which in vivo acts as an anticoagulant by causing defibrination and also by stimulating fibrinolytic system. In vitro, it has a direct anticoagulant effect on fibrinogen by forming imperfect fibrin polymer.
- *Arvin or ancord.* It is purified preparation of snake venom. It is a glycoprotein in nature and is administered by injection.

5. Cold. Keeping blood cold (at 5–10°C) can retard the process of coagulation but cannot absolutely prevent it.

Because of this reason blood is stored in blood banks at low temperature.

Note. It looks paradoxical that whenever bleeding occurs, ice is applied to arrest the haemorrhage (while cold delays coagulation); in fact, when ice is applied on the surface it prevents bleeding by inducing reflex vasoconstriction.

BLEEDING DISORDERS

Bleeding disorders are characterized by spontaneous escape of blood from blood vessels (in the tissues, inside the body cavities or on few surfaces like skin and mucous membrane or persistent and/or excessive bleeding following minor injuries like tooth extraction etc.

Classification of bleeding disorders

I. Platelet disorders

A. *Deficiency of blood platelets. Thrombocytopenic purpura,* *see page 161.*

B. *Functional disorder of platelets.*

II. Coagulation disorder or defective coagulation mechanism

1. Deficiency of clotting factors (see Table 3.5-2)
2. Vitamin K deficiency (see page 157)
3. Anticoagulant overdose
4. Disseminated intravascular clotting.

III. Vascular disorders. Damage of capillary endothelium (Non-thrombocytopenic purpura)

- Due to infection by bacteria and their toxins,
- Due to toxic effects of drugs and chemicals,
- Due to avitaminosis C,
- Allergic purpura and
- Connective tissue diseases.

Only a few important bleeding disorders, the purpura and haemophilia, are described briefly.

PURPURA

Purpura is a group of bleeding disorder occurring due to various causes. The term purpura is derived from purple-coloured petechial haemorrhages and bruises in the skin. The blood that leaks out changes colour from red to blue to dark blue and green over a period of time.

Thrombocytopenic purpura

Decrease in the platelet count below 1.5 lac/μL is called thrombocytopenia. Thrombocytopenic purpura may be:

- *Primary thrombocytopenic purpura* (idiopathic, cause not known) and
- *Secondary thrombocytopenic purpura.* The causes of platelet deficiency are:
  - Bone marrow depression due to effects of various cytotoxic drugs. Whole body irradiations and hypoplastic and aplastic anaemia.
  - Leukaemia and secondary deposits of malignancies in the bone marrow.
  - Acute septicaemia, toxaemia and uraemia.
  - Hypersplenism.

Relation between platelet count and bleeding is as follows:

- Above 100,000/μL : No clinical symptoms; bleeding is rare.
- From 50,000–100,000/μL : Bleeding may occur after major surgery.
- From 20,000–50,000/μL : Bleeding occurs with minor trauma in everyday life.
- Below 20,000/μL : Spontaneous haemorrhage in urinary tract, GI tract, nose bleeds, etc.
- At very low counts : Fatal haemorrhage may occur in the brain.

Table 3.5-2 Deficiency of clotting factors

Deficiency of factor	Clinical syndrome	Cause
• Factor I	Fibrinogenopenia Afibrinogenaemia	Depletion during pregnancy with premature separation of placenta. Congenital (rare)
• Factor II	Hypoprothrombinaemia (Haemorrhagic tendency in liver disease)	Decreased hepatic synthesis, usually secondary to vitamin K deficiency
• Factor V	Parahaemophilia	Congenital
• Factor VII	Hypoconvertinaemia	Congenital
• Factor VIII	Haemophilia A or classical haemophilia	Congenital defect due to abnormal gene on X chromosome
• Factor IX	Haemophilia B or Christmas disease	Congenital
• Factor X	Stuart-Prower factor deficiency	Congenital
• Factor XI	PTA deficiency	Congenital
• Factor XII	Hageman trait (does not produce bleeding)	Congenital
• Von Willebrand's factor	Von Willebrand's disease	Congenital

Non-thrombocytopenic purpura

Non-thrombocytopenic purpura occurs due to vessel wall defects, the platelet counts are normal, but bleeding time is prolonged and capillary fragility test is positive.

Causes of non-thrombocytopenic purpura are:

1. *Drug-induced damage to capillary wall* is seen in patients with prolonged treatment with corticosteroids, penicillin, sulpha drugs and aspirin.
2. *Deficiency of vitamin C (scurvy)* causes failure of collagen formation and associated with impaired hydroxyproline synthesis. Petechiae and bleeding from gums occur due to decreased intercellular substance and less stable capillary basement membrane.
3. *Allergic purpura* occurs due to damage to capillary walls by antibodies.
4. *Infections* such as typhus, bacterial endocarditis and haemolytic streptococcus may be associated with capillary wall damage.
5. *Senile purpura* refers to purpuric haemorrhagic spots seen on the back of hands and forearm due to prolonged pressure or mild trauma. Small vessels in old age rupture due to increased mobility of skin resulting from less of elastic and connective tissues around the blood vessels.
6. *Connective tissue diseases* are also sometimes associated with damage to capillary walls and purpuric haemorrhages.

HAEMOPHILIA

Haemophilia is the name given to a group of disorders occurring due to *hereditary deficiency of coagulation* and characterized by bleeding tendencies associated with increased clotting time. The haemophilia includes:

1. Haemophilia A

- Haemophilia A, also known as true or classical haemophilia, occurs due to *deficiency of factor VIII,* i.e. antihaemophilic globulin (AHG). It occurs in 83% cases.
- Being sex linked recessive disease it affects males exclusively and females act as carriers (Fig. 3.5-10).
- The majority of patients with haemophilia A has blood levels of factor VIII below 5% and so usually bleed severely on minor trauma.
- Clinical features (bleeding tendency) are not apparent since birth but generally start early in life (within first 3 years).
  - The haemophilics have a tendency to bleed into soft tissues, muscles, joints, GI tract, urinary tract and from nose.

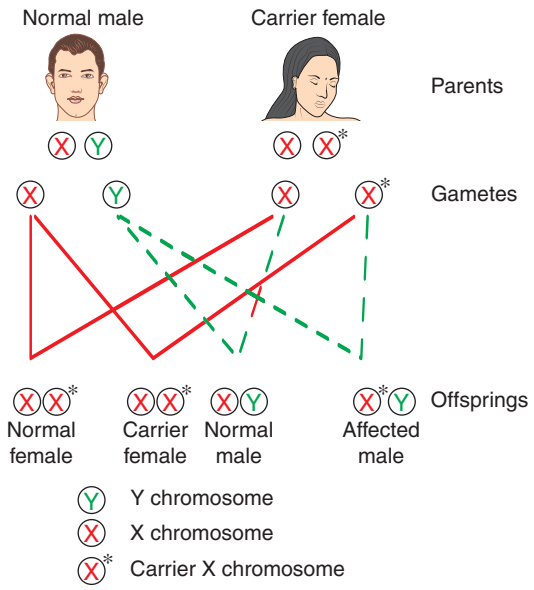


Fig. 3.5-10 Sex linked inheritance of haemophilia A.

  - Joints of haemophilic patients become severely damaged due to repeated joint haemorrhage.
  - Haemorrhage into the soft tissue around the floor of the mouth may cause respiratory obstruction and death by suffocation.
- Haemophilics have normal bleeding time, platelet count and prothrombin time (PT). Coagulation time is increased and typically, the patients have prolonged partial thromboplastin time (PTT).

2. Haemophilia B

- Haemophilia B, also known as *Christmas disease,* occurs due to deficiency of factor IX (Christmas factor or plasma thromboplastin component, PTC). It was discovered in a family with surname Christmas.
- Like haemophilia A, haemophilia B is also a recessive X-linked disease that occurs in males and is transmitted by females.

3. Haemophilia C

- Haemophilia C refers to deficiency of PTA (factor XI).
- It is inherited as *Mendelian dominant* and affects both males and females.
- Clotting time in this condition may be prolonged or may be within normal limits.

DISSEMINATED INTRAVASCULAR COAGULATION

- Disseminated intravascular coagulation (DIC), as the name indicates, refers to the condition when clotting

mechanism becomes activated in widespread areas of the circulation.
- Due to widespread intravascular coagulation, there occurs plugging of small vessels with clots resulting into decreased O_2 and nutrient supply to its tissues causing multiple organ damage.
- The widespread intravascular coagulation uses up most of the coagulation factors and platelets present in the blood resulting in the failure of haemostatic mechanism. The patient thus develops bleeding tendencies. Hence, the condition is also called *consumption coagulopathy.*

LABORATORY TESTS IN BLEEDING DISORDERS

Bleeding time

Definition. Bleeding occurs from the skin when it is pricked with a needle, which normally stops of its own within a few minutes. The time lapse between the skin prick and the arrest of bleeding is called *bleeding time* (BT).

Normal BT *by* Duke's method varies from 1 to 6 min. Normal BT indicates that platelets count and their function as well as health of capillaries are normal.

Prolonged BT occurs in purpura, while it is normal in haemophilia.

Capillary fragility test of Hess or Tourniquet test

Tourniquet test is performed to assess the mechanical fragility of the capillaries by raising pressure within them. It may demonstrate latent purpura.

Platelet count

Normal platelet count varies from 1.5 lac to 4.5 lac/μL with an average 2.5 lac/μL. Platelet count is decreased in primary and secondary thrombocytopenic purpura (page 161).

Coagulation time

Definition. Coagulation time refers to the time taken by the fresh fluid blood to get coagulated (demonstrated) by the formation of fibrin threads. It is abnormally prolonged when the coagulation factors are seriously deficient.

Importance of coagulation time (CT). The CT is prolonged in haemophilia and other clotting disorders because thrombin cannot be normally generated; however, the BT, which reflects vasoconstriction and platelet plug formation independently of clot formation, is normal since CT can increase due to deficiency of any of the factors so it is a non-specific test. More specific tests include *prothrombin time, partial thromboplastin time, thoromboplastin generation test, thrombin time,* etc. Specific tests can pinpoint the particular deficient factor.

- *Physiologically* clotting time is reduced during menstruation and before and during parturition.
- *Pathologically CT is prolonged in* haemophilia, liver diseases, afibrinogenaemia, Christmas disease, vitamin K deficiency and DIC.

Prothrombin time

Procedure. Quick's one stage method, now is the standard method to measure prothrombin time (PT). In this method, oxalated or citrated plasma of the patient are added to tissue thromboplastin (commercially available) and calcium chloride solution (to provide calcium ion) and the mixture is incubated at 37°C. The end point is conversion of fluid plasma into a gel (due to formation of fibrin). Normal PT is 11–16 s. Intrinsic system is not involved in this test since plasma does not contain platelets. Obviously it is extrinsic system which is tested.

- PT is increased in patients on oral anticoagulants, liver failure, vitamin K deficiency, abnormally low fibrinogen concentration, deficiency of factors II, V, VII and X.
- PT is normal in haemophilia and Christmas disease.

Partial thromboplastin time

Partial thromboplastin time (PTT) also known as kaolin cephalin clotting time (KCCT).

Procedure. To the oxalated or citrated plasma of the patient are added kaolin (to provide surface contact), cephalin (to provide phospholipids) and calcium chloride (to provide calcium ions), and the mixture is incubated at 37°C. The end point is formation of plasma gel. Normal PTT is about 40 s.

Importance of PTT. It is used to monitor the heparin therapy.

PTT is prolonged in haemophilia, von Willebrand's disease, liver failure, deficiency of contact factor XII, anticoagulant therapy and intravascular clotting.

Thromboplastin generation test

Thromboplastin generation test (TGT) measures generation of thromboplastin, i.e. the efficiency of a part of the intrinsic mechanism of coagulation. Normal value of TGT is 12 s or less. Prolonged TGT indicates deficiency of factors needed to form prothrombin activator by the intrinsic mechanism, i.e. factors VIII, IX, X and V.

Comments. From values of PT and TGT, following inferences can be drawn:

- In haemophilia, PT is normal but TGT is prolonged,
- In pure factor VII deficiency, PT is prolonged, but TGT will remain normal and
- In factor X deficiency, both PT and TGT will be abnormal.

Thrombin time (TT)

This test measures the final step in coagulation, i.e. functional fibrinogen available. In this test, thrombin is added to plasma, which will convert fibrinogen present in the plasma to fibrin. Normally a clot is formed in about 10 s, which is the end point. *TT is prolonged* in hypofibrinogenaemia, dysfibrinogenaemia, DIC and heparin treatment.

Clot retraction test

Clot retraction test measures time needed for contraction of an undisturbed clot. It indicates function and number of platelets. Normally clot retraction begins within 2 h and is completed within 24 h.

- *Clot retraction is retarded* in thrombocytopenia.
- Clot is small and soft in thromboasthenia, i.e. functional disturbance of platelets.

Chapter 3.6

Blood Groups and Blood Transfusion

BLOOD GROUPS

INTRODUCTION

Agglutinogens and agglutinins

Agglutinogens refer to the antigens present on the cell membranes of RBCs. A variety of antigens are present on the cell membrane, but only a few of them are of practical significance.

Agglutinins refer to the antibodies against the agglutinogens. These are present in the plasma.

Agglutination of RBCs can be caused by the antigens present on their cell membranes in the presence of suitable agglutinins (antibodies). That is why, these antigens are called agglutinogens.

Blood grouping systems

Depending upon the type of agglutinogen present or absent on the red cell membranes, various blood grouping systems are known, which can be classified as:

Major blood group systems are based on the presence of agglutinogens which are widely prevalent in the population and are known to cause worst transfusion reactions. These include:

- The classical ABO blood grouping system and
- Rhesus (Rh), (CDE) blood grouping system.

Minor blood group systems are based on the presence of agglutinogens which are found only in small proportion of the population and occasionally produce mild transfusion reactions. These include:

- M and N blood grouping and
- P blood group system

Note. From clinical point of view only major blood group systems, i.e. classical ABO and Rh (CDE) blood grouping systems are important and so will be discussed in detail.

Landsteiner law

Karl Landsteiner in 1900, framed a law in relation to agglutinogens and agglutinins, which states that:

If an agglutinogen is present on the red cell membrane of an individual, the corresponding agglutinin must be absent in the plasma and

If an agglutinogen is absent from the cell membrane of RBCs of an individual, the corresponding agglutinin must be present in the plasma. It is important to note that:

- The Landsteiner law is applicable to ABO blood group system only.
- The law is not applicable to other blood group systems because there are no naturally occurring agglutinins in these systems.

CLASSICAL ABO BLOOD GROUPING SYSTEM

A AND B AGGLUTINOGENS

The classical ABO blood grouping system is based on the presence of A and B agglutinogens on the cell membrane of RBCs.

- A and B agglutinogens are complex oligosaccharides differing in their terminal sugars.
- The A and B antigens present on the membranes of RBCs are also present in many other tissues like salivary glands, pancreas, kidney, liver, lungs and testis; and also in body fluids like saliva, semen and amniotic fluid.
- The antigens on RBCs membrane are glycolipids, while in the tissues and body fluids they are soluble glycoproteins.

ANTI-A AND ANTI-B AGGLUTININS

- Anti-A agglutinin and anti-B agglutinin refer to the antibody, i.e. which reacts with or acts on the antigen A and antigen B, respectively.
- There are two types of α agglutinins: the α_1 and α proper.
- The α and β agglutinins are globulins of IgM type and *cannot cross the placenta.*
- The α and β agglutinins act best at low temperature (5–20°C) and are therefore also called as cold antibodies.

TYPES OF ABO BLOOD GROUPS

Depending upon the presence or absence of A and B agglutinogens and α and β agglutinins, there are four types of blood groups:

Blood group A is characterized by:

- Presence of A agglutinogen and absence of B agglutinogen on the cell membrane of RBCs.
- Presence of anti-B agglutinin and absence of anti-A agglutinin from the plasma.

Blood group B is characterized by:

- Presence of B agglutinogen and absence of A agglutinogen on the cell membrane of RBCs and
- Presence of anti-A agglutinin and absence of anti-B agglutinin from the plasma.

Blood group AB is characterized by:

- Presence of both A and B agglutinogens on the cell membrane of RBCs and
- Absence of both anti-A and anti-B or agglutinins from the plasma.

Table 3.6-1 Population distribution of ABO blood groups in India vis-a-vis in Britain

Blood group	India (%)	Britain (%)
A	20	42
B	40	9
AB	8	3
O	32	46

Blood group O is characterized by:

- Absence of both A and B agglutinogens on the cell membrane of RBCs and
- Presence of both anti-A and anti-B agglutinins in the plasma.

POPULATION DISTRIBUTION OF ABO BLOOD GROUPS (TABLE 3.6-1)

INHERITANCE OF ABO BLOOD GROUPS

Agglutinogens A and B or the non-antigenic substances which determine the blood groups are genetically inherited as Mendelian dominant in the classical Mendelian pattern.

The ABO phenotypes and possible genotypes are as under:

Phenotype	Genotype	
Blood group A	AA	AO
Blood group B	BB	BO
Blood group AB	AB	
Blood group O	OO	

Inheritance of classical ABO blood grouping (A, B, AB and O) depend upon three genes, A, B and O (named after A, B, and O factors). The blood group of offspring depends on two genes which are inherited from each parent. The possible blood groups (genotype and phenotype) of the offspring are shown in Table 3.6-2.

Agglutinogens A and B first appear in the sixth week of fetal life. Their concentration at birth is one-fifth of adult level and it progressively rises during puberty and adolescence.

Anti-A (or α) and Anti-B (or β) agglutinins (specific blood group antibodies) are absent at birth, but they appear 10–15 days after birth and reach a maximum concentration by the age of 10 years. The probable mechanism of appearance of α and β agglutinins is described. Antigens very similar to A and B antigens are commonly present in the intestinal bacteria and foods. When the newborn is exposed to these antigens, these are absorbed into the blood and stimulate the formation of antibodies against the antigens

Table 3.6-2 The possible blood groups (genotype and phenotype) of offspring when the blood group of father is B and that of mother is A

	Father				Mother				
Phenotype	B				A				
Genotype	BB		BO		AA		AO		
Inherited gene	B	B	B	O	A	A	A	O	
Offsprings									Offsprings
I	AB	AB	AB	BO					Genotype
	AB	AB	AB	B					Phenotype
II	AB	AB	AB	BO					Genotype
	AB	AB	AB	B					Phenotype
III	AB	AB	AB	BO					Genotype
	AB	AB	AB	B					Phenotype
IV	AO	AO	AO	OO					Genotype
	A	A	A	O					Phenotype

Table 3.6-3 Determination of blood group of an individual

Blood group of RBCs suspension	Agglutination with	
	Antiserum A (containing α agglutinins)	Antiserum B (containing β agglutinins)
A	+	–
B	–	+
AB	+	+
O	–	–

Note. + sign indicates agglutination and – sign indicates no agglutination.

recognized as non-self (i.e. not present in the own body) by the immune system.

DETERMINATION OF ABO BLOOD GROUPS

The ABO blood group of an individual can be determined by mixing one drop of suspension of the red cells (in isotonic saline) with a drop each of antiserum A (containing α agglutinins) and antiserum B (containing β agglutinins) separately on a glass slide. The antiserum A will cause agglutination (clumping of RBCs having A antigens) and antiserum B will cause agglutination of RBCs having B antigens). The blood group of the individual will be shown by the presence of agglutination with one, both or none of the sera (Table 3.6-3 and Fig. 3.6-1).

Note. The antisera A and B are available commercially. For a quick identification, the anti-A serum is tinted blue and anti-B serum is tinted yellow.

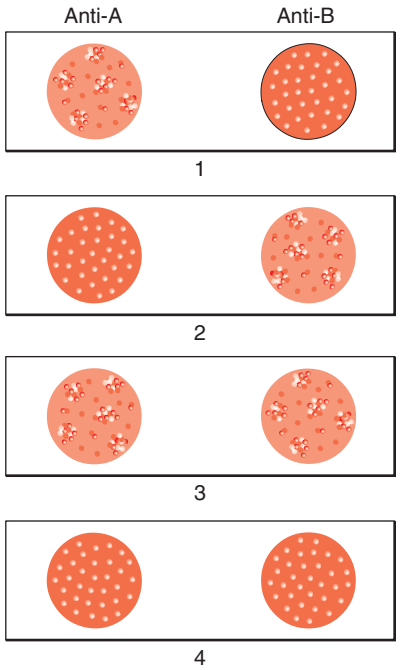


Fig. 3.6-1 Determination of blood groups—the RBCs showing agglutination with antisera are: 1, of blood group 'A' with antisera A; 2, of blood group B with antisera B; 3, of blood group 'AB' with antisera A and B (both) and 4, of blood group 'O' with none.

RHESUS (Rh) BLOOD GROUPING SYSTEM

Rh ANTIGENS

- The antigens responsible for this blood grouping system are called *Rh antigens* or Rh agglutinogens or Rh factor because these were first discovered in the RBCs of

rhesus monkeys. Based on the presence of Rh antigen, two types of blood groups are described:

- Rh positive blood group and
- Rh negative blood group.
- The Rh antigens were discovered by Landsteiner and Weiner in 1940. They noticed that when RBCs of rhesus monkey (monkey with red ischial callosity) were injected into a rabbit, antibodies were formed against these RBCs. When such rabbit's serum was tested against human red cells, agglutination occurred in 85% of the cases, i.e. these person's RBCs contained antigen which reacted with antibodies formed against rhesus monkey RBCs. They labelled this antigen as Rh antigens and such persons as Rh +ve. The remaining 15% were labelled as Rh –ve.
- Three types of Rh antigens, viz. C, D and E have been recognized. However, D antigen is commonest and produces worst transfusion reactions. Therefore, for all practical purposes, the term Rh antigen refers to D antigen.
- Rh antigens are integral membrane proteins.

Note. Rh agglutinogen has not been detected in other tissues and body fluids like A and B agglutinogens.

Rh ANTIBODIES

There are no natural antibodies of Rh antigens, while in ABO system of blood grouping α or β antibodies are always present naturally if the appropriate antigen is absent.

Rh antibodies (also called anti-D) are produced only when an Rh –ve individual is transfused with Rh +ve blood or when a Rh –ve mother gives birth to Rh +ve baby (Rh +ve RBCs of foetus enter into the maternal circulation), Rh antibodies are of IgG type and can cross the placenta. Since these react best at body temperature so are also called warm antibodies.

Once produced, the Rh antibodies persist in the blood for years and can produce serious reactions during the second transfusion.

INHERITANCE OF Rh ANTIGENS

- The Rh antigen (D antigen) is inherited as dominant gene D. When gene D is absent from a chromosome, its place is occupied by the alternate form (allelomorph) called 'd'. Rh gene is inherited from both the father and the mother.
- Rh +ve individual may have two genotypes. DD (homozygous) or Dd (heterozygous) of 85% Rh +ve individuals about 35% have DD genotype and 50% have Dd genotype.
- The genotype of Rh –ve individual is dd.
- Therefore, the genotype (gene composition) of offspring will be:
  - DD when gene D is carried by both sperm and ovum
  - Dd when one gamete carries D and other d and
  - dd, when both the gametes carry gene d. Inheritance of Rh antigen is summarized in Fig. 3.6-2.

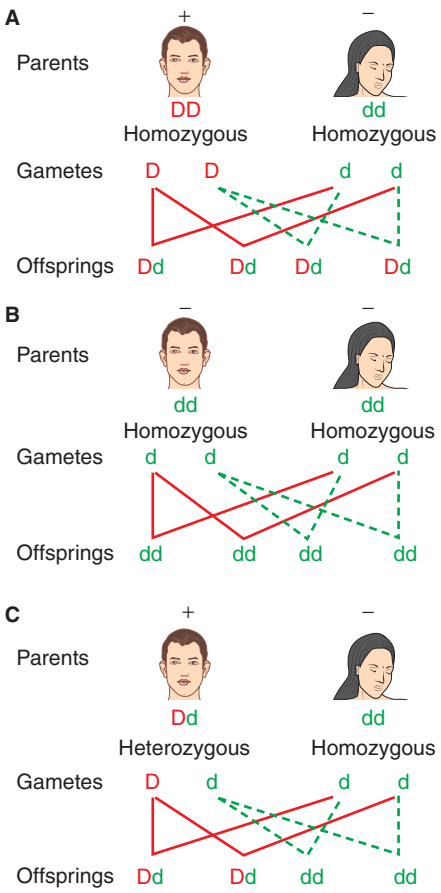


Fig. 3.6-2 Inheritance of Rh antigen: A, when father is homozygous Rh +ve and mother is homozygous Rh –ve; then all the offsprings are heterozygous Rh +ve; B, when father and mother both are homozygous Rh –ve, then all the offsprings are homozygous Rh –ve and C, when father is heterozygous Rh +ve and mother is homozygous Rh –ve, then 50% of offsprings are Rh +ve (heterozygous) and other 50% are Rh –ve (homozygous).

HAEMOLYTIC DISEASE OF NEWBORN

Haemolytic disease of newborn occurs as a result of incompatibility of Rh blood groups between the mother and the fetus.

Mechanism of haemolytic disease of newborn in Rh incompatibility

Mechanism of development of haemolytic disease of the newborn can be described under following steps:

1. Entrance of Rh +ve fetal RBCs into Rh –ve mother's circulation during first pregnancy. At the time of delivery, the fetal RBCs enter maternal circulation because of severance of umbilical cord. Before delivery, usually the foetal and

maternal circulation do not mix. Since the Rh +ve RBCs enter maternal circulation during delivery, so the first child is usually normal.

2. Production of Rh antibodies (anti-D) in mother. During postpartum period, i.e. within a month after delivery, the mother develops Rh antibodies in her blood. As mentioned earlier, the Rh antibodies are of IgG type and are able to cross the placental barrier. Once formed the Rh antibodies persist for a long period in mother's blood.

3. Rh incompatibility reaction during second pregnancy. When the Rh –ve mother in the second pregnancy also bears a Rh +ve child, the Rh antibodies present in the mother's blood enter the fetal circulation by crossing the placental barrier and cause agglutination of fetal RBCs leading to haemolytic disease of newborn.

Manifestations of haemolytic disease of newborn

Depending upon the severity, the haemolytic disease of newborn may manifest as:

- Erythroblastosis fetalis,
- Icterus gravis neonatorum,
- Kernicterus and
- Hydrops fetalis.

1. Erythroblastosis fetalis is characterized by:
- *Erythroblastosis,* i.e. appearance of large number of erythroblasts in the peripheral blood.
- *Anaemia* occurs due to excessive haemolysis of RBCs by Rh antibodies. Infant may even die of severe anaemia.

2. Icterus gravis neonatorum
- *Jaundice* may occur within 24 h of birth due to excessive formation of bilirubin as a result of excessive haemolysis of RBCs.
- *Liver and spleen* are enlarged.

3. Kernicterus. It is a neurological syndrome occurring in newborns with severe haemolysis. The excessive bilirubin formed may enter the brain tissue as the blood–brain barrier is not well developed in infants and cause damage. The bilirubin mostly affects the basal ganglia producing disturbance of motor activities. It usually develops when serum bilirubin level exceeds 18 mg/dL.

4. Hydrops fetalis, i.e. the fetus is grossly oedematous. It occurs when haemolysis is very severe. Usually, there occurs intrauterine death of fetus or if born prematurely or even at term, the infant dies within a few hours.

Prevention and treatment

Prevention of haemolytic disease of newborn. The haemolytic disease in the newborn during second pregnancy can be prevented by injecting single dose of Rh antibodies (anti-D) in the form of Rh-immunoglobulin to mother soon after child birth (1st pregnancy). These antibodies will destroy the Rh +ve RBCs of the fetus which have gained access to maternal circulation. In this way active antibodies will not be formed by the mother.

Treatment of haemolytic disease of newborn. Treatment of haemolytic disease of the newborn is replacement of baby's Rh +ve blood with Rh –ve blood exchange transfusion.

CLINICAL APPLICATIONS OF BLOOD GROUPING

1. In blood transfusion. Before blood transfusion always crossmatching (see page 170) is done.

2. In preventing haemolytic disease in newborn due to Rh incompatibility (as discussed above).

3. In paternity disputes. The ABO and Rh blood grouping is helpful in settling cases of disputed paternity.

Antigens A and B are dominant, whereas O is recessive. It is possible to prove that a person could not have been the father but not that he was or is father; e.g.

- If the child's blood group is O, whatever the blood group of the mother, a person with blood group AB cannot be father.
- If the child's blood group is AB, whatever the blood group of the mother, a person with blood group O cannot be the father (Table 3.6-4). The predictive value of such a test is strengthened further if several blood group systems are considered. DNA fingerprinting can prove or disprove fatherhood with 100% certainty.

4. In medicolegal cases. Any red stain on clothing may be claimed to be blood by a supposed victim. Therefore, it is first confirmed that it really is blood by preparing hemin crystals from the stain extract.

Table 3.6-4 Predictive blood groups of parents in paternity disputes

Blood group of child	Parents must have given blood group	Mother's blood group	Father not have been of blood group
O	O+O	No matter which	AB
AB	A+B	No matter which	O
A	A+O or A+A	B or O	B or O
B	B+O or B+B	A or O	A or O

Note. The child's true ABO typing may not be set until one year of age.

5. In knowing susceptibility to diseases. The incidence of certain diseases is related to blood groups, e.g.

- Individuals with blood group O (non-secretors) are said to be more susceptible to duodenal ulcer (peptic ulcer).
- Individuals with blood group A are more susceptible to carcinoma of stomach.

BLOOD TRANSFUSION

INDICATIONS

Blood transfusion is a life saving measure and should be carried out when it is absolutely essential. Common situations in which blood transfusion is indicated are:

1. ***Blood loss.*** Severe blood loss is the most important indication for blood transfusion.
2. ***For quick restoration of haemoglobin*** in patients with severe anaemia which is required in situations like pregnancy and emergency surgery.
3. ***Exchange transfusion*** is required in haemolytic disease of newborn.
4. ***Blood diseases*** like aplastic anaemia, agranulocytosis, leukaemias, haemophilia, purpura and clotting defects may require blood transfusion.
5. ***Acute poisoning,*** e.g. carbon monoxide poisoning.

DONOR AND RECIPIENT

Donor refers to a person who donates the blood and the person who receives blood is a recipient.

Precautions to be taken while selecting a donor are:

- Donor should be healthy and aged between 18 and 60 years.
- Pregnant and lactating mothers preferably should not donate blood.
- Donor should be screened to exclude the diseases which are spread through blood such as AIDS, viral hepatitis, malaria and syphilis.
- Haemoglobin and packed cell volume (PCV) of the donor should be within normal range. Its approximate concentration is tested.

Universal donor. Blood of the individuals with blood group O does not contain any agglutinogen. So when this blood is transfused to a person with any blood group (A, B, AB or O), theoretically its RBCs will not be agglutinated. Because of this fact, an individual with blood group O is called universal donor. However, practically this term is no longer valid, as it ignores the complications produced by existence of Rh factor and other blood group systems.

Universal recipient. Blood of an individual with blood group AB does not contain any agglutinins. So, theoretically when such an individual receives blood from the individual with any blood group (A, B, AB or O), there should be no transfusion reaction. Because of this fact an individual with AB blood group is called universal recipient. However, practically this term is no more valid because it ignores the complications produced by the existence of Rh factor and other blood group systems.

PRECAUTIONS TO BE OBSERVED DURING BLOOD TRANSFUSION

1. ***Absolute indication*** should always be there for the transfusion of blood.
2. ***Crossmatching*** should always be done before the blood transfusion. For it blood is collected from donor as well as recipient. Plasma and RBCs are separated in each. The crossmatching involves two steps: major and minor crossmatching.
   - *Major crossmatching involves mixing of donor's cells with recipient's plasma.* This is called major crossmatching because of the fact that when mismatched blood is transfused in a recipient, the donor's cells get agglutinated as against their agglutinogen there is sufficiently high concentration of agglutinins in the recipient's plasma.
   - *Minor crossmatching involves mixing of recipient's cells with donor's plasma.* This is called minor crossmatch due of the fact that reaction of donor's plasma and recipient's cells usually does not occur or is very very mild on giving mismatched blood transfusion because:
   - Firstly, the donor's plasma in the transfusion (about 250 mL) is usually so diluted by the much larger volume of recipient's blood (about 5 L) that it rarely causes agglutination even when the titre of agglutinins against the recipient's cell is high, and
   - Secondly, donor's agglutinins are also neutralized by soluble agglutinogen which are found free in the recipient's body fluid.
3. ***Rh +ve blood should never be transfused to Rh –ve person.*** It is particularly must for females at any age before menopause, because once she is sensitized by the Rh antigen, the anti-D antibodies are formed and she will not be able to bear a Rh +ve fetus. In other words, Rh +ve transfusion may make a woman permanently childless.
4. ***Donor's blood should always be screened*** for diseases which are spread through blood, such as AIDS, hepatitis B, malaria and syphilis.
5. ***Blood bag/bottle should be checked*** for the name of recipient and blood group on the label before starting the blood transfusion.
6. ***Blood transfusion should be given at slow rate.*** If rapid transfusion is given, citrate present in stored blood may

cause chelation of calcium ions leading to decreased serum calcium level and tetany.

7. ***Proper aseptic measures*** must be taken during transfusion of blood.
8. ***Careful watch on recipient's condition*** is must for the first 10–15 min of starting the transfusion and from time to time later.

HAZARDS OF BLOOD TRANSFUSION

1. ***Mismatched transfusion reactions.*** Mismatched transfusion reaction is the most serious and potentially fatal hazard of blood transfusion. It is characterized by showing effects of inter-group blood transfusion:
   - *Agglutination* of donor's red blood cells in the recipient circulation (Table 3.6-5).
   - *Tissue ischaemia* occurs due to blockage of certain vessels by the agglutinated cells. Soon patient complains of violent pain in back or elsewhere and tightness of chest.
   - *Haemolysis* of agglutinated red cells occurs rapidly releasing large amount of haemoglobin in circulation (haemoglobinaemia).
   - *Haemolytic jaundice* may occur due to excessive formation of bilirubin from haemoglobin released by haemolysed RBCs.
   - *Renal vasoconstriction* is caused by toxic substances released from the haemolyzed RBCs.
   - *Circulatory shock* occurs due to loss of circulating red cells and release of toxic substances leading to fall in arterial blood pressure and decreased renal blood flow.
   - *Haemoglobinuria* occurs when total free haemoglobin becomes more than that can bind with haptoglobin (plasma protein binding haemoglobin). The extra free haemoglobin leaks through glomerular membrane and is passed in urine producing haemoglobinuria.
   - *Renal tubular damage.* If urine is acidic and glomerular filtration is slow, the free haemoglobin passing through glomeruli is precipitated in the tubules as acid haematin. This obstructs the lumen of tubules producing renal tubular damage.
   - *Acute renal shut down (anuria)* sets in ultimately due to the combined effects of renal vasoconstriction, circulatory shock, hypotension and renal tubular damage. Acute renal shut down usually occurs within a few minutes to few hours after transfusion of mismatched blood and continues.
   - *Uraemia* (increased nitrogenous substances and potassium in the body) results due to acute renal failure, soon producing coma and death.
2. ***Circulatory overload*** due to hypervolaemia may occur following blood transfusion when the transfusion is rapid specially in patients with cardiac diseases.
3. ***Transmission of blood-borne infections*** such as AIDS, viral hepatitis, malaria, syphilis, etc. may occur to recipient from the infected donor.
4. ***Pyrogenic reaction*** characterized by fever and chills may occur probably due to destruction of leucocytes and platelets by antibodies against them.
5. ***Allergic reactions*** such as skin rashes and asthma may occur if donor blood contains substances to which patient is allergic.
6. ***Hyperkalaemia*** may occur after excessive transfusion because K^+ concentration in stored blood is high. Owing to leakage of K^+ from the RBCs into the plasma.
7. ***Hypocalcaemia*** producing tetany may occur following massive transfusion of citrated blood.

Table 3.6-5 Effects of intergroup blood transfusion

Blood group of recipient	Agglutinin in plasma	Donor's RBCs (antigen)			
		AB	A	B	O
AB	Nil	–	–	–	–
A	β	+	–	+	–
B	α	+	+	–	–
O	αβ	+	+	+	–

Note. Agglutination (+) of RBCs (incompatibility). No agglutination (–) of RBCs (compatibility).

AUTOLOGOUS BLOOD TRANSFUSION

Autologous blood transfusion refers to transfusion of an individual's own blood which has been withdrawn and stored. Autologous transfusion is done under the following situations:

For elective surgery, a self-predonation is a common practice in some hospitals.

During surgery, the cell-saver machine when used sucks up the blood from the wound, recycles it and returns it to the patient's body.

STORAGE OF BLOOD FOR TRANSFUSION

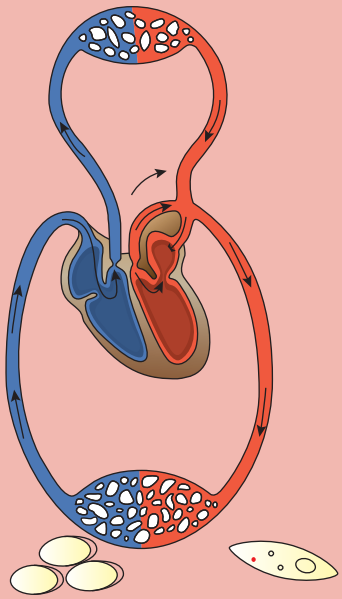
Some facts about the storage of donated blood are:

- *One unit of blood* (420 mL) can be collected from a donor at a time under all aseptic measures. An individual can safely donate one unit of blood every 6 months. Acid-citrate-dextrose (ACD) mixture (120 mL) is added to blood and is stored in sterile container.
- *Contents of ACD mixture* are:
  - Acid citrate (monohydrous), 0.48 g,
  - Trisodium citrate, 1.32 g,
  - Dextrose 1.47 g and
  - Distilled water 100 mL.

- *Dextrose (glucose) present in ACD mixture provides* energy for maintenance of sodium–potassium pump activity.
- *Anticoagulant activity* is provided by the citrates present in the ACD mixture which also decreases the pH of blood.
- The blood can be stored under above conditions *up to 21 days.*
- *The RBCs in the stored blood* swell up due to the following changes as a result of decreased cell metabolism in cold storage.
  - *Loss of intracellular* K^+, which increases plasma K^+ concentration from 4–5 mEq/L to 20–30 mEq/L,
  - *Increase in intracellular* Na^+ from 12 mEq/L to 30–40 mEq/L
  - *Increase in intracellular water content.* Because of the above changes the RBCs become more spherocytic and their haemoglobin in hypotonic solution increases. Such cells may rupture in vitro even in 0.8% NaCl solution. With reference to Na^+ and K^+ content, volume, shape and saline fragility, the RBCs become normal within 48 h of transfusion.
- *WBCs and platelets in stored blood* are virtually absent after 24 h of storage. Therefore, stored blood is not a suitable medium for transferring WBCs and platelets to a recipient.
- *After transfusion of stored blood,* 80% RBCs survive for 24 h and thereafter surviving cells are destroyed at a rate of 1% per day.

SECTION 4

Cardiovascular System



The cardiovascular system consists of the heart and the blood vessels. The *heart* acts as a system of two pumps working in series and forms the driving force for blood flow. The blood vessels that take blood from the heart to various tissues are called *arteries*. The smallest arteries are called *arterioles*. Arterioles open into a network of *capillaries* which constitute the microcirculation. The most important function of blood vessels, i.e. the rapid exchange of materials between the blood and extracellular fluid bathing the tissue cells is served by the capillaries. In other words, capillaries serve as the *exchange region*. In some situations, capillaries are replaced by slightly different vessels called *sinusoids*. Blood from capillaries (or from sinusoids) is collected by small *venules* which join to form *veins*. Veins serve as the blood reservoir and return the collected blood to the heart.

FUNCTIONS OF CARDIOVASCULAR SYSTEM

Primary functions of the cardiovascular system are:

1. Distribution of nutrients and oxygen (O_2) to all body cells and
2. Collection of waste products and CO_2 from different body cells and to carry them to excretory organs for excretion.

Secondary functions that are subserved by the cardiovascular system are:

1. Thermoregulation,
2. Distribution of hormones to the target tissues and
3. Delivery of antibodies, platelets and leucocytes to aid body defence mechanism.

PHYSIOLOGY OF CARDIOVASCULAR SYSTEM

Physiology of cardiovascular system includes various aspects of physiology of heart as a pump and physiology of two main divisions of blood circulation—the pulmonary and systemic circulation.

- The heart consists of two pumps in series (right and left halves) that are connected by pulmonary and systemic circulation.
- In systemic circulation, the various systemic organs receive blood through parallel distribution channels. The parallel arrangement of vessels supply the body organs with blood of the same arterial composition (i.e. same O_2 and CO_2 tension, pH, glucose level) and essentially the same arterial pressure.
- Since the pulmonary and systemic circulation divisions are arranged in series, so both ventricles must pump the same amount of blood over any significant time period. Such a balanced output is achieved by an intrinsic property of cardiac muscle known as *Frank-Starling* mechanism.

Chapter 4.1

Functional Anatomy of Heart and Physiology of Cardiac Muscle

FUNCTIONAL ANATOMY OF HEART
- Chambers of heart
- Valves of heart
- Structure of the walls of heart

PHYSIOLOGY OF CARDIAC MUSCLE
- Structural organization of cardiac muscle
- Structure of a cardiac muscle fibre
- Sarcotubular system
- **Process of excitability and contractility**
  - Electrical potentials in cardiac muscle
  - Excitation–contraction coupling phenomenon in cardiac muscles
  - Process of cardiac muscle contraction
  - Relaxation of cardiac muscle
- **Properties of cardiac muscle**
  - Excitability
  - Contractility

FUNCTIONAL ANATOMY OF HEART

The heart is a muscular pump designed to ensure the circulation of blood through the tissues of the body. The human heart weighs approximately 300 g and it consists of two halves, right and left. The right heart circulates blood through the lungs for the purpose of oxygenation (i.e. through pulmonary circulation). The left heart circulates blood to the tissues of the entire body (i.e. through the systemic circulation).

CHAMBERS OF HEART

Each half of the heart consists of an inflow chamber called the *atrium* and an outflow chamber called the *ventricle* (Fig. 4.1-1). Thus, there are four chambers in the heart.

Atria

Interatrial septum separates the right and left atria which are thin walled chambers.

Right atrium receives deoxygenated blood from the tissues of the entire body through the *superior and inferior vena cavae.* This blood passes into the right ventricle through the right *atrioventricular orifice* which is guarded by a *tricuspid valve.* The right atrium has got the pacemaker known as *sinoatrial node* that produces cardiac impulses and *atrioventricular* node that conducts these impulses to the ventricles.

Left atrium receives oxygenated blood from the lungs through the four *pulmonary veins* (two right and two left). This blood passes into left ventricle through the *left atrioventricular orifice* which is guarded by the mitral valve.

Ventricles

Interventricular septum separates the right ventricle from the left ventricle.

Interior of each ventricle has an inflow part and an outflow part. *Papillary muscles* are finger-like processes attached to the ventricular wall at one end but free at the other. They are functionally related to the atrioventricular valves.

Right ventricle receives blood from the right atrium and pumps through the pulmonary trunk (which divides into

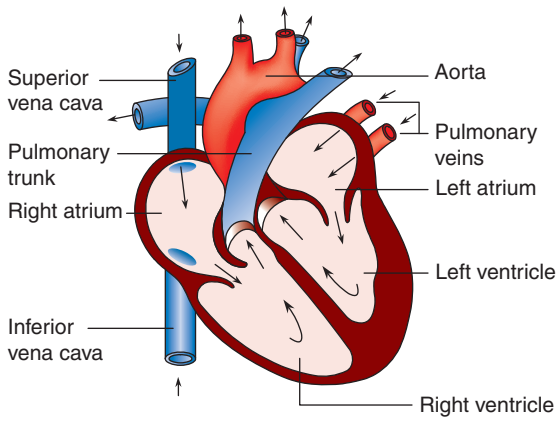


Fig. 4.1-1 Schematic diagram of the heart to show its chambers.

right and left pulmonary arteries) into the lungs. The *pulmonary valve* is present at the junction of right ventricle and pulmonary trunk.

Left ventricle receives blood from the left atrium and pumps out into systemic circulation through the aorta. *Aortic* valve is present at the junction of left ventricle and the *ascending aorta.*

The wall of left ventricle is three times thicker than of the right ventricle (Physiological hypertrophy) as left ventricle has to do more work to pump the blood to whole body.

VALVES OF HEART

There are four valves in a human heart, two atrioventricular valves and two semilunar valves. Valves allow unidirectional flow of blood.

Atrioventricular valves

The atrioventricular valves open towards the ventricles and close towards the atria. They allow blood to flow from atria to ventricles. But when ventricles contract, they are closed and thus prevent backflow of blood from ventricles to atria. These valves passively open and close due to pressure gradient.

- The *right atrioventricular valve* is known as *tricuspid valve* and is made of three cusps: anterior, posterior and septal (Fig. 4.1-2).
- The left atrioventricular valve is called *mitral valve* or *bicuspid valve* and is made of two cusps: anterior and posterior (Figs 4.1-2 and 4.1-3).
- At the periphery, the cusps (flaps) of the atrioventricular valves are attached to the *atrioventricular ring,* which is the fibrous connection between the atria and ventricles. The free edges of the cusps are attached to the *papillary muscles* through the cord-like structures called the *chordae tendineae* (Fig. 4.1-3).
- *Papillary muscles* arise from the inner surface of ventricles and contract when the ventricular walls contract. They do not help the valves to close but prevent the bulging of the valves into the atria when ventricles contract.

Semilunar valves

- *Aortic valve* is the semilunar valve present at the opening of aorta in the left ventricle. It is made of three semilunar cusps: one anterior and two posterior (Fig. 4.1-2). These valves are adapted to withstand physical trauma of high pressure in aorta and high velocity of blood flow during the ventricular systole (rapid ejection phase).
- *Pulmonary valve* is the semilunar valve present at the opening of pulmonary trunk into the right ventricle. It is also made of three semilunar cusps: one posterior and two anterior (Fig. 4.1-2).
- *Semilunar valves open* away from the ventricles and close towards the ventricles. These valves open when ventricles contract allowing the blood to flow from the left ventricle to aorta and from the right ventricle to the pulmonary trunk.
- *Semilunar valves close* when ventricles relax thus preventing backflow of blood from aorta or pulmonary trunk into the ventricles.
- Opening of the semilunar valves is a slow process. While closure is a sudden process causing neighbouring fluid to vibrate resulting in noise which is heard as heart sounds.

STRUCTURE OF THE WALLS OF HEART

Walls of the heart are composed of thick layer of cardiac muscle, the *myocardium*, covered externally by the epicardium and lined internally by the *endocardium.*

- Walls of the atrial portion of the heart are thin.
- Walls of the ventricular portion of the heart are thick.

Skeleton of the heart consists of fibrous rings that surround the atrioventricular, pulmonary and aortic orifices and are continuous with the membranous part of the ventricular septum. The fibrous rings around the atrioventricular orifices separate the muscular walls of the atria from those of the ventricles but provide attachment for the muscle fibres. The fibrous rings support the bases of the valve cusps and prevent the valves from stretching and becoming incompetent.

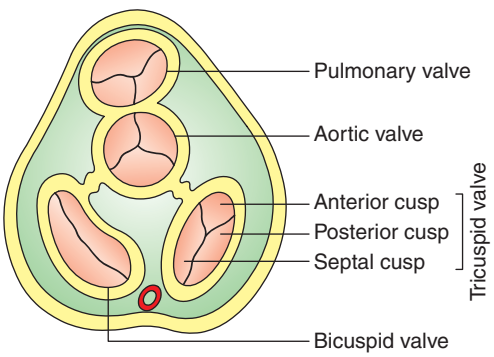


Fig. 4.1-2 Valves of the heart.

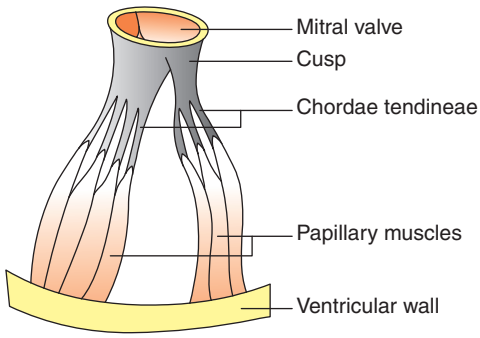


Fig. 4.1-3 Bicuspid valve attached with papillary muscles and chordae tendineae.

Pericardium

The heart and roots of the great vessels are enclosed by a fibroserous sac called pericardium. Its function is to restrict excessive movements of the heart as a whole and to serve as a lubricated container in which different parts of the heart can contract. Pericardium consists of two layers: outer fibrous and inner serous (Fig. 4.1-4).

Fibrous pericardium surrounds the heart like a bag and is attached with the surrounding structures.

Serous pericardium has parietal and visceral layers. The *parietal layer* of serous pericardium lines the fibrous pericardium and is reflected around the roots of the great vessels to become continuous with the *visceral layer of serous pericardium* that closely cover the heart and is often called the *epicardium.* The slit-like space between the parietal and the visceral layers of the serous pericardium is called *pericardial cavity* which contains small amount of *pericardial fluid* (5–30 mL) that acts as a lubricant to facilitate movement of the heart.

Myocardium

The myocardium (muscular tissue of the heart) is the main tissue constituting the walls of the heart. It consists of three types of muscle fibres:

- *Cardiac muscles* forming the walls of the atria and ventricles.
- Muscle fibres forming the *pacemaker* which is the site of origin of cardiac impulse.
- Muscle fibres forming the conducting system which transmits the impulse to the various parts of the heart.

Endocardium

Endocardium is thin, smooth and glistening membrane lining the myocardium internally. It consists of a single layer of endothelial cells. The endocardium continues as the endothelium of great vessels opening in the heart.

PHYSIOLOGY OF CARDIAC MUSCLE

STRUCTURAL ORGANIZATION OF CARDIAC MUSCLE

- The cardiac muscle fibres are *striated* and resemble quite a lot to the skeletal muscle fibres in structure. However, unlike the skeletal muscles, the cardiac muscles are *involuntary* (like smooth muscles). Thus cardiac muscles share some characteristics with the skeletal muscles and others with the smooth muscles.
- The cardiac muscle fibres are *ribbon-like* rather than cylindrical. These are *branched* and interdigitate freely with each other, but each fibre is a completely separate unit. The branches from the neighbouring fibres join together. At the point of contact of two muscle fibres, the membranes of both the muscle fibres are fused together and thrown into an extensive infolding forming the so-called *intercalated disc* (Fig. 4.1-5). These discs form tight junctions between the muscle fibres and do not allow the ions to pass through. However, the intercalated discs provide a strong union between fibres and thus play an important role during the contraction of muscle fibres by transmitting pull of one contractile unit along its axis to the next, thereby increasing force of contraction.
- Along the sides near the outer border of intercalated disc, the two adjacent muscle fibres are connected with each other through the *gap junctions.* The action potential passes from one cardiac muscle cell to the other through gap junctions which provide *low resistance bridges* and thus the cardiac muscle acts as *a functional syncytium* of many cardiac cells. In this way, the cardiac impulse spreads throughout the muscle mass quickly resulting in a co-ordinated

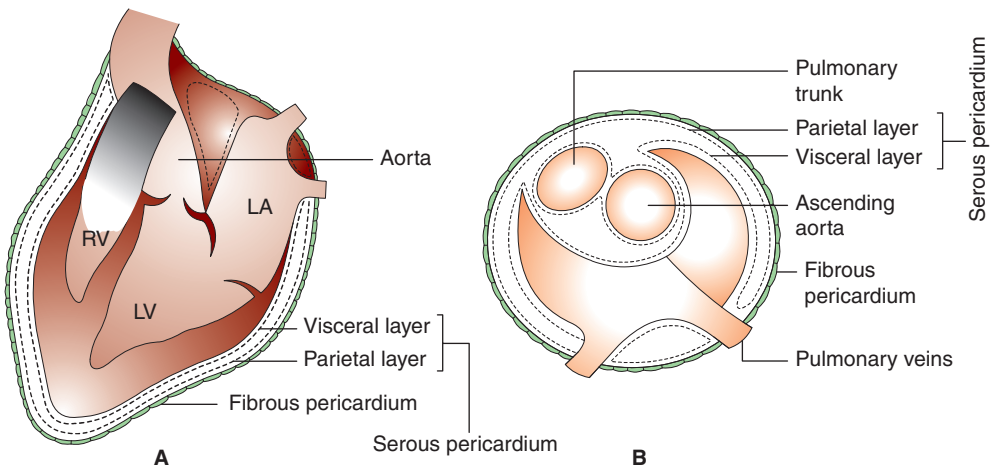


Fig. 4.1-4 Schematic sagittal section of the heart showing fibrous and serous pericardium.

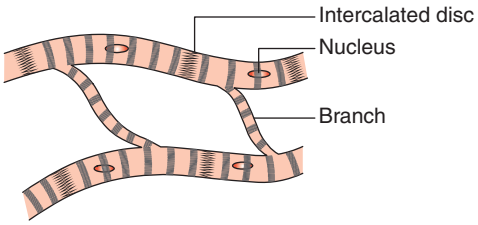


Fig. 4.1-5 Structure of cardiac muscle.

contraction of the whole tissue. In the heart, the cardiac muscle forms two separate syncytia, i.e. the *atrial syncytium* (walls of two atria) and the *ventricular syncytium* (walls of the two ventricles). Action potential is conducted from the atrial syncytium to the ventricular syncytium by way of specialized conducting system. Each syncytium obeys all or none law. Because the atrial and the ventricular syncytium are two separate syncytia, therefore, atria contract a short time ahead of the ventricular contraction.

- The cardiac muscle fibres are richly supplied by the capillaries (one capillary/fibre).

STRUCTURE OF A CARDIAC MUSCLE FIBRE

Each muscle fibre is about 80–100 µm long and about 15 µm broad. Its cell membrane is called *sarcolemma* and the cytoplasm is called *sarcoplasm.* The sarcoplasm is in abundance and contains all the cell organelles, a well-developed sarcoplasmic reticulum and a centrally placed nucleus. Each muscle fibre is made up of number of *myofibrils* which lie parallel to each other. Each myofibril is 2 µm is diameter.

Myofibril. Each myofibril consists of thick and thin filaments. Essentially, the structure and striations as seen under light microscope and the detailed electron microscopic structure are similar to that of a skeletal muscle (see page 67).

SARCOTUBULAR SYSTEM

The sarcotubular system in the cardiac muscles is well developed like that of the skeletal muscle (page 69). However, the tubules of the T-system penetrate the sarcomere at Z-line (Fig. 4.1-6). Therefore, in cardiac muscles, there is only one triad per sarcomere as compared to two in skeletal muscle (page 70).

PROCESS OF EXCITABILITY AND CONTRACTILITY: AN ELECTROMECHANICAL PHENOMENON

The cardiac muscle being an excitable tissue produces an action potential *(electrical phenomenon)* when stimulated

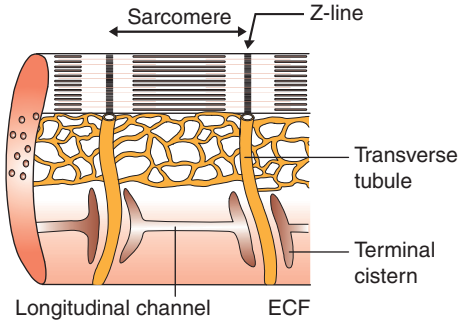


Fig. 4.1-6 Sarcotubular system in the cardiac muscle.

and responds by contracting *(mechanical phenomenon).* The events which link the electrical phenomenon with mechanical phenomenon constitute the *excitation–contraction coupling phenomenon.* These three phenomena are discussed separately.

ELECTRICAL POTENTIALS IN CARDIAC MUSCLE

Resting membrane potential

The resting membrane potential of a normal cardiac muscle fibre is –85 to –95 mV (negative interior with reference to exterior).

Action potential

When stimulated, each cardiac muscle fibre shows an electrical activity known as propagated action potential. It is different from the electrocardiogram, which refers to the extracellular recording of the summed electrical events of all the cardiac muscle fibres generated with each heart beat.

The action potential recorded from a single cardiac muscle fibre is unusually long and can be divided into five distinct phases (Fig. 4.1-7):

Phase 0: Rapid depolarization. The phase 0 (upstroke) is characterized by the depolarization which proceeds rapidly, an overshoot is present, as in skeletal muscle and nerve. In mammalian heart, depolarization lasts about 2 ms. In this phase, amplitude of potential reaches up to +20 to +30 mV (positive interior with reference to exterior).

Ionic basis. The initial rapid depolarization and the overshoot are due to the rapid opening of voltage-gated Na^+ channels and rapid influx of Na^+ ions similar to that occurring in the nerve and the skeletal muscle.

At –30 to –40 mV membrane potential the calcium channels also open up and influx of Ca^{2+} ions also contributes in this phase.

Duration of depolarization is 2 ms and is followed by repolarization which occurs in three phases.

Phase 1: Initial rapid repolarization. Rapid depolarization is followed by a very short-lived slight rapid repolarization.

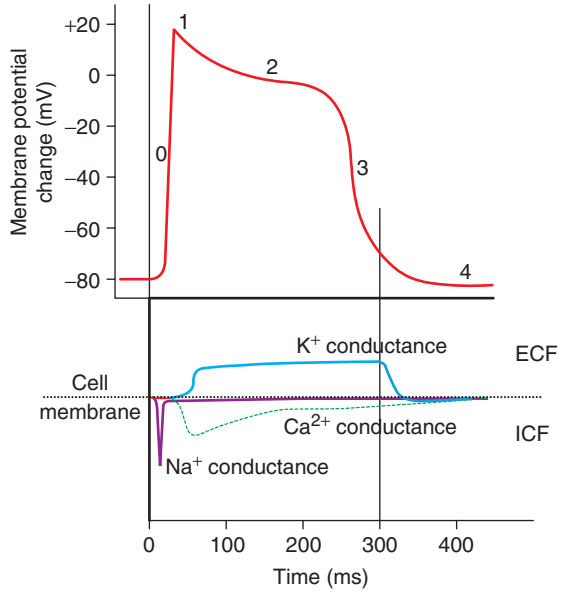


Fig. 4.1-7 Various phases of action potential and ion conductance: Phase 0 = depolarization; Phase 1 = rapid repolarization; Phase 2 = plateau phase; Phase 3 = late rapid repolarization and Phase 4 = resting potential.

The membrane potential reaches from +30 mV to –10 mV during this phase.

Ionic basis. The initial rapid repolarization is due to closure of Na^+ channels and opening of K^+ channels resulting in transient outward current.

Phase 2: Plateau. During plateau phase, the cardiac muscle fibre remains in the depolarized state. The membrane potential falls very slowly only to –40 mV during this phase. The plateau lasts for about 100–200 ms. This plateau in action potential explains the 5–15 times longer contraction time of the cardiac muscle as compared to skeletal muscle.

Ionic basis. Very slow repolarization during the plateau phase is due to:

- Slow influx of Ca^{2+} ions resulting from opening of sarcolemmal *L-type* Ca^{2+} channels.
- Closure of a distinct set of K^+ channels called the inward rectifying K^+ channels.

Phase 3: Repolarization. During this phase, complete repolarization occurs and the membrane potential falls to the approximate resting value. This phase lasts for about 50 ms.

Ionic basis. The slow repolarization results from the closing of Ca^{2+} channels and opening of following two types of K^+ channels:

- *Delayed outward rectifying* K^+ channels, which are voltage-gated and are activated slowly.
- *Ca^{2+} activated channels* which are activated by the elevated sarcoplasmic Ca^{2+} levels.

Phase 4: Resting potential. In this phase of resting membrane potential (also called as polarised state), the potential is maintained at –90 mV

Ionic basis. The resting membrane potential is maintained by a resting K^+ current, the largest contributor to which is the inward rectifying K^+ current. The resting ionic composition is restored by Na^+–K^+ ATPase pump.

Duration of action potential

The duration of action potential is about 250 ms at a heart rate 75 beats/min. The duration of action potential decreases with increased heart rate (150 ms at a heart rate of 200 beats/min). This type of action potential found in contractile myocardial cells of the ventricles is referred to as *fast response.*

Spread of action potential through cardiac muscle

The cardiac muscle acts as a physiological syncytium due to the presence of gap junctions amongst the cardiac muscle fibres. Because of this, the action potential spreads through the cardiac muscles very rapidly. Further, as there are two syncytia (the atrial and the ventricular) in the heart, so the action potential is transmitted from atria to ventricles only through the fibres of specialized conductive system.

EXCITATION–CONTRACTION COUPLING PHENOMENON IN CARDIAC MUSCLES

Excitation–contraction coupling refers to the sequence of events by which an excited plasma membrane of a muscle fibre leads to cross-bridge activity by increasing sarcoplasmic calcium concentration.

The sequence of events during excitation–contraction coupling in the cardiac muscle is similar to those observed in a skeletal muscle (see page 70) with the following exception:

In cardiac muscle (as against that in skeletal muscle), extra calcium ions diffuse into the sarcoplasm from T-tubules (Fig. 4.1-8) without which the contraction strength would be considerably reduced. The T-tubules of cardiac muscle contain mucopolysaccharides, which are negatively charged and bind an abundant store of calcium ions. T-tubules open directly to the exterior and therefore, Ca^{2+} ions in them directly come from the extracellular fluid (ECF). These Ca^{2+} ions diffuse into the sarcoplasm when action potential propagates along the T-tubules (Fig. 4.1-8A, B). Because of this, strength of cardiac muscle contraction depends to a great extent on Ca^{2+} concentration in the ECF. Whereas, the skeletal muscle contraction is hardly affected by calcium concentration in the ECF.

PROCESS OF CARDIAC MUSCLE CONTRACTION

The molecular mechanism of cardiac muscles contraction by cross-bridge cycling and sliding of filaments primarily

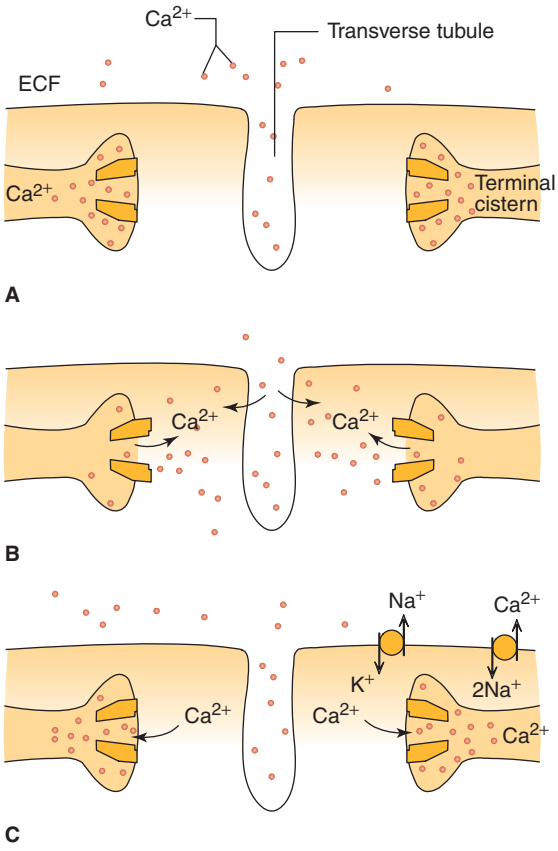


Fig. 4.1-8 Dynamics of Ca^{2+} during excitation–contraction coupling phenomenon and relaxation in cardiac muscle: A, in resting state; B, calcium-induced calcium release during excitation and contraction state and C, during relaxation state.

similar to that of skeletal muscles (page 71) and smooth muscles. However, in cardiac muscle:

- Troponin–tropomyosin complex controls the onset and offset of cross-bridge cycling, similar to that in the skeletal muscles.
- Like smooth muscles, the contractility of cardiac muscle is sensitive to *phosphorylation.*

RELAXATION OF CARDIAC MUSCLE

Relaxation of cardiac muscle (diastole) occurs when levels of Ca^{2+} ions fall in the cardiac muscle fibres. During diastole, the Ca^{2+} ions are extruded out of the cardiac muscle fibre by a carrier system operating at the sarcolemma in which two Na^+ ions are exchanged for each Ca^{2+} ion extruded (Fig. 4.1-8C). Thus, the rate of Ca^{2+} ion extrusion depends on the gradient of Na^+ created by Na^+–K^+ ATPase.

Inhibition of this secondary active transport of Ca^{2+} ions, e.g. by digitalis or other cardiac glycosides, raises the intracellular Ca^{2+} concentration and thereby increases myocardial contractility. This effect is utilized in the patients with congestive heart failure.

PROPERTIES OF CARDIAC MUSCLE

The basic properties of a cardiac muscle include:

- Automaticity,
- Rhythmicity (chronotropism),
- Conductivity (dromotropism),
- Excitability (bathmotropism) and
- Contractility (inotropism).

Some of the properties of cardiac muscle, viz. automaticity, rhythmicity and conductivity are discussed in Chapter 4.2 (see page 185).

The characteristics of excitability and contractility are described here.

EXCITABILITY

Excitability (bathmotropism) is the property by which tissues respond to stimuli. The cardiac muscle responds by the development of action potential. The essential features of the resting membrane potential and action potential of cardiac muscle have been discussed on page 178.

The characteristic of a cardiac muscle excitability which needs a special emphasis is its refractory period.

Refractory period

Refractory period refers to the period following action potential during which the cardiac muscle does not respond to a stimulus. Cardiac muscle has a long refractory period (250–300 ms in ventricles and about 150 ms in atria). It is of two types:

1. Absolute refractory period (ARP). During this period, the cardiac muscle does not show any response at all. It extends from phase 0 to half of phase 3 of action potential, i.e. until the membrane potential reaches approximately 250 mV during repolarization (Fig. 4.1-9). Normal duration of ARP in the ventricles is about 180–200 ms.

2. Relative refractory period. During this period, the muscle shows response if the strength of stimulus is increased to maximum. It extends from second half of the phase 3 to phase 4 of the action potential. Normal duration of relative refractory period in ventricles is about 50 ms.

Experimental demonstration of refractory period in heart

Experimental demonstration of refractory period in heart can be done both in a beating heart as well as in a quiescent heart.

Refractory period in a beating heart of a pithed frog can be demonstrated during recording of a cardiogram. As shown in Fig. 4.1-10, *when the electrical stimulus is applied to the base of the ventricle during systole,* no response is seen, depicting thereby that the heart is in an absolute refractory period during systole.

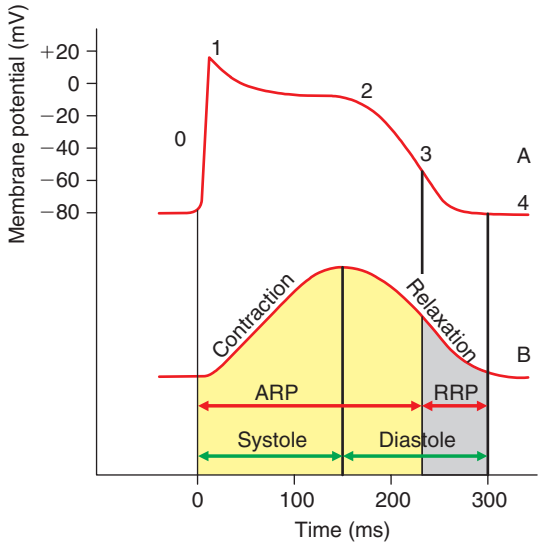


Fig. 4.1-9 Record of action potential (A) and mechanical response (B) from the cardiac muscle fibre shown on same time scale depicting the significance of long refractory period. (ARP=Absolute refractory period; RRP=relative refractory period.)

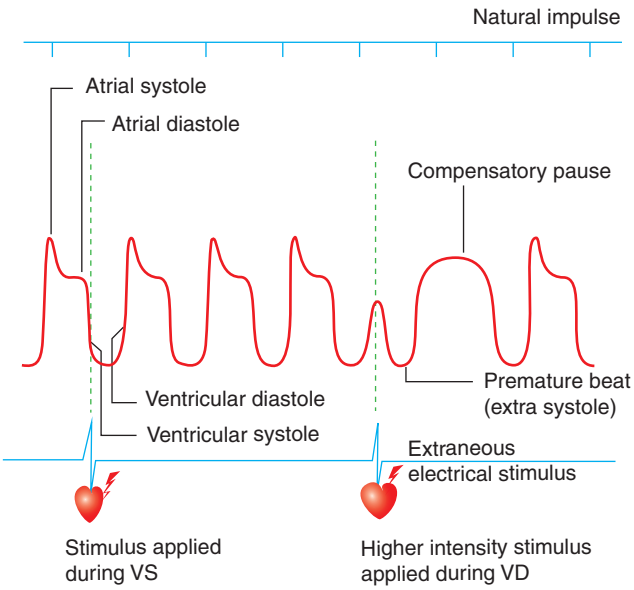


Fig. 4.1-10 Demonstration of refractory period in beating heart of frog. (VS=ventricular systole; VD=ventricular diastole.)

- *When the stimulus is applied during diastole*, the heart responds by a premature contraction producing *extra systole* followed by a *compensatory pause.* The total duration of the extra systole and the compensatory pause is equivalent to the duration of two cardiac cycles. These events can be explained as:
  - When the stimulus is applied during diastole, the premature contraction occurs since the heart is in the phase of *relative refractory period.*
  - The natural impulse from the sinus venosus arrives at the time of premature contraction (*absolute refractory period*) and cannot produce contraction and thus the heart has to wait for the arrival of next natural impulse. The heart stops during this period in relaxation producing the so-called *compensatory pause.*

Significance of long refractory period in cardiac muscle

As shown in Fig. 4.1-9, the cardiac muscle is refractory to any stimulus during the contraction phase (systole), therefore, the complete summation of contractions and thus tetanus cannot be produced in the cardiac muscle. This property is very useful. Since the heart has to function as a pump, it must relax, get filled up with blood and then contract to pump out the blood. A tetanized heart would be useless as a pump.

CONTRACTILITY

Contractility is the ability of the cardiac muscle to actively generate force to shorten and thicken to do work when sufficient stimulus is applied.

Process of myocardial contractility is discussed on page 179.

Mechanical response in cardiac muscle fibre begins just after the start of depolarization and lasts about 1.5 times as long as the action potential (Fig. 4.1-11B). Thus the mechanical response (300 ms) overlaps the electrical response (200 ms) for the whole period. This is in contrast to the skeletal muscle, where the mechanical response begins a few milliseconds after the end of repolarization and lasts for 30–50 ms in the mammalian skeletal muscle and 100 ms in the amphibian skeletal muscle (Fig. 4.1-11A).

Characteristic features of myocardial contractility and factors affecting are:

- All or none law,
- Staircase phenomenon,
- Summation of subminimal stimuli,
- Effect of preload,
- Effect of afterload,
- Effect of ions (see page 204) and
- Effect of temperature.

1. All or none law

The response of a cardiac muscle to a stimulus is all or none in character, i.e. when a stimulus is applied either the heart does not contract at all (none response), or contracts to its maximum ability (all response). This is because of the syncytial arrangement of the cardiac muscle fibres. Therefore, the 'all or none' law in heart is applicable to whole of functional syncytial unit, i.e. the entire atria or entire ventricle. While

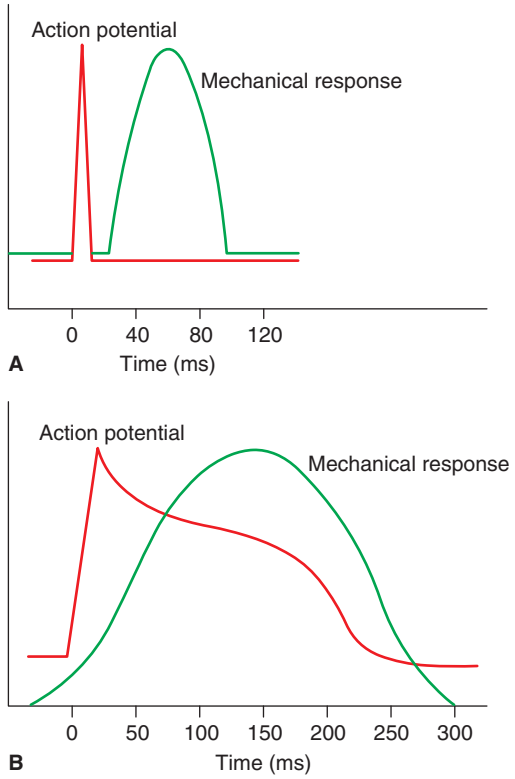


Fig. 4.1-11 Relationship of action potential and mechanical response: A, in skeletal muscle and B, in cardiac muscle.

in skeletal muscle, it is applicable only to a single muscle fibre.

2. Staircase phenomenon (Treppe)

Staircase phenomenon or effect refers to the successive increase in the force of cardiac contractions in first few (4–5) contractions after the quiescent heart starts beating, e.g. as seen after vagal stimulation (Fig. 4.1-12). This is because of the beneficial effect.

Cause of beneficial effect. When the heart stops, there occurs increase in Na^+ and decrease in K^+ concentration inside the cell, this increases Ca^{2+} influx. Thus, there occurs progressive increase in the Ca^{2+} concentration in the sarcoplasm due to increase in Ca^{2+} influx with each action potential. This produces a progressive increase in strength of the few (4–5) cardiac muscle contractions.

3. Summation of subminimal stimuli

When a subthreshold stimulus is applied to the quiescent heart, there occurs no response. However, when subthreshold stimuli are applied repeatedly at an interval of one half to one second, there occurs a contraction of the heart after about 10–20 stimuli. This phenomenon is called *temporal summation* of subminimal stimuli (Fig. 4.1-13).

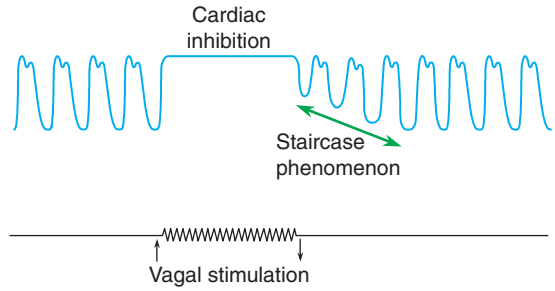


Fig. 4.1-12 Staircase phenomenon (treppe) in the cardiac muscle after vagal stimulation in a frog.

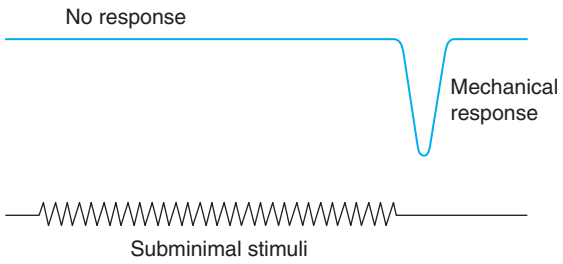


Fig. 4.1-13 Temporal summation of subminimal stimuli in quiescent heart of frog.

4. Effect of preload

A load which starts acting on a muscle before it starts to contract is *called preload.* The preload increases the initial length of the muscle. According to Starling's law, the force of contraction is the function of the initial length of the muscle fibres and up to physiological limits the greater the initial length, greater is the force of contraction. In the case of heart muscle, the end diastolic volume forms the preload. The effect of changing end diastolic volume on force of cardiac contraction has been studied by Frank and Starling in 1910. *The Frank–Starling law of heart* states that within physiological limits the force of cardiac contraction is proportional to its end diastolic volume.

Length–tension relationship

Length–tension relationship, i.e. the relation between the initial fibre length and total tension in cardiac muscle is basically similar to that in the skeletal muscle. In cardiac muscle, the length–tension relationship graph is plotted with the end diastolic volume in mL (representing initial length) along the horizontal axis and the pressure developed in the ventricle in mm Hg (representing tension) along the vertical axis (Fig. 4.1-14). The following inferences can be drawn from this graph (Starling's curve).

Diastolic intraventricular pressure represents the passive tension and it increases with the increase in end diastolic volume (i.e. with the passively increased muscle length). It is important to note that the pressure–volume curve for ventricles in diastole is initially quite flat, indicating

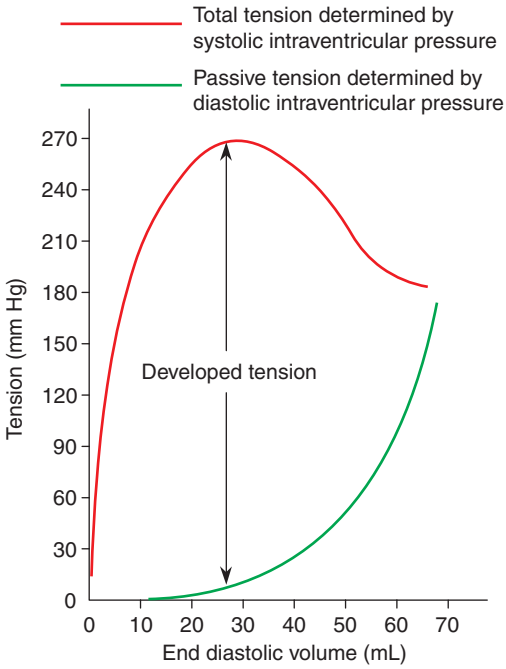


Fig. 4.1-14 Length–tension relationship in cardiac muscle.

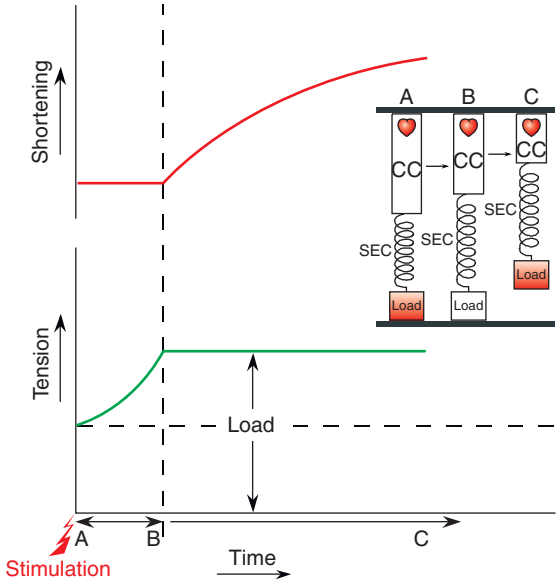


Fig. 4.1-15 Model of contraction of cardiac muscle in after-loaded condition: A, resting phase; B, isometric contraction phase and C, isotonic contraction phase.

that large increase in volume can be accommodated with only small increase in pressure.

Systolic ventricular pressure represents the active tension developed (isometric tension) which is proportionate to the degree of diastolic filling of the heart (initial length of muscle fibres). The graph (Fig. 4.1-14) shows that the developed tension increases as the diastolic volume increases until it reaches a maximum (ascending limb of Starling curve), then tends to decrease (descending limb of Starling's curve). The descending limb is instead due to the beginning of disruption of the myocardial fibres.

Clinical significance of the Frank–Starling law of heart and its role in control of cardiac output is discussed in Chapter 4.3 on 'Heart as a Pump' (see page 218).

5. Effect of afterload

Afterload refers to the load which acts on the muscle after the beginning of muscular contraction. The afterload affecting the force of contraction of cardiac muscle is represented by the resistance against which the ventricles pump the blood. The afterload (resistance) for right ventricle is low in pulmonary artery due to its intrathoracic location. The afterload (resistance) for the left ventricle is high in the aorta due to resistance to blood flow through the aortic valves and systemic blood vessels called *peripheral resistance.*

Cardiac muscle contraction with afterload. Figure 4.1-15A is the model for contraction of afterloaded cardiac muscle. It shows two phases:

Isometric contraction phase. In this phase, the muscle contracts but there occurs no shortening of the muscle. The contraction of contractile component (CC) of the muscle stretches the series elastic component (SEC) and there occurs a rise in tension (Fig. 4.1-15B). The rise in tension continues till it equals the load. This is the end point of isometric contraction.

Isotonic contraction phase starts when the muscle tension exceeds the load and it starts moving. There occurs shortening of the muscle without further stretching of SEC (Fig. 4.1–15C). The performance of contractile component is given by the force–velocity relationship.

Force–velocity relationship

The force–velocity curve is plotted by noting the velocity of muscle contraction with progressively increasing load on the muscle. In the heart, load is represented by the resistance against which the ventricles pump the blood and velocity of muscle contraction is represented by the stroke output. Following inferences can be drawn from the force–velocity curve (Fig. 4.1-16A).

- When the load is zero, the muscle contracts rapidly and the velocity of muscle shortening is maximum (Vmax).
- As the load increases progressively, the velocity of shortening decreases till it reaches zero. At this point, force developed is called maximum isometric force and is represented by Po. Therefore, during muscle contraction, the velocity of shortening and force developed are inversely related. The force–velocity relationship curve is influenced by change in initial length of the muscle and the effect of catecholamines.
  - *Effects of change in initial length on force–velocity relationship curve.* An increase in change in the initial length

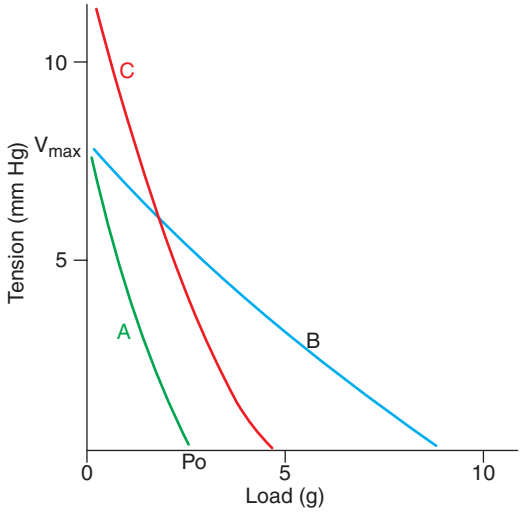


Fig. 4.1-16 Force–velocity curve in cardiac muscle (A), effect of change of initial length (B) and effect of catecholamines or increased Ca^{2+} concentration in ECF (C) on it.

(within physiological limits) increases the force of contraction (Po) without changing the velocity (Vmax), i.e. the relationship shifts to the right (Fig. 4.1-16B).

– *Effect of catecholamines or increased calcium concentration in ECF.* The catecholamines or increased Ca^{2+} concentration in ECF, both cause an increase in Po as well as Vmax (Fig. 4.1-16C).

Significance of force–velocity relationship. The cardiac muscle can alter its work and power (rate of working) at any given load and muscle length by nature of its changing force–velocity relationship in different conditions.

- When the pressure against which the heart is pumping the blood is raised, the heart strokes out less blood than it receives for several beats. Consequently, blood accumulates in the ventricle increasing the end-diastolic volume which increases the initial length of muscle fibres (i.e. size of the heart). The distended heart beats more forcefully and output returns to its previous level, i.e. by an increase in the initial length of cardiac muscle, the Po is increased and Vmax is achieved.
- Conversely, when the pressure against which the heart is pumping the blood is reduced, the stroke output rises transiently but the size of heart decreases and the stroke output falls to the previously constant level.

Chapter 4.2

Origin and Spread of Cardiac Impulse and Electrocardiography

ORIGIN AND SPREAD OF CARDIAC IMPULSE

- **Introduction**
- **Anatomic consideration**
  - Conducting system of the heart
  - Characteristic histological features of conducting system
  - Innervational characteristics of the heart
- **Mechanism of origin of rhythmic cardiac impulse**
  - Pacemaker
  - Electrical potential in pacemaker tissue
  - Role of autonomic nervous system in controlling heart rhythm
- **Spread of cardiac impulse**

ELECTROCARDIOGRAPHY

- **Introduction**
- **Recording of ECG**
  - ECG leads
  - Electrocardiograph
  - Normal electrocardiogram
  - Waves of ECG
  - Intervals and segments of ECG
  - Characteristic features of ECG complex in unipolar chest leads and limb leads
- **Vectorial analysis of electrocardiogram and vector cardiography**
  - Concept of cardiac vectors
  - Mean electrical axis
  - Vector cardiography
- **His bundle electrogram**
- **Clinical applications of ECG**
  - Cardiac arrhythmias
  - Myocardial infarction
  - Ventricular hypertrophy
  - Effect of changes in the ionic composition of blood on electrical activity of heart

ORIGIN AND SPREAD OF CARDIAC IMPULSE

INTRODUCTION

The cardiac muscle possesses special properties which include autorhythmicity, conductivity, excitability and contractility.

Autorhythmicity, refers to the property of the cardiac muscle which enables the heart to initiate its own impulse at constant rhythmical intervals. Because of this property, the heart continues to beat even after all nerves to it are sectioned. This is because of the presence of a specialized *pacemaker tissue* in the heart that can initiate repetitive action potentials. The pacemaker tissue makes a *conduction system* that normally spreads impulses through the heart.

ANATOMIC CONSIDERATION

CONDUCTING SYSTEM OF THE HEART

The conducting system of the heart consists of specialized fibres of the heart muscle present as the sinoatrial node, the interatrial tract, the internodal tracts, the atrioventricular (AV) node, the AV bundle of His and its right and left terminal branches, and the subendocardial plexuses of the Purkinje fibres (Fig. 4.2-1).

1. Sinoatrial node. Sinoatrial (SA) node is located in the wall of right atrium, just right to the opening of superior vena cava. Its dimensions are about 15 mm length, 2 mm width and 1 mm thickness. Spontaneous rhythmical electrical impulses arise from the SA node and spread in all directions to:

- Cardiac muscles of atria,
- Interatrial tract to left atrium and
- Internodal tracts to AV node.

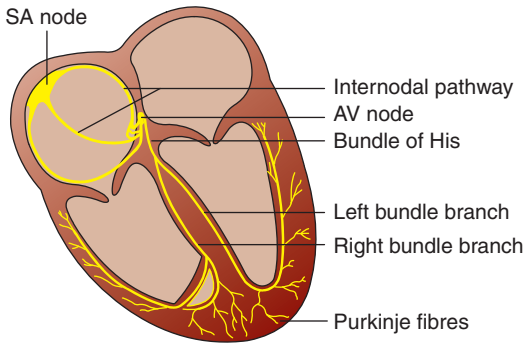


Fig. 4.2-1 Specialized conducting tissues of the heart.

2. Interatrial tract (Bachman's bundle). It is a band of specialized muscle fibres that run from the SA node to the left atrium. It causes simultaneous depolarization of the atria.

3. Internodal conduction pathway. Three internodal conduction paths have been described (Fig. 4.2-2):

- *Anterior internodal pathway of Bachman* leaves the anterior end of the SA node and passes anterior to the superior vena cava opening. It then descends on the atrial septum and ends in the AV node.
- *Middle internodal pathway of Wenckebach* leaves the posterior end of the SA node and passes posterior to the superior vena cava opening. It then descends on the atrial septum to the end in the AV node.
- *Posterior internodal pathway of Thorel* leaves the posterior part of the SA node and descends through the crista terminalis and the valve of inferior vena cava to the AV node.

4. Atrioventricular node. The AV node is located just beneath the endocardium on the right side of lower part of the atrial septum, near the tricuspid valve. It is stimulated by the excitation wave that travels through the internodal tracts and the atrial myocardium. From it, the cardiac impulse is conducted to the ventricles by the AV bundle.

5. Atrioventricular bundle of His. The AV bundle arises from the AV node, descends through the fibrous skeleton of the heart and divides into right bundle branch for the right ventricle and the left bundle branch for the left ventricle. The branches break up and become continuous with the plexus of Purkinje fibres.

6. Purkinje fibres. These are spread out deep to the endocardium and reach all parts of the ventricles including the bases of papillary muscles.

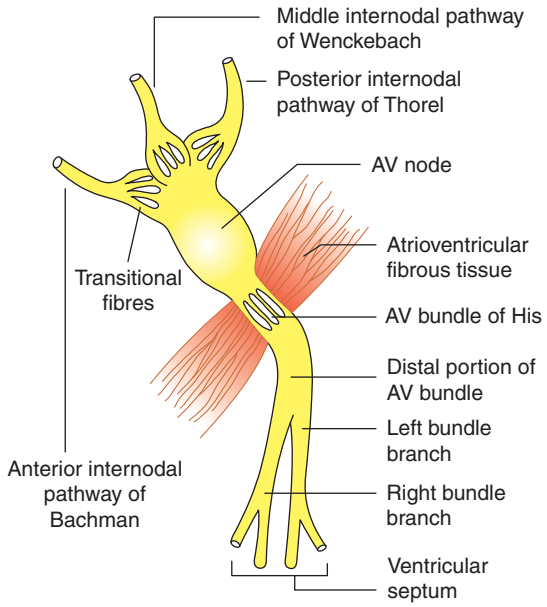


Fig. 4.2-2 Internodal conduction pathways.

CHARACTERISTIC HISTOLOGICAL FEATURES OF CONDUCTING SYSTEM

- The conduction system of the heart is composed of modified cardiac muscle that has *fewer striations* and *indistinct boundaries.*
- The SA node and, to a lesser extent, the AV node, also contain *small round cells* with few organelles, which are connected by *gap junctions.* These are probably the actual pacemaker cells, and therefore they are called P cells.

INNERVATIONAL CHARACTERISTICS OF THE HEART

- Both SA node and AV node are richly supplied by the sympathetic as well as the parasympathetic nerves. Parasympathetic fibres come from the vagus nerve and most sympathetic fibres come from the stellate ganglion.
- The SA node is supplied by the right vagus nerve and right-sided sympathetic nerves because it develops from the structures on the right side of embryo.
- The AV node is supplied by the left vagus and left-sided sympathetic nerves because it develops from the structures on the left side of embryo.

IMPORTANT NOTE

- Noradrenergic fibres are epicardial, whereas the vagal fibres are endocardial.
- Connections exist for the reciprocal inhibitory effects of sympathetic and parasympathetic innervation of the heart on each. Thus, acetylcholine acts presynaptically to reduce norepinephrine release from the sympathetic nerves, and conversely, neuropeptide Y released from the noradrenergic endings may inhibit the release of acetylcholine.

MECHANISM OF ORIGIN OF RHYTHMIC CARDIAC IMPULSE

PACEMAKER

The part of the heart from which rhythmic impulses for heart beat are produced is called *pacemaker.* In mammalian heart, though the other parts of the heart like AV node, atria and ventricle can also produce the impulse but *SA node acts as a pacemaker* because the rate of impulse generation by the SA node is highest. However, when there occurs blockage of transmission of impulse from the SA node to the AV node, the pacemaker activity may shift from the SA node to other sites, e.g. the AV node. When pacemaker is other than the SA node, it is called as *ectopic pacemaker.* Ectopic pacemaker causes abnormal sequence of contraction of different parts of the heart.

Rate of production of rhythmic impulses by different parts of the heart is:

- SA node: 70–80/min,
- AV node: 40–60/min,

- Atrial muscle: 40–60/min and
- Ventricular muscles: 20–40/min.

ELECTRICAL POTENTIAL IN PACEMAKER TISSUE

The electrical potential in a cardiac muscle (contractile myocardial cells—CMC) has been described on page 178. As shown in Fig. 4.1-7 in phase 4 of action potential of cardiac muscle (CMC), there exist a constant resting membrane potential of –85 to –90 mV. In pacemaker (SA node) fibres, however, the resting membrane potential is only of –55 to –60 mV; and that this is not steady, i.e. it shows a slow rise in the resting membrane potential due to slow depolarization (Fig. 4.2-3). Due to this slow depolarization, the threshold level –40 mV is reached very slowly. Once the threshold level of –40 mV is reached, there occurs a rapid depolarization up to +5 mV followed by a rapid repolarization, i.e. there occurs action potential and generation of an impulse. After a rapid repolarization (phase 3 of action potential), once again the resting membrane potential (phase 4 of action potential) is reached, which is not stable and starts rising slowly to again reach at threshold level to produce the second impulse. This slow rising resting membrane potential in between the action potentials is called *prepotential* or *pacemaker potential* (Fig. 4.2-3). Presence of this unique feature in the cells of pacemaker tissue is the underlying mechanism responsible for self-generation of rhythmic impulses (autorhythmicity).

Ionic basis of pacemaker potential and action potential in SA node

The myocardial cells present in the SA node and the AV node are called *slow fibres* and the other myocardial cells are called *fast fibres* depending on the membrane potential and the speed of conduction velocity of the action potential (Fig. 4.2-4).

- The slow fibres of the pacemaker tissue have a unique feature, i.e. leakage of resting membrane for sodium. This occurs due to the activation of *'h'* channels, (also called as 'f' channels) as a result of unusual or funny activation following hyperpolarization of the membrane (while the resting membrane of fast fibres is relatively impermeable to Na^+). This causes *slow diffusion of Na^+ into the SA nodal fibres under resting condition. This slow entry of Na^+ in the cells slowly raises the potential to –55 mV* (i.e. causes slow depolarization). This slow depolarization forms the initial part of pacemaker potential (Fig. 4.2-4).
- Then the 'T (transient) calcium channels' open up and there is slow influx of Ca^{2+} causing further depolarization in the same at slower rate till a threshold level of –40 mV is reached. Thus, calcium current (Ica) due to opening of 'T calcium channels' forms the later part of the pacemaker potential. In addition, local release of Ca^{2+} from the sarcoplasmic reticulum (Ca^{2+} sparks) as reported to occur during prepotential.
- At the threshold level (–40 mV) the 'L (long lasting) calcium channels' open up and the action potential starts with a rapid depolarization due to influx of Ca^{2+}. Thus, it is important to note that the depolarization in SA node is mainly due to influx of Ca^{2+} rather than Na^+. Consequently, the depolarization is not as sharp as in the other myocardial fibres.
- At the end of depolarization, potassium channels open up and calcium channels close. This causes K^+ to diffuse out of the fibres resulting in a rapid repolarization from –55 to –60 mV.
- Again, due to an unique feature of slow fibres of the SA node (i.e. *leakage of resting membrane to Na^+*), *the resting*

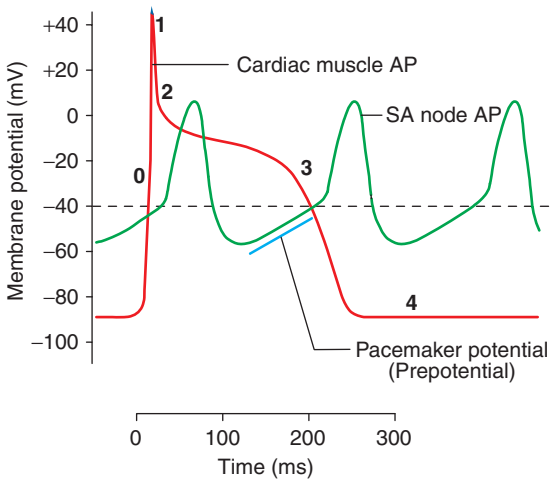


Fig. 4.2-3 Phases of action potential in a cardiac muscle fibre (0, 1, 2, 3, 4) and sinoatrial node showing pacemaker potential. (AP = Action potential.)

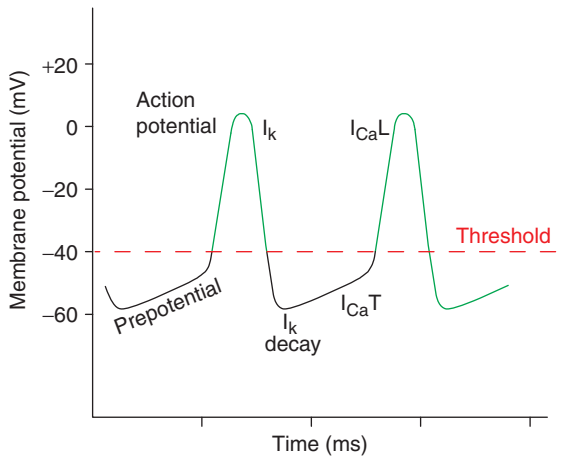


Fig. 4.2-4 Pacemaker potential and its ionic basis. $I_{Ca}T$: Ca^{2+} conductance through transient channels; $I_{Ca}L$: Ca^{2+} conductance through long lasting channels; I_k: potassium conductance.

potential does not become stable but slow depolarization starts due to slow influx of Na^+ making initial part of pre-potential. And ultimately, due to repetition of the above described steps another action potential is initiated. In this way, impulses are generated at regular intervals of time *(autorhythmicity).*

ROLE OF AUTONOMIC NERVOUS SYSTEM IN CONTROLLING HEART RHYTHM

Vagal tone. SA node is richly innervated by the parasympathetic fibres from the right vagus. Normal activity of vagus liberates acetylcholine from its nerve endings, which increases the permeability of the SA nodal fibres for potassium producing hyperpolarization (due to rapid efflux of K^+). This hyperpolarization slows the firing rate of the SA node from its automatic rate of 90–120 impulses/min to the actual heart rate of about 72 beats/min. The normal vagal activity is called vagal tone.

Effect of parasympathetic stimulation. Parasympathetic stimulation causes release of acetylcholine at the vagal nerve endings and by the above described mechanism causes (Fig. 4.2-5A):

- Decrease in the heart rate by decrease in the rate of sinus rhythm and
- Decrease in the rate of transmission of impulses to ventricles due to decreased excitation of the conducting system. Strong parasympathetic stimulation may even completely block the transmission and ventricles may stop beating for 4–10 s. If it happens, the Purkinje system initiates the rhythm causing the ventricular contraction at a rate of 15–40/min. This phenomenon is called *vagal escape.*

Effect of sympathetic stimulation. Stimulation of the sympathetic nerves causes release of norepinephrine at the nerve endings. Probably, this increases permeability of cardiac muscle fibres to calcium by opening up L calcium channels. This increases the rate of sinus rhythm and rate of conduction of impulse as well as excitability in all the portions of heart. Force of contraction of atria and ventricles also increases greatly (Fig. 4.2-5B).

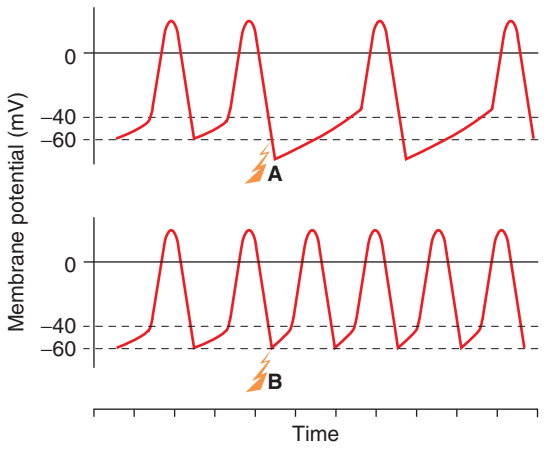


Fig. 4.2-5 Effect of parasympathetic A, and sympathetic B, stimulation on pacemaker potential.

SPREAD OF CARDIAC IMPULSE

The cardiac impulse which originates in the SA node in the form of action potential spreads throughout the heart through the conduction system (properties of which are summarized in Table 4.2-1) in the sequence given:

SA node and atria. The impulse travels over the muscle fibres of atria from the SA nodal fibres and through the interatrial tract to the left atrium. Conduction through these fibres causes simultaneous depolarization of both the atria. Atrial depolarization is completed in about 0.1 s.

AV node. *Conduction through AV node* is slow, there is a delay of about 0.1 s. *The causes of AV nodal delay are:*

- Transitional fibres connecting internodal tracts and the AV node are very small and conduct the impulse at a very slow rate, i.e. 0.02–0.05 m/s. The AV nodal fibres also conduct the impulse at a very slow rate (0.02–0.05 m/s), and.
- There are very few gap junctions connecting successive fibres in the pathway.

The ability of the AV node to slow and to block the rapid impulse is called *detrimental contraction.* This *AV nodal delay is useful,* for it provides time for completion of the atrial contraction and their emptying, (i.e. ventricular filling) before the ventricles contract.

Ventricular conduction. The impulses conducted through the AV node are distributed to ventricles through bundle of His,

Table 4.2-1 Properties of the conduction system

Tissue	Fibres diameter (μm)	Resting membrane potential (mV)	Conduction velocity (m/s)
SA node	—	–40–50	0.05
Atrial muscle	8–10	–70–80	0.3–0.5
Interatrial and internodal tract	15–20	–80–90	1.0
AV node	Variable	–50	0.02–0.05
Purkinje fibres	70–80	–70	2.0–4.0
Ventricular muscle	10–16	–80	<1.0

its branches and Purkinje fibres in 0.08–0.15 s. In humans, depolarization of the ventricular muscle proceeds as (Fig. 4.2-6):

- Starts at the left side of the interventricular septum.
- Moves first to the right across the middle portion of the septum.
- The wave of depolarization then spreads down the septum to apex of the heart.
- It, then, returns along the ventricular walls to the AV groove, proceeding from the endocardial to the epicardial surface.
- The last parts of the heart to be depolarized are the posterobasal portion of the left ventricles, the pulmonary conus and the upper most portion of the septum.

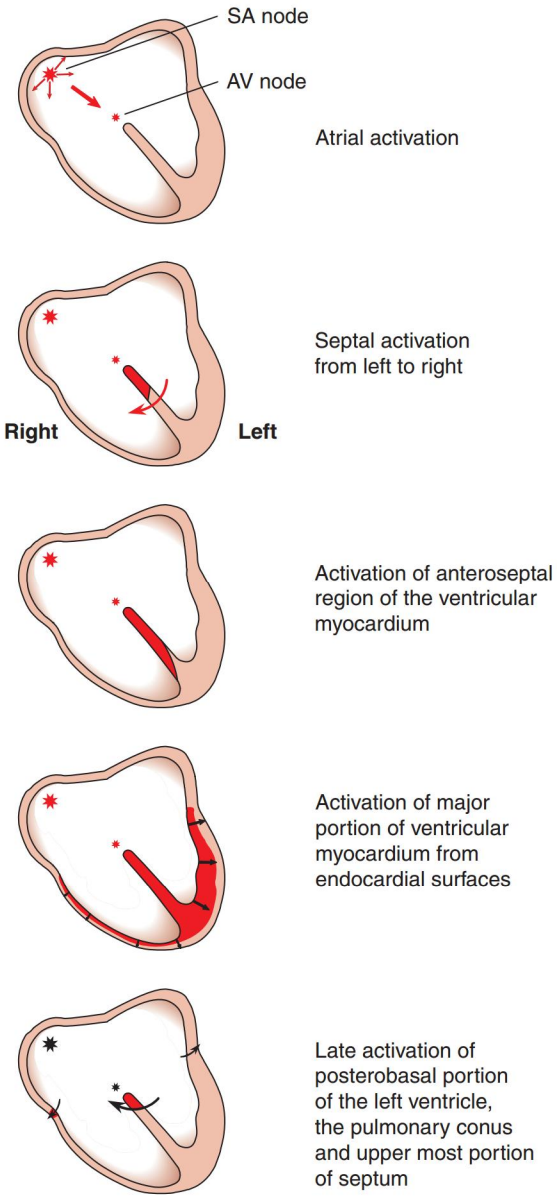
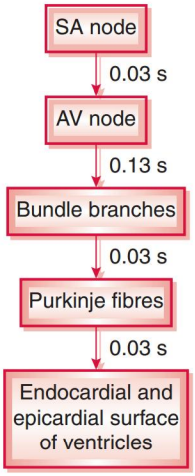


Fig. 4.2-6 Spread of cardiac impulse.

Time taken for impulse to travel through different tissue is depicted below:



Thus, total time required for conduction from the SA node to the endocardial surface is 0.22 s.

IMPORTANT NOTE

The AV bundle is the only connecting tissue between atria and ventricles. Therefore, if there is destruction of the AV node, then atria and ventricles beat independent of one another, i.e. atria at the rate of 72 beats/min and ventricles at a slower rate (30–40 beats/min). There is a complete dissociation of atrial and ventricular beating and is called *Idioventricular rhythm*.

ELECTROCARDIOGRAPHY

INTRODUCTION

Willem Einthoven, a Dutch physiologist, originally developed the technique of electrocardiography. He was awarded Nobel prize in 1924 for his contribution and is called the father of modern electrocardiography.

Electrocardiography refers to the extracellular recording of the summed-up electrical events of all the cardiac muscle fibres generated with each heart beat. Electrically, heart behaves as a *dipole,* i.e. two terminal battery in which the excited part (depolarized segment) forms a negative pole and the non-excited part forms the positive pole (Fig. 4.2-7) lying in the volume conductor (body fluids are good conductor of electricity). Therefore, electrical changes occurring in the heart during each beat are picked up by the surface electrodes. Thus, electrocardiogram (ECG) is the surface recording of the potential difference between the two poles of the heart dipole at a given time. The record of the potential fluctuations during the cardiac cycle is called the *electrocardiogram*. The machine used to record these potential fluctuations is called *electrocardiograph,* which is essentially a sensitive galvanometer.

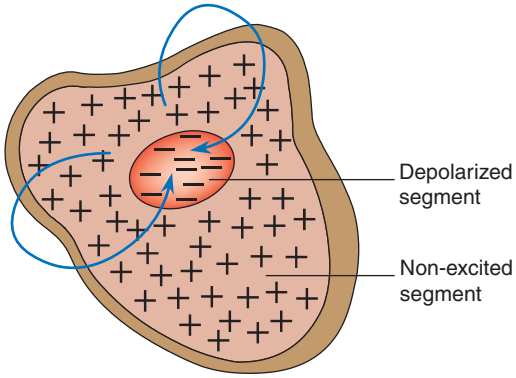


Fig. 4.2-7 The heart as a dipole.

RECORDING OF ECG

ECG LEADS

ECG leads refer to the two electrodes which are placed on the body surface and connected to ECG machine for measuring the potential fluctuations between only two points. ECG is recorded using two types of lead systems, the bipolar leads and unipolar leads.

Bipolar leads

In bipolar recording *both the electrodes are active* and one of the active electrode is connected to negative terminal of the ECG machine and the other to the positive terminal. Three standard limb leads used in the bipolar recording are based on Einthoven's assumption that the body is like an electrically homogeneous plate in which the right and left shoulders and the pubic region form the corners of an equilateral triangle with heart in its centre (Einthoven's triangle) and that two active electrodes need to be placed at two corners of this triangle (Fig. 4.2-8). However, for convenience, the electrodes are connected to the left arm (LA), right arm (RA) and left foot (LF) instead of the shoulders and the pubic region (Fig. 4.2-8). Practically, it does not make any difference whether the electrodes are placed in proximal or distal part of the extremities because the current flows in the body fluids and so the records obtained are similar. In three standard limb leads, the two active electrodes are connected as (Fig. 4.2-8):

Lead I (LI). In lead I, the two active electrodes are connected to LA and RA.

- Left arm (LA) electrode is connected to positive terminal, and
- Right arm (RA) electrode is connected to negative terminal of the ECG machine.

Lead II (LII). In lead II, the electrodes are connected to RA and LF.

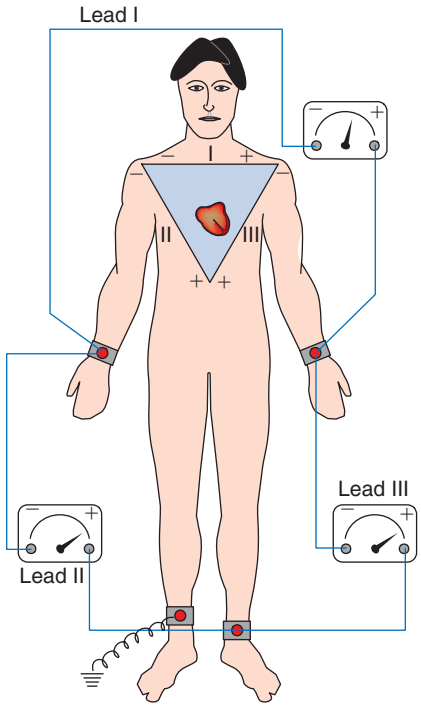


Fig. 4.2-8 Einthoven's triangle and position of electrodes for standard limb leads (I, II and III).

- Left leg (LF) electrode is connected to *positive* terminal, and
- Right arm (RA) electrode is connected to *negative* terminal of the ECG machine.

Lead III (LIII). In lead III, the electrodes are connected to LA and LF.

- Left leg (LF) electrode is connected to positive terminal, and
- Left arm (LA) electrode is connected to negative terminal of the ECG machine.

Unipolar leads

In unipolar recording, one electrode is an *active* or the *exploring* electrode and the other is an *indifferent electrode* at zero potential. Since the potential at the indifferent electrode remains zero, so in unipolar recording the records obtained represent the potential fluctuations occurring at the site of exploring electrode.

In a volume conductor, the sum of potentials at the points of an equilateral triangle with a current source at the centre is zero at all times. Therefore, if the three electrodes (placed on left arm, right arm and on the left leg) are connected to a common terminal, through a resistance, an indifferent electrode that stays near zero potential is obtained. In clinical electrocardiography, two types of unipolar leads are used.

Unipolar chest leads. There are six unipolar chest leads (precordial leads) designated V_1–V_6. The indifferent electrode is obtained as described above and the active electrode is placed on six points on the chest as (Fig. 4.2-9):

- *Lead V_1*: In the right fourth intercostal space, just near the sternum.
- *Lead V_2*: In the left fourth intercostal space, just near the sternum.
- *Lead V_3*: Halfway between V_2 and V_4.
- *Lead V_4*: In the left fifth intercostal space at mid-clavicular line.
- *Lead V_5*: In the left fifth intercostal space at anterior axillary line.
- *Lead V_6*: In the left fifth intercostal space at mid-axillary line.

Unipolar limb leads. These include lead VL, VF and VR. In unipolar limb leads one *exploring* (active) *electrode* is placed over a limb (In lead VL over the left arm, in VF over the left foot and in lead VR over the right arm) and is connected to the positive terminal of the electrocardiograph. The *indifferent electrode* is obtained as described above and is connected to the negative terminal of the electrocardiograph. These leads are not used and have been replaced by the augmented limb leads.

Augmented unipolar limb leads. Generally, augmented unipolar limb leads designated as aVR, aVL and aVF are used. In augmented leads, the size of potential is increased by 50% without any change in the configuration from the non-augmented record. The *active electrode* is from one of the limbs and the indifferent electrode is obtained by connecting the other two limbs through 5000 Ω resistance as:

- Lead aVR : Active electrode is from RA and indifferent electrode is from LA + LF.
- Lead aVL : Active electrode is from LA and indifferent electrode is from RA + LF.
- Lead aVF : Active electrode is from LF and indifferent electrode is from RA + LA.

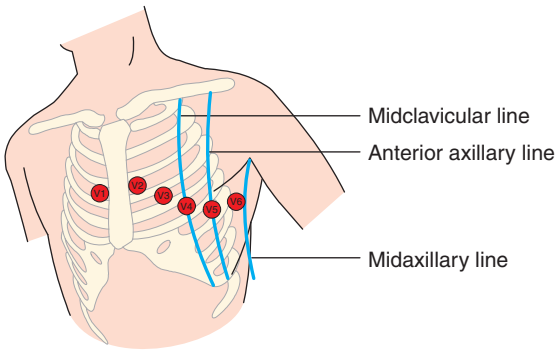


Fig. 4.2-9 Position of electrodes for chest leads (V_1–V_6).

ELECTROCARDIOGRAPH

The electrocardiograph (ECG machine) is essentially a sophisticated string galvanometer. A modern electrocardiograph amplifies and records the potential fluctuations on a moving strip of paper. Special paper is used which turns black on exposure to heat. The stylus (recording pen) is made hot by the electrical current flowing through its tip.

Calibration of time and voltage on ECG paper

- The special ECG paper having 1 mm and 5 mm squares (Fig. 4.2-10) is used. The tracing is usually made at a standard recording speed of 25 mm/s.
- On horizontal axis, therefore, each millimetre represents 0.04 s (1/25).
- The sensitivity of electrocardiograph is adjusted in such a way that a potential fluctuation of 1 mV causes a vertical deflection of 1 cm. Thus, on vertical axis, each millimetre represents 0.1 mV magnitude of potential.

NORMAL ELECTROCARDIOGRAM

Electrocardiogram refers to the record of the potential fluctuations during the cardiac cycle. As a result of sequential spread of the excitation in the atria, the interventricular septum and the ventricular walls (Fig. 4.2-6) and finally repolarization of the myocardium, a series of positive and negative waves designated as P, Q, R, S and T are recorded during each cardiac cycle. Depolarization moving towards an active electrode in a volume conductor produces a positive deflection, whereas depolarization moving in the opposite direction produces a negative deflection. Therefore, the shape and polarity of P, Q, R, S and T waves will vary in different

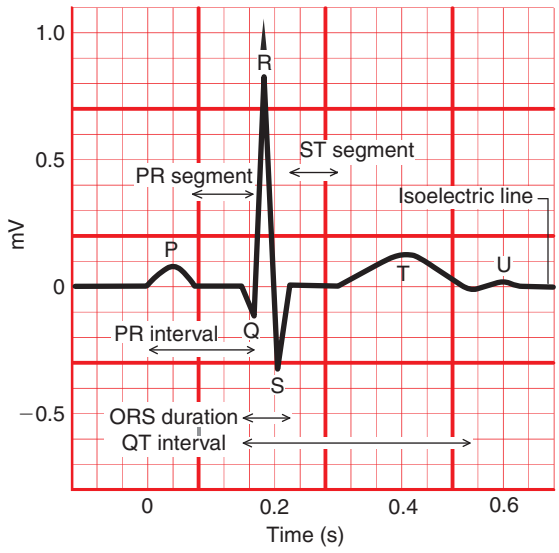


Fig. 4.2-10 Calibration of time and voltage (amplitude) on special ECG paper.

leads due to differences in the orientation of each lead with respect to the heart (Fig. 4.2-11). Configuration of a typical electrocardiogram from a bipolar limb lead II (L II) is described below (Fig. 4.2-10):

WAVES OF ECG

P wave

Configuration. P wave is the positive (upright rounded) deflection.

Cause. It is produced by the depolarization of the atrial musculature so also called atrial complex.

Duration of P wave is not more than 0.1 s.

Amplitude of P wave is from 0.1 to 0.12 mV.

Clinical significance. Magnitude of P wave is a guide to the functional activity of atria.

- In mitral stenosis, the left atrium is hypertrophied and P wave becomes larger and prolonged.
- In tricuspid stenosis, the right atrium is hypertrophied and P wave becomes tall (0.5 mV) and peaked with normal duration.

QRS complex

Configuration. QRS complex consists of three consecutive waves. Q wave is a small negative wave which may be absent normally (quite often). It is continued as a tall positive R wave which is followed by a small negative S wave.

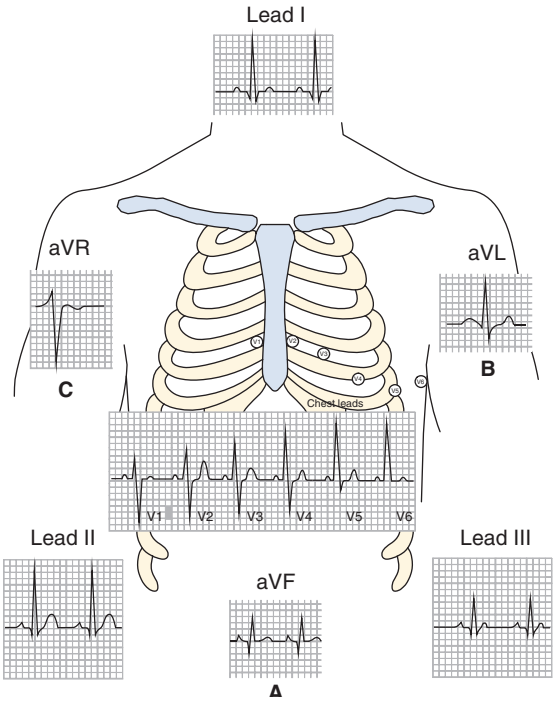


Fig. 4.2-11 The electrocardiographic complexes recorded from different leads.

Cause. The QRS complex is caused by a ventricular depolarization.

Duration of QRS complex is normally less than 0.08 s. It is a measure of an intraventricular conduction time.

Amplitude of Q wave is 0.1–0.2 mV, R wave is 1.0 mV and S wave is 0.4 mV (Total 1.5–1.6 mV).

Clinical significance. QRS complex from the precordial leads are more important than the limb leads.

- Deep Q wave (more than 0.2 mV) along with other changes is an important sign of myocardial infarction.
- Tall R wave (more than 1.3 mV) is seen in ventricular hypertrophy.
- Low voltage QRS complex (total sum less than 1.5 mV) is seen in hypothyroidism and pericardial effusion.
- QRS complex is prolonged in bundle branch block.

T wave

Configuration. T wave is the last, positive, dome-shaped deflection. Normally, it is in the same direction as the QRS complex, because the ventricular repolarization follows a path opposite to depolarization.

Cause. T wave represents ventricular repolarization.

Duration of T wave is approximately 0.27 s.

Amplitude of T wave is about 0.3 mV.

Clinical significance

- Inverted T wave is an important sign of myocardial ischaemia or infarction.
- Tall and peaked T wave occurs in hyperkalaemia.

U wave

Configuration. It is a small round positive wave.

Cause. It occurs due to slow repolarization of papillary muscle.

Duration of U wave when present is 0.08 s.

Amplitude of U wave is about 0.2 mV.

Significance. It is rarely seen normally. It becomes prominent in hypokalaemia.

Note. Since atrial repolarization coincides with ventricular depolarization, so it is merged with the QRS complex and thus not recorded as a separate wave.

INTERVALS AND SEGMENTS OF ECG

P–R interval

It is measured from the onset of P wave to the onset of the QRS complex. Actually, it is PQ interval but Q wave is frequently absent therefore it is called P–R interval.

- It measures the *AV conduction time,* including the AV nodal delay.
- Its *duration* varies from 0.12 to 0.21 s depending on the heart rate.
- *Clinical significance.* Prolonged P–R interval indicates AV conduction block.

J point

J point refers to the point on ECG which coincides with the end of depolarization and start of repolarization of ventricles, i.e. it occurs at the end of the QRS complex. At this point, since all parts of the ventricles are depolarized so no current is flowing around the heart.

QT interval

It is the time from the start of the QRS complex to the end of T wave.

- It indicates total *systolic time of ventricles,* i.e. ventricular depolarization and repolarization.
- *Duration* of QT interval is about 0.4 s (QRS duration and ST segment duration).
- *Clinical significance.* Ischaemia and any ventricular conduction defects prolong the QT interval. In hypocalcaemia also QT interval is prolonged.

TP interval

It is measured from the end of T wave to the beginning of P wave.

- It measures the *diastolic period* of the heart.
- Variable TP interval indicates AV dissociation.

P–P interval

P–P interval is the interval between two successive P waves. Equal P–P intervals indicate rhythmic depolarization of the atria.

ST segment

It is an isoelectric period between the end of the QRS complex and the beginning of T wave.

- Its *duration* is about 0.04–0.08 s.
- It corresponds with the *ventricular repolarization.*
- *Clinical significance.* ST segment is elevated in the patients with myocardial infarction.

ST interval

It is the time from the end of S wave to the end of T wave.

- Normal duration of ST interval is 0.32 s.
- It represents the ventricular repolarization.

CHARACTERISTIC FEATURES OF ECG COMPLEX IN UNIPOLAR CHEST LEADS AND LIMB LEADS

Chest leads

The ECG complex produced in unipolar chest leads (V_1–V_6) represents the electrical activity of the part of the heart which lies nearest to the active electrode (Fig. 4.2-12).

P wave is positive in all the leads because the excitation wave moves from the SA node to the AV node, i.e. posterior to anterior.

QRS complex. It represents the electrical activity of ventricles and so its configuration changes in different leads are as below:

- In V_1 and V_2 (which reflect right ventricular activity), the main QRS complex is *negative.*
- In V_3 and V_4 (which reflect activity of both ventricles including interventricular septum), the main QRS complex is *biphasic.*
- In V_5 and V_6 (which reflect left ventricular activity), mainly the main QRS complex is *positive.* Thus as shown in Fig. 4.2-12.
- *R wave* gradually increases in size from V_1 to V_6 leads. In leads V_1, R wave represents the activity of right ventricle and in V_6 of left ventricle.
- *S wave* gradually decreases in size from lead V_1 to V_6. In lead V_1, S wave represents activity of left ventricle and in lead V_6 of right ventricle.

Limb leads

Lead VF and aVF. These leads reflect the electrical activity of inferior surface of the heart which is formed by parts of both right and left ventricles and interventricular septum. Therefore, the QRS complex in these leads like that of V_3 and V_4 is predominantly *biphasic* (Fig. 4.2-11A).

Lead VL and aVL. These leads reflect the electrical activity of left outer side of the heart, which is mainly formed by the left ventricle. Therefore, the QRS complex in these leads like that of V_6 is predominantly *positive* (Fig. 4.2.11B).

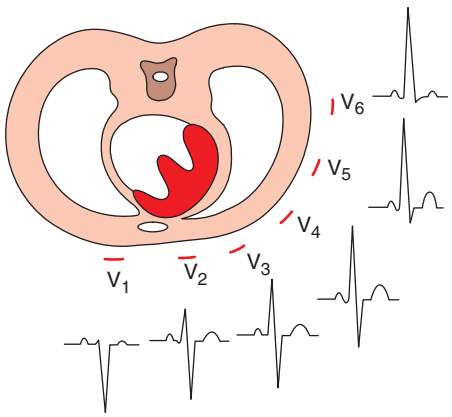


Fig. 4.2-12 Pattern of QRS complex in chest leads (V_1–V_6).

Lead VR and aVR. These leads reflect the activity of the cavity of the ventricles, irrespective of the position of heart. Therefore, *P wave, QRS complex and T wave* all are *negative deflection* (Fig. 4.2-11C).

IMPORTANT NOTE

- In common practice it is necessary to record 12 leads ECG because limb leads i.e. I, II, III, aVL and aVF represent depolarization of heart in the vertical frontal plane and chest leads represent depolarization of heart in horizontal plane.
- Standard limb lead II is commonly used for cardiac monitoring because the position of electrodes in this lead resembles the pathway of current flow.

VECTORIAL ANALYSIS OF ELECTROCARDIOGRAM AND VECTOR CARDIOGRAPHY

In the discussion until this point, we have studied the configuration of various positive and negative waves of the ECG complex and their clinical significance. In addition to the information gained from the changes in the configuration of various waves, the vectorial analysis of the electrocardiogram also provides many useful information pertaining to cardiac abnormalities. The concept of cardiac vector, methods of vector analysis and their clinical significance is discussed.

CONCEPT OF CARDIAC VECTORS

During cardiac cycle (depolarization and repolarization of heart), current flows in the heart at every instant. The magnitude and direction of the potential generated can be represented in the form of an arrow which is called a vector. By convention, the arrowhead points towards the direction and the length of the arrow is drawn proportional to the voltage of the potential.

During most of the cycle of ventricular depolarization, direction of electrical potential (negative to positive) is from the base of ventricles towards the apex. This preponderant direction of potential during depolarization is called the *mean QRS vector* (mean electrical axis (MEA) of the heart) and is drawn through the centre of the ventricles in a direction from the base of heart towards the apex (Fig. 4.2-13).

The *instant vector,* however, represents the magnitude and direction of potential at a particular instant during the cardiac cycle. The instantaneous vectors of five different instants during the process of ventricular depolarization are shown in Fig. 4.2-14.

CALCULATING THE MEAN ELECTRICAL AXIS FROM STANDARD LEAD ELECTROCARDIOGRAM

As discussed earlier, MEA refers to the *mean vector* produced during a cardiac cycle (i.e. by P, QRS and TP waves of ECG).

In the frontal plane, MEA can be calculated from any two standard (i.e. bipolar) limb leads (using Einthoven's triangle or triaxial reference system) or any two augmented limb leads (using hexa-axial reference system).

The MEA in the horizontal plane is derived using the precordial (chest) leads.

Triaxial reference system involves moving the sides of Einthoven's triangle so that they intersect at the centre of triangle. Lead I then divides the system into an upper (negative) and a lower (positive) hemisphere. The triaxial system of representation reveals that (Fig. 4.2-15B):

- The axis of lead I is 0° because the electrodes lie in the horizontal direction with a positive electrode to the left.
- The axis of lead II is about 60° as the electrodes are placed on the right arm (–ve) and left leg (+ve) and
- The axis of lead III is about 120° as the electrodes are placed on the left arm (–ve) and left leg (+ve).

The hexa-axial reference system involves superimposition of the axis of augmented limb leads on the triaxial system. The hexa-axial system of vector representation reveals that (Fig. 4.2-15C):

- The axis of lead aVF is 90°,
- The axis of lead aVR is 210° and
- The axis of lead aVL is –30°.

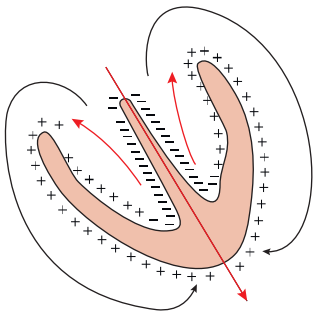


Fig. 4.2-13 Instantaneous mean vector during ventricular depolarization.

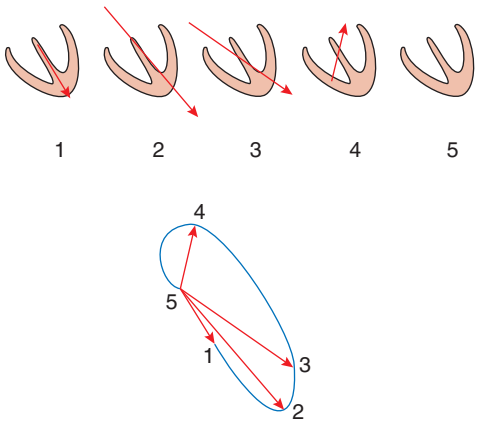


Fig. 4.2-14 The instantaneous mean vector of five different instants during process of ventricular depolarization and construction of QRS vector cardiogram.

Clinically, the frontal MEA of the QRS complex is more useful and is determined from the two standard limb leads as:

- The Q and S waves (negative values) are added algebraically to the R wave (positive value) for each of two leads.

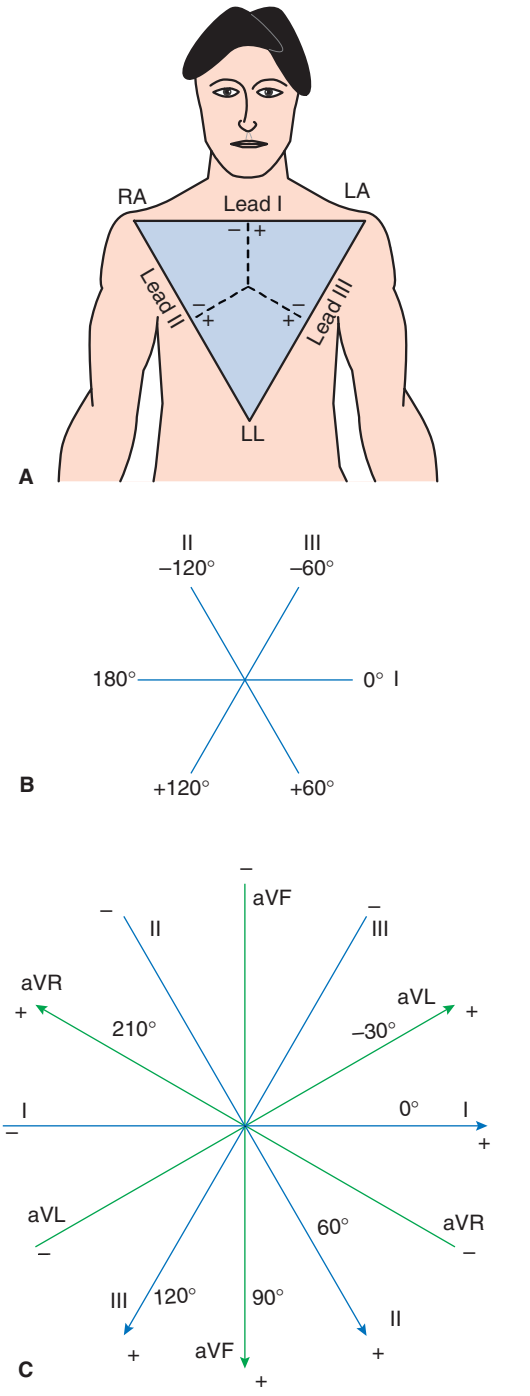


Fig. 4.2-15 Schematic drawing for calculating the mean electrical axis (MEA): A, Einthoven's triangle and connections of bipolar limb leads; B, the triaxial reference system in which lead I, II and III collapsed into their respective zero points; and C, the hexa-axial reference system obtained by adding augmented unipolar limb leads (aVR, aVL and aVF) to triaxial system.

The result gives the magnitude and the direction of the QRS vector (+or –) in each lead. For example:

In lead I, the algebraic sum of Q (–3), R (+13), S (–5) = –3 + 13 –5 = 5 mV and

In lead II, the algebraic sum of the Q (–1), R (+15), S (–0) = –1 + 15 – 0 = +14 mV.

The QRS magnitude is plotted along the respective leads from the zero point towards the appropriate polarity with the arrow pointing towards the correct pole (positive or negative using the Einthoven's triangle (Fig. 4.2-16) or the triaxial reference system).

- Perpendiculars are then drawn from the leads of the two vectors until they intersect.
- The intersection marks the head of the mean electrical vector. The MEA is then drawn from the centre of the triangle (electrical zero for the system) to the perpendiculars' intersection. The length of this arrow represents the magnitude of the MEA and its direction (in degrees represents the electrical axis in the frontal plane.

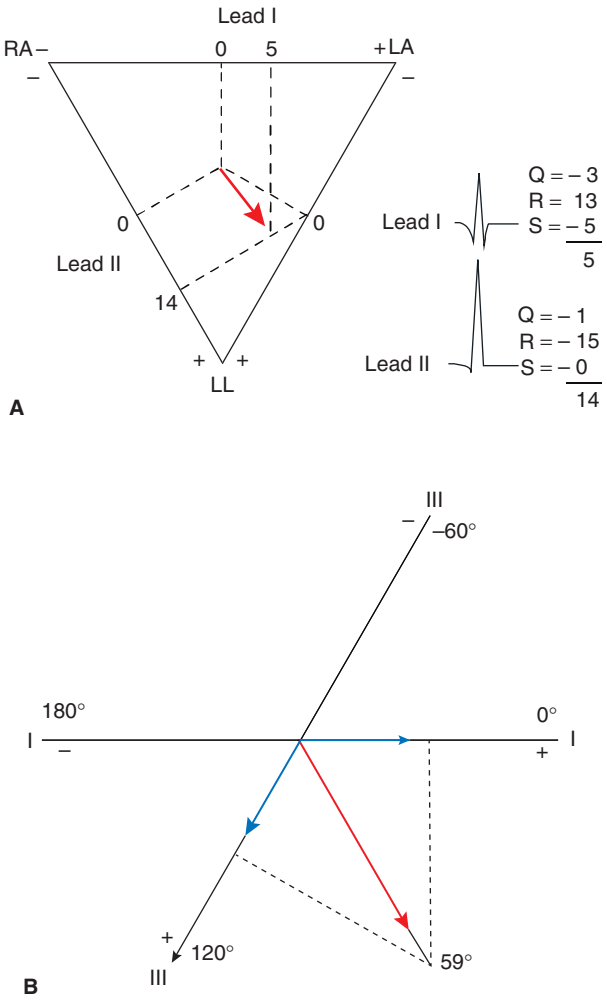


Fig. 4.2-16 Mean electrical axis (MEA) determined by vector analysis, A and normal MEA (in degrees) determined by hexa-axial system, B.

- The MEA of normal ventricles lies between –30° and +120° when plotted on the hexa-axial reference system or (2 and 7 O'clock, respectively, if the hexa-axial system is considered a clock face (Fig. 4.2-17A)). Usually, the normal MEA is 59°, i.e. 5 O'clock positive (Fig. 4.2-17B).

Abnormalities of MEA

1. Right axis deviation (RAD) is said to be present when the MEA lies between +120° and +180° (7 and 9 O'clock position). Figure 4.2-17 shows the QRS complex that occurs in RAD.

Causes of RAD include:

- Right ventricular hypertrophy secondary to chronic lung disease or pulmonary valve stenosis.

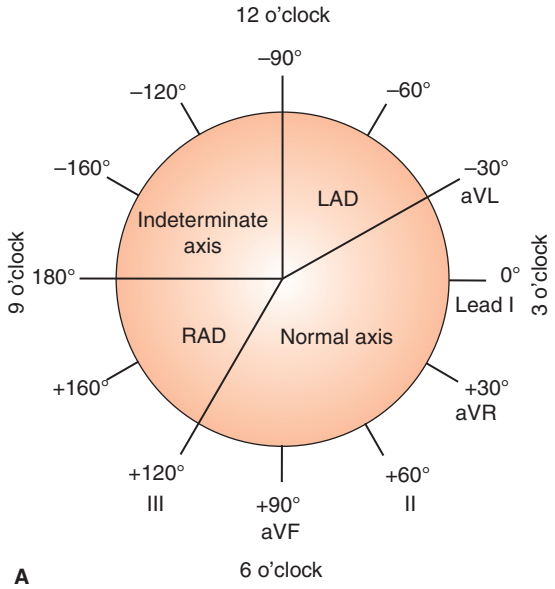


Fig. 4.2-17 A, Mean electrical axis in the frontal plane (using hexa-axial system) showing normal axis, right axis deviation (RAD) and left axis deviation (LAD); and B, QRS complexes (normal, in RAD and in LAD) seen in leads I, II and III.

- Right bundle branch block causing delayed activation of the right ventricle and
- Posterior (or inferior) myocardial infarction (MI).

2. Left axis deviation (LAD) is said to be present when the MEA lies between –30° and –90° (or 2 and 12 O'clock). Figure 4.2-17 shows typical QRS complex that occur in LAD.

Causes of LAD are:

- Left ventricular hypertrophy,
- Obesity,
- Left bundle branch block and
- Anterolateral myocardial infarction.

VECTOR CARDIOGRAPHY

As discussed earlier, the vector of current flow through the heart changes rapidly as the impulse spreads through the myocardium. The vector changes into two aspects:

- Increase or decrease in length corresponding to change in voltage (magnitude) and
- Changes its direction because of changes in the average direction of the electrical potential of the heart.

Figures 4.2-14 and 4.2-18 show the instant vectors of eight different successive instants during the process of ventricular depolarization (i.e. in a normal QRS complex). On joining the positive ends of the vectors a loop is obtained which is called *vector cardiogram*.

A continuous record of all the vectors is made using an *oscilloscope*. The procedure is called vector cardiography and the record obtained in the form of a loop called vector cardiogram (similar to P, QRS and T that described above). Three loops can be recorded during one cardiac cycle (Fig. 4.2-19). An atrial depolarization cannot be recorded with standard technique of vector cardiography because of the prolonged time course and small voltages involved.

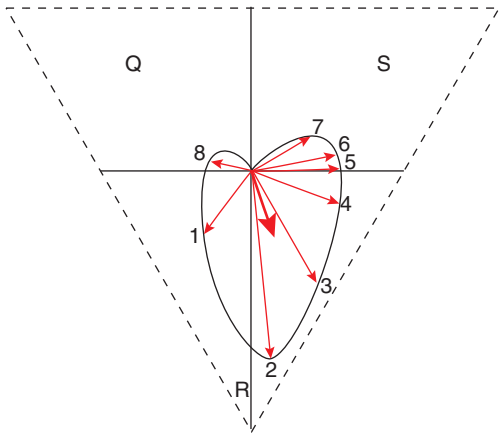


Fig. 4.2-18 Instant vector of eight arbitrary stages of ventricular depolarization and reconstruction of QRS complex from QRS loop for three bipolar limb leads. Arrows indicate the direction of loop recording.

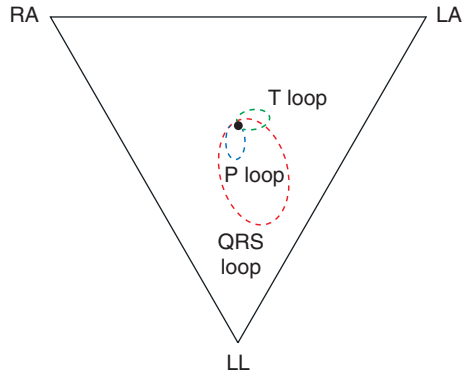


Fig. 4.2-19 The vector cardiographic loops P, QRS and T.

1. P loop. It is caused by atrial depolarization. It is small and is directed leftward and inferiorly resulting in a positive P wave in three bipolar limb leads.

2. QRS loop. It is caused by ventricular depolarization. The normal QRS loop is inscribed counterclockwise and is directed leftward, inferior and posterior. Figure 4.2-18 shows how the QRS loop generates the QRS complex in three limb leads.

3. T loop. It results from ventricular repolarization, which is roughly opposite in direction to the depolarization. This reversal of direction results in T wave that normally is in the same direction as the QRS complex.

HIS BUNDLE ELECTROGRAM

His bundle electrogram (HBE) refers to the recording the electrical activity of the heart obtained through the intracardiac ring electrodes placed near the tricuspid valve.

It is accomplished with a catheter containing ring electrodes at its tip that is passed through a vein to the right side of heart and manipulated into a position close to the tricuspid valve. Three or more standard electrocardiographic leads are recorded simultaneously.

Normal HBE shows following deflections (Fig. 4.2-20):

- *A deflection,* which corresponds to the activation of AV node,
- *H spike* is due to transmission of impulse through the His bundle and
- *V deflection* is produced during the ventricular depolarization.

Uses of HBE

It is specially useful in patients with heart blocks. From the HBE and ECG from standard leads, it is possible to accurately time the following three intervals:

PA interval. It is the time from the first appearance of atrial depolarization to the A wave in the HBE. It represents

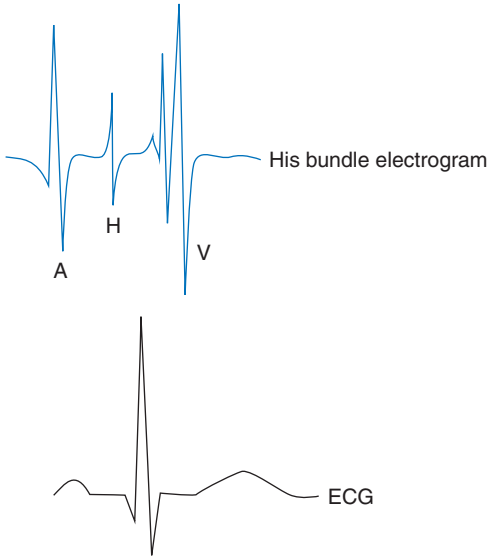


Fig. 4.2-20 Normal His bundle electrogram (HBE) with simultaneously recorded ECG.

conduction time from the SA node to the AV node. Normal value of PA interval is 27 ms.

AH interval. It is the time from the A wave to the start of H spike in the HBE. It represents the AV nodal conduction time. Normal value of AH interval is 92 ms; the higher value of AH interval shows relative slowness of conduction in the AV node.

HV interval. It is the time from the start of the H spike in the HBE to the start of the QRS complex deflection in the ECG. It represents conduction in the bundle of His and the bundle branches. Normal value of HV interval is 43 ms.

CLINICAL APPLICATIONS OF ELECTROCARDIOGRAPHY

Electrocardiography is an indispensable tool in the diagnosis, prognosis and planning treatment in most of the cardiac disorders. The important applied aspects which need special mention are:

- Cardiac arrhythmias,
- Myocardial infarction,
- Hypertrophy of various cardiac chambers and
- Effects on ECG of changes in the ionic composition of blood.

CARDIAC ARRHYTHMIAS

Cardiac arrhythmias refer to the disruption of the normal cardiac rhythm. The normal cardiac rhythm implies a regular sinus rhythm with a normal cardiac rate, between 60 and 100 beats/min (average 72 beats/min). *Sinus rhythm* is said to be present when the SA node is pacemaker, and each P wave is followed by a normal QRS complex, the P–R and

QT intervals are normal, and R–R interval is regular. Cardiac arrhythmias may be discussed as:

- Abnormal sinus rhythm.
- Conduction disturbances (heart blocks).
- Ectopic cardiac rhythm.

Abnormal sinus rhythm

Sinus arrhythmia

- Sinus arrhythmia (Fig. 4.2-21B) is characterized by a normal sinus rhythm except for the R–R interval (cardiac rate) which varies in a set pattern. Usually, heart rate increases during inspiration and decreases during expiration, as a result of variations in vagal tone that affect the SA node.
- Sinus arrhythmia is common in children and in endurance athletes with slow heart rates.

Sinus tachycardia

- Sinus tachycardia (Fig. 4.2-21C) is characterized by a normal sinus rhythm except for the increased heart rate (i.e. decreased but regular R–R interval). Tachycardia is labelled when heart rate is more than 100 beats/min.
- Sinus tachycardia is a normal response to exercise and is also associated with:
  - Fever
  - Hyperthyroidism
  - As a reflex response to low arterial pressure

Sinus bradycardia

- Sinus bradycardia (Fig. 4.2-21D) is characterized by a normal sinus rhythm except for the decreased heart rate (i.e. increased but regular R–R interval). Bradycardia is labelled when heart rate becomes less than 60 beats/min.

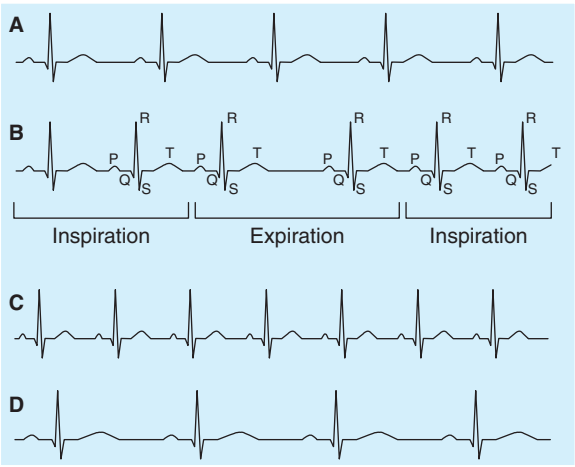


Fig. 4.2-21 Electrocardiogram tracings showing: A, normal sinus rhythm; B, sinus arrhythmia; C, sinus tachycardia and D, sinus bradycardia.

- Sinus bradycardia is more commonly seen in highly trained endurance athletes, sometimes it may be abnormal.

Sick sinus syndrome. Sick sinus syndrome refers to a condition characterized by marked bradycardia accompanied by dizziness and syncope.

Causes of sick sinus syndrome include:

- Sinus bradycardia that does not improve with sympathetic stimulation or vagal inhibition.
- SA nodal block (see below) and
- Sinus arrest, i.e. complete stoppage of sinus discharge.

Treatment, when the condition causes severe symptoms, consists of implantation of an artificial pacemaker. Sinus node dysfunction accounts for over half of the pacemaker implants.

Conduction disturbances (Heart blocks)

Heart blocks refer to the slowing down or blockage of cardiac impulse (generated from SA node) along the cardiac conductive pathway. Conduction blockage may occur as:

- SA nodal block
- AV nodal block
- Bundle branch block

SA nodal block

- SA nodal block or the so-called sinoatrial block is characterized by blockage of impulse conduction from SA node to atria.
- Sinoatrial block may manifest as *sick sinus syndrome.*
- AV nodal rhythm also called *junctional rhythm* is characterized by an inverted P wave and normal QRS complex, and the rate is slower than the sinus rhythm.

AV nodal block

AV nodal blockage may occur as an *incomplete heart* block (which includes first degree and second degree heart blocks) or complete heart block (third degree heart block).

First degree AV nodal block. First degree AV nodal (or heart) block is characterized by the slowing of conduction at the level of AV node. Though all the atrial impulses reach the ventricles but the PR interval is abnormally long, i.e. more than 0.21 s (Fig. 4.2-22B).

Second degree AV nodal block. In second degree AV nodal block (Fig. 4.2-22C and D), not all atrial impulses are conducted to ventricles. It is usually associated with organic heart diseases. Consequently, there may be one ventricular contraction after every 2, 3 or 4 atrial contractions producing the so-called 2:1, 3:1 or 4:1 block (*constant block).* Other forms of second degree heart blocks are:

- *Wenckebach phenomenon (Mobitz type I block).* It is characterized by a progressive lengthening of the P–R

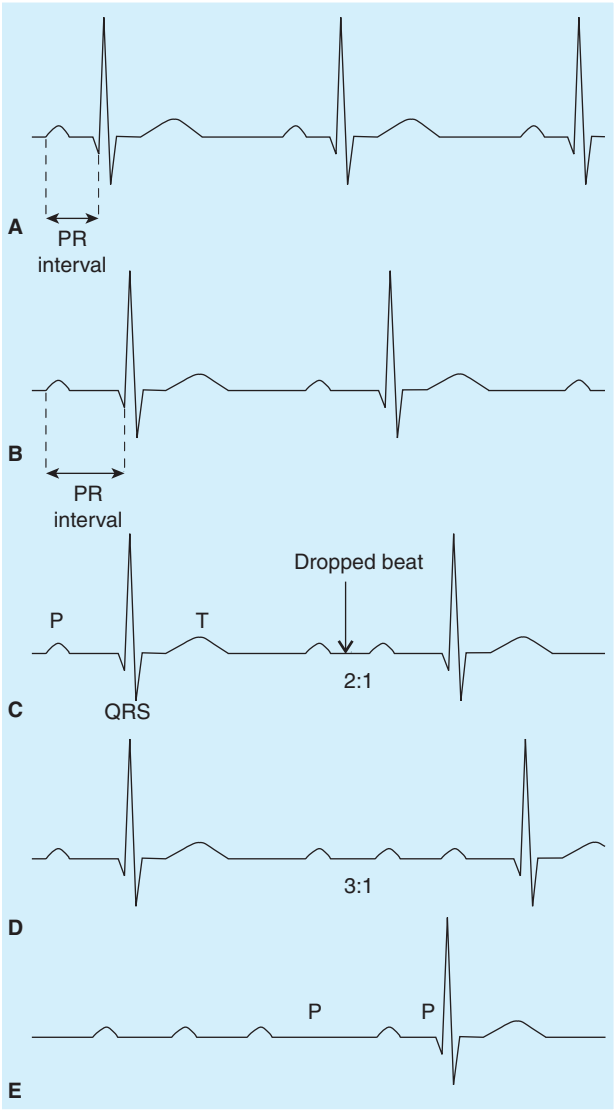


Fig. 4.2-22 Various types of AV nodal blocks: A, normal sinus rhythm; B, first degree AV block; C, second degree AV block (2:1); D, 3:1 block and E, complete AV block (third degree).

interval in successive beats and finally a failure of one impulse to be transmitted.

- *Periodic block (Mobitz type II).* It is characterized by an occasional failure of conduction that results in an atrial to ventricular rate of for example 6:5 or 8:7. The P–R interval is constant.

Third degree (complete) AV nodal block

- In third degree, (complete AV block) no impulse from atria can pass to the ventricles.
- Therefore, ventricles start beating at their own rhythm (about 40 beats/min) called idioventricular rhythm.
- The atria, however, continue to beat at the normal sinus rhythm of about 72 beats/min. Thus, ECG shows that there is complete dissociation between P waves and QRS complexes called *atrioventricular dissociation* (Fig. 4.2-22E).

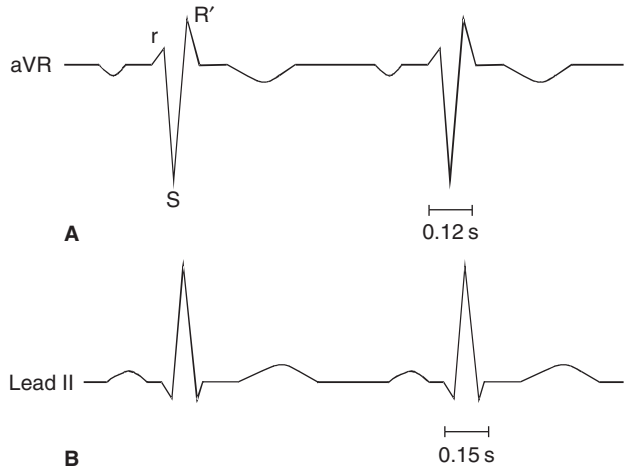


Fig. 4.2-23 Electrocardiogram characteristics in bundle branch blocks: A, right bundle branch block and B, left bundle branch block.

Bundle branch block

- Bundle branch block refers to the conduction blocks in one or more branches of the bundle of His.
- In this condition, excitation passes normally down the bundle on the intact side and then sweeps back through the muscle to activate the ventricle on the blocked side.
- Therefore, the ventricular rate is normal, but the QRS complexes are prolonged (beyond 0.12 s) and deformed Fig. 4.2-23. The characteristic features of the branch involved are:
  - *Right bundle branch block* may occur in otherwise healthy individuals or secondary to chronic pulmonary disease. The activation of right ventricle is delayed and ECG may show features of RAD (Fig. 4.2-23A).
  - *Left bundle branch block* is usually associated with the organic heart disease. It is best diagnosed using left precordial leads (Fig. 4.2-23B).

Note. His bundle electrogram is useful for detailed analysis of the site of block when there is a defect in the conduction system.

Ectopic cardiac rhythm

Ectopic cardiac rhythm refers to the abnormal cardiac excitation produced either by an ectopic focus or a re-entry phenomenon.

MECHANISMS OF DEVELOPMENT OF CARDIAC ARRHYTHMIAS

Cardiac arrhythmias may result from the *ectopic foci of excitation* and/or re-entry mechanism.

1. Ectopic foci of excitation

Under normal circumstances, SA node acts as a pacemaker, however, in certain abnormal conditions, the bundle of His

or Purkinje fibres or the myocardial fibres become hyperexcitable and discharge spontaneously. The site in the heart which becomes hyperexcitable is called an *ectopic focus* which may behave as:

Single discharge. When the irritable ectopic focus discharges once, an *extra systole* or *premature* beat is caused before the next normal beat. Depending upon the site of ectopic focus the premature beat may be atrial, nodal or ventricular.

Repetitive discharge. If the ectopic focus discharges impulses repeatedly at a rate higher than that of the SA node, the tachycardia with very high rate (tachyarrhythmias) result.

2. Re-entry mechanism

Re-entry mechanism or the circus movement refers to a phenomenon in which the wave of excitation propagates repeatedly (continuously) within a closed circuit. It is a more common cause of tachyarrhythmias. Re-entry of excitation wave is known to occur under two situations: (i) in the presence of transient block in the conduction pathway and (ii) in the presence of an abnormal extra bundle of conducting tissue called *bundle of Kent.*

Re-entry due to transient block in the conduction system. Normally, during depolarization of a ring of cardiac tissue, the impulse spreads in both directions of the ring (Fig. 4.2-24A) and the tissue behind each branch of the impulse is refractory and thus the impulse cannot go down the other side.

When there is a transient block on one side, the impulse can go down the other side of ring (Fig. 4.2-24B), because this portion is not depolarized and so not refractory.

If the transient block is worn off, the impulse from retrograde direction is conducted through this (previously blocked) area and then continues to circle indefinitely. This phenomenon is called circus movement or re-entry phenomenon (Fig. 4.2-24C).

The site of re-entry keeps on producing impulses continuously. If the re-entry is in AV node, the re-entrant activity depolarizes the atrium and the resulting atrial beat is called an *echo beat.* In addition, the re-entrant activity in the node propagates back down to the ventricles producing paroxysmal nodal tachycardia. The re-entrant activity can also become established in the atrial muscle fibres (producing atrial tachycardias, flutter or fibrillation) and in the ventricular muscle fibres (producing ventricular tachycardia or ventricular fibrillation).

Re-entrant activity in the presence of bundle of Kent. Bundle of Kent is an abnormal extra bundle of the conducting tissue present in some individuals. This bundle connects the atria and ventricles directly, so the conduction is very rapid than through the regular conductive system.

If a transient block develops in the normal conductive system, the impulse from the SA node reaches the ventricle through the bundle of Kent and produces excitation. If the blockage in the normal conduction system worns off then the excitation wave from the ventricle travels in the opposite direction and re-enter the AV node and a circus movement is established (Fig. 4.2-25).

This re-entrant activity produces an echo beat in atria and nodal paroxysmal tachycardia or the so-called supraventricular tachycardia.

The nodal paroxysmal tachycardia occurring in the patients with bundle of Kent is called Wolf–Parkinson–White syndrome. Producing short PR interval, prolonged slurred QRS deflection but normal PJ interval (start of P wave to end of QRS complex).

SALIENT FEATURES OF CARDIAC ARRHYTHMIAS

Extra systole

Extra systole (premature beat or premature contraction or ectopic beat) refers to the contraction of the heart prior to the time that normal contraction would have been expected.

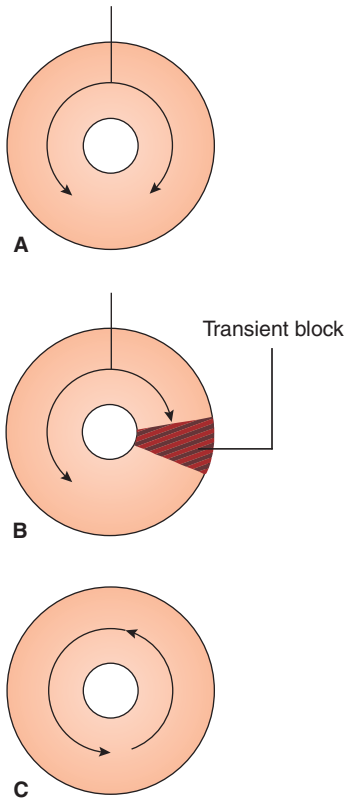


Fig. 4.2-24 Re-entry phenomenon or circus movements, a cause of cardiac arrhythmias; A, normal depolarization of a ring cardiac tissue; B, spread of wave of excitation in presence of transient block and C, circus movement.

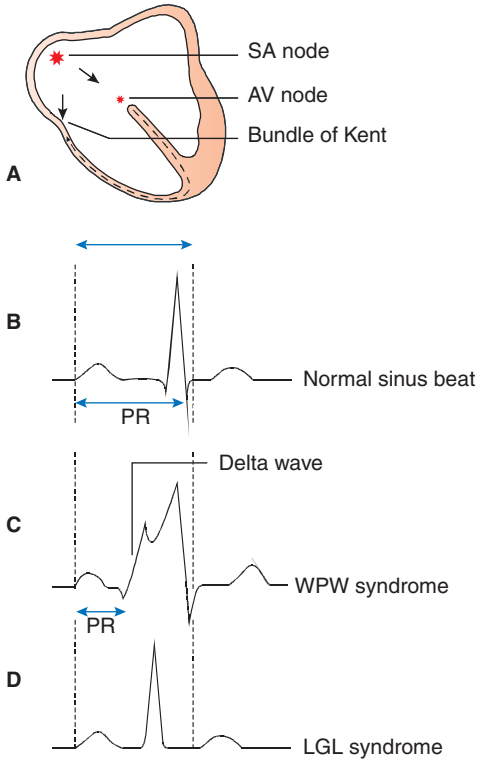


Fig. 4.2-25 Re-entry phenomenon in the presence of bundle of Kent (A) and electrocardiogram record in a patient with bundle of Kent showing short PR interval, wide and slurred QRS complex with normal PJ interval [Wolff-Parkinson-White (WPW) syndrome] (C) and Lown-Ganong-Levine (LGL) syndrome (D).

It is caused by some ectopic focus in the atria or ventricles and thus the premature beat may be atrial or ventricular.

Atrial extra systole (premature beat)

ECG appearance (Fig. 4.2-26B) of atrial premature beat is characterized by:

- A premature P wave which has an aberrant configuration and an abnormal PR interval because of the different path of atrial depolarization.
- QRS complex and T wave are normal.
- The subsequent cardiac rhythm is shifted and reset because the premature beat discharges the SA node, which then repolarizes and fires after the normal interval.

Significance. Since atrial extra systole occurs normally, the patient may or may not be aware of an occasional irregularity in the cardiac rhythm.

Ventricular extra systole

Ventricular extra systoles can arise from any portion of the ventricular myocardium.

ECG appearance of ventricular premature beat is characterized by (Fig. 4.2-26C):

- Absence of P wave preceding the QRS complex.

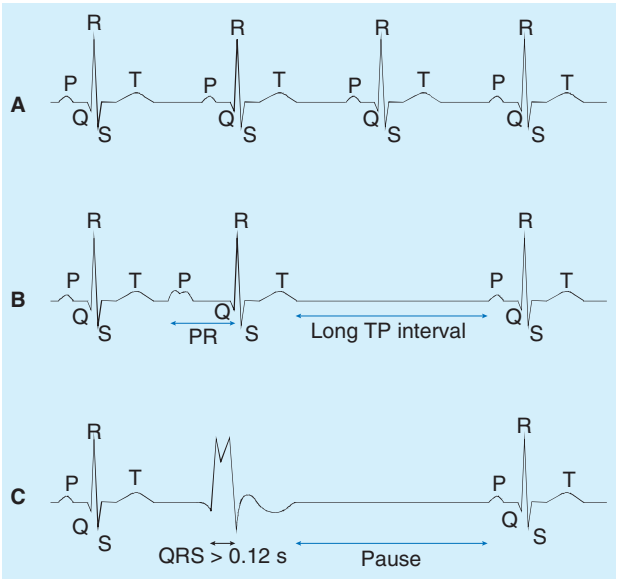


Fig. 4.2-26 Electrocadiographic record in extrasystole: A, normal ECG; B, ECG with atrial premature beat (atrial extra systole) and C, ECG with ventricular premature beat.

- QRS complex is prolonged and bizarre shaped because of the slow spread of the impulse from the ectopic focus through the ventricular muscle to the rest of the ventricle.
- T wave is usually appositely directed from the QRS complex.
- Compensatory pause is often long. Since retrograde transmission of depolarization to the atria usually does not occur with a premature ventricular beat, so the atrial rate remains unaltered. The atrial depolarization that follows the premature ventricular beat arrives while the AV node is still refractory and, therefore, it is not conducted to the ventricles, creating a pause in the ventricular rhythm.
- *The beat following ventricular premature beat* is usually stronger than the normal because of the added stroke volume and thus usually detected by the patient.
- ***Pulse deficit.*** Ventricles contract ahead of time in the atrial and ventricular premature beats. Sometimes by that time the ventricles are not filled with blood and stroke volume output during the contraction is therefore decreased or even absent. During such a contraction, pulse wave passing to the periphery may be so weak that it is not felt at the radial artery. A deficit in the number of pulses felt in the radial pulse in relation to number of contraction in the heart is called pulse deficit.

Atrial arrhythmias

Atrial tachycardia (Fig. 4.2-27A) occurs when an atrial site (outside the SA node) becomes the dominant pacemaker. It is characterized by very regular rates ranging from 140 to 220 beats/min. Atrial tachycardia may be caused by an overindulgence in caffeine, nicotine or alcohol and may also occur during anxiety attack.

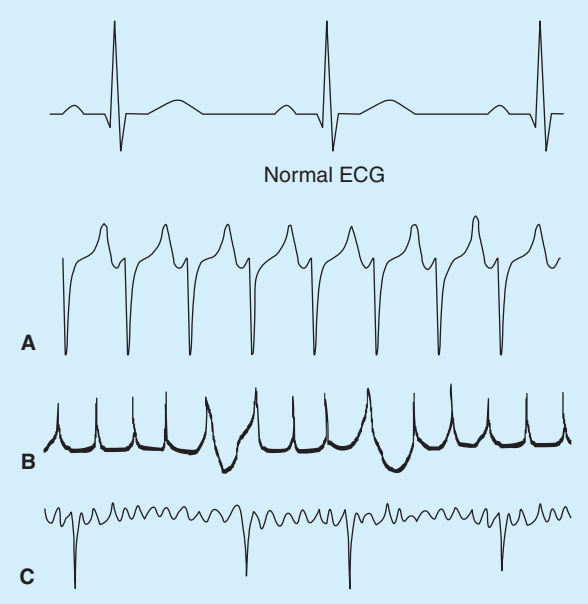


Fig. 4.2-27 Electrocardiographic record in atrial arrhythmia: A, atrial tachycardia; B, atrial flutter and C, atrial fibrillation.

Paroxysmal atrial tachycardia as the name indicates occurs in paroxysms, which usually begin suddenly and lasts for few seconds.

Atrial flutter (Fig. 4.2-27B) is said to occur with atrial rates of 220–350 beats/min. During atrial flutter, AV node is unable to transmit all of the atrial impulses and therefore the ventricular rate may be half, one-third or one-fourth of the atrial rates.

Atrial fibrillation (Fig. 4.2-27C) is characterized by a totally irregular, rapid rate (350–500 beats/min). In it, *Ventricular rate* is completely irregular because only a fraction of the atrial impulses that reach the AV node are transmitted to the ventricles.

ECG appearance is characterized by:

- Small irregular oscillations called F waves. There are no recognizable P waves.
- R–R interval is irregularly irregular.
- The QRS complex and T wave are normal because the impulses that are transmitted through the AV node are conducted normally through the ventricles.

Nodal arrhythmia

- ***Nodal paroxysmal tachycardia*** is similar to the atrial tachycardia and may be indistinguishable on ECG. They are called *supraventricular tachycardia.*
- ***Wolff–Parkinson–White syndrome*** (Fig. 4.2-25C), also known as accelerated AV conduction, refers to the occurrence of repeated attacks of nodal paroxysmal tachycardia due to the presence of bundle of Kent.

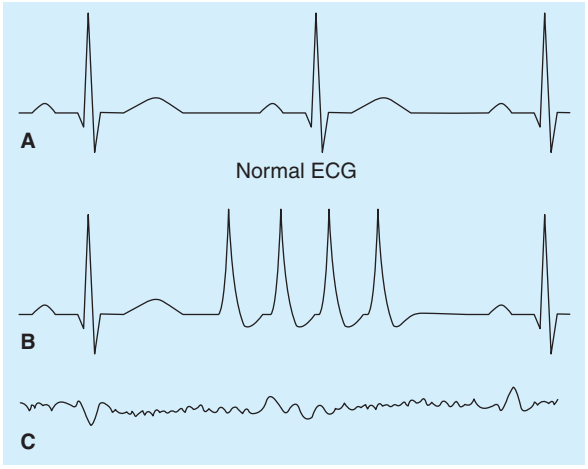


Fig. 4.2-28 Electrocardiographic record in ventricular arrhythmias: A, normal; B, paroxysmal ventricular tachycardia and C, ventricular fibrillation.

- ***Lown–Ganong–Levine syndrome*** (Fig. 4.2-25D). It is characterized by attacks of paroxysmal supraventricular tachycardia, usual nodal tachycardia in individuals with short PR intervals and normal QRS complexes. In this condition, depolarization presumably passes from the atria to the ventricles via an *aberrant bundle* that bypasses the AV node but enters the intraventricular conducting system distal to node.

Ventricular arrhythmias

Paroxysmal ventricular tachycardia occurs when a ventricular site discharges rapidly and repetitively.

Causes. Ventricular tachycardias are usually associated with serious heart disease or drug toxicity.

ECG appearance is characterized by (Fig. 4.2-28B):

- Wide, bizarre QRS complexes that occur at a rapid rate.
- P waves are usually indistinguishable.

Significance. Ventricular tachycardia is more serious because cardiac output is decreased, sustained ventricular tachycardia can be a life-threatening when it degenerates into a ventricular fibrillation.

MYOCARDIAL INFARCTION

Myocardial infarction refers to the ischaemic necrosis of a part of myocardium which occurs when the coronary blood flow ceases or is reached below a critical level. The electrocardiography is very useful for diagnosing and localising areas of myocardial infarction.

ECG appearance. The ECG undergoes a series of changes following the myocardial infarction. These changes must be

recorded daily with ECG tracing for diagnostic purpose. The *hallmark of acute myocardial infarction* is:

- *Elevation of ST segment* (Fig. 4.2-29 LI) in the leads overlying the area of infarct and
- *Depression of ST segment* (Fig. 4.2-29 LII) in the leads on the opposite side of area of infarct.

ECG appearance in old cases of myocardial infarct is characterized by:

- ST segment returns to normal,
- Appearance of Q wave (Fig. 4.2-29 LI and LII) in some of the leads in which it was not previously present and
- An increase in the size of normal Q wave in some of the other leads.

Physiological basis of ECG changes in acute myocardial infarction

The alterations in ECG pattern seen in acute myocardial infarction are attributed to *injury current,* which flows from the affected to the unaffected part of myocardium. This happens because of the fact that the affected part of myocardium gets depolarized partly or completely but does not get repolarized rapidly. The three major abnormalities that cause ECG changes (ST segment elevation) in acute myocardial infarction are:

1. Decline in resting membrane potential. The ischaemic necrosis of the myocardial fibres results in breakdown of cell membrane producing increased K^+ efflux and increase in Na^+ influx. Therefore, inside of the infarcted cells becomes less negative as compared to the unaffected area. Therefore, during the ventricular repolarization the *current flows into the infarct* from the unaffected area (Fig. 4.2-30B). This results in a depression of TQ segment of the ECG in the leads overlying the infarcted area. However, the electronic arrangement in the electrocardiographic recorders is such that TQ segment depression is recorded as ST segment elevation.

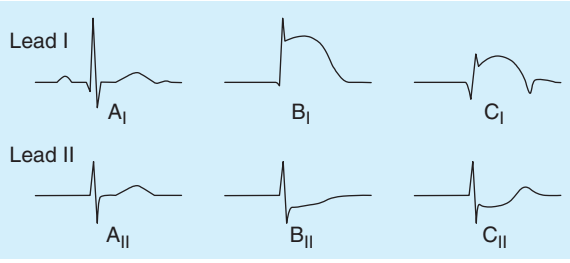


Fig. 4.2-29 Electrocardiographic record in anterior wall ischaemia in lead I and II, respectively: A_I and A_{II}, normal ECG; B_I and B_{II}, ECG within few hours of ischaemia, note ST segment elevation in lead I and depression in lead II (reciprocal) and C_I and C_{II}, ECG after several weeks, note ST segment returns to normal and Q wave appears.

2. Delayed depolarization of infarcted cells causes the infarcted area to be positive relative to the unaffected area. Therefore, current flows out of the infarcted area into the unaffected area (Fig. 4.2-30C).

3. Rapid repolarization. The repolarization in the infarcted area occurs rapidly as compared to the unaffected area due to accelerated opening of K^+ channels. Because of the rapid repolarization in the infarct, the membrane potential of the affected area becomes greater than that of the unaffected area. Extracellularly, current therefore flows out of the infarct into normal unaffected area (Fig. 4.2-30D). This current flow towards electrodes over the injured area, causing increased positivity between the S and T waves of ECG. Consequently, leads on the opposite side of the heart show ST segment depression.

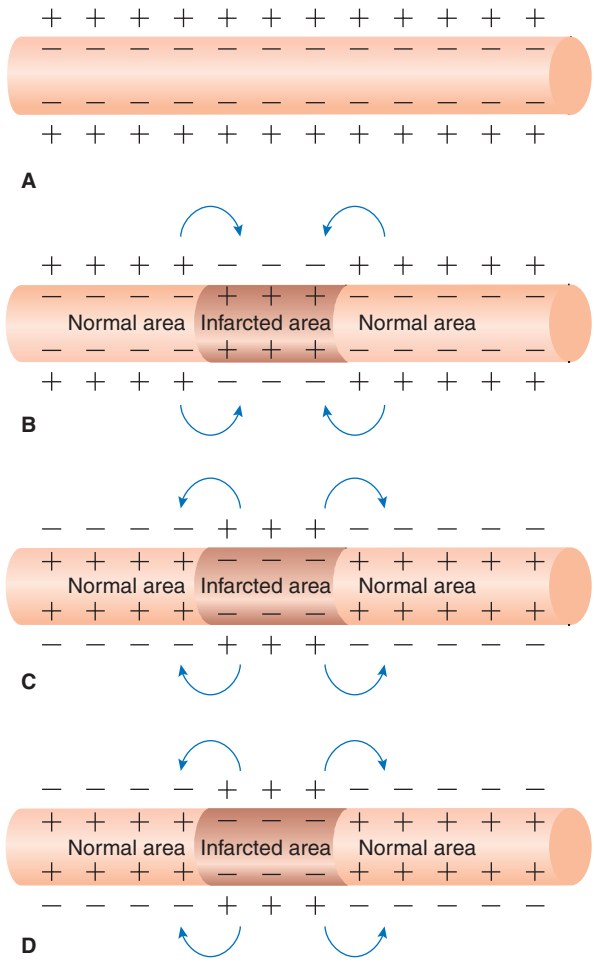


Fig. 4.2-30 Physiological basis of ECG changes (ST segment elevation) in acute myocardial infarction: A, normal resting state; B, decline in resting membrane potential in infarct area as compared to normal neighbouring (unaffected) region; C, delayed depolarization of infarct area as compared to neighbouring areas and D, rapid repolarization of infarct area in comparison to normal neighbouring areas.

Physiological basis of ECG changes in old cases of myocardial infarction

After some days or weeks of infarction, the dead myocardium and scar tissue become electrically silent. Therefore, the affected area becomes negative relative to the unaffected normal myocardium during depolarization, and it fails to contribute its share of positivity to the electrocardiographic complexes. The occurrence of Q wave in some leads (which normally lack) and deepening of Q wave in other leads is one of the manifestations of this negativity.

Localization of area of myocardial infarction

1. Anterior myocardial infarction
- Leads showing changes of MI are LI, aVL and V_3–V_5
- Leads showing reciprocal changes are LII, LIII and aVF.

2. Posterior (inferior) myocardial infarction
- Leads showing changes of MI include LII, LIII and aVF. $1–V_6$.
- Leads showing reciprocal changes are LI, aVR, aVL and V_1–V_6.

3. Lateral myocardial infarction
- Leads showing changes of MI include LI, aVL and V_6 and.
- Leads showing reciprocal change are LII, LIII, aVF and V_1.

4. Septal myocardial infarction
- Leads showing changes of MI are V_1–V_3.

VENTRICULAR HYPERTROPHY

Ventricular hypertrophy occurs when the work of the ventricle is increased sufficiently. In ventricular hypertrophy, the number of myocardial cells remains the same, but the diameter of the individual cells increases.

Left ventricular hypertrophy (LVH) occurs in the patients with systemic hypertension or aortic valve stenosis. Right ventricular hypertrophy (RVH) occurs in patients with pulmonary hypertension, pulmonary valve stenosis and some congenital heart disease.

ECG appearance is characterized by:

1. R wave. There is direct correlation between the thickness of ventricular wall and the height of R wave in the overlying leads. Therefore:
- In LVH, the R wave is tall in leads I, aVL, V_5 and V_6.
- In RVH, the R wave is tall in lead III, aVR, V_1 and V_2.

2. QRS duration is increased slightly due to the increased muscle mass. It is usually less than 0.12 s, but occasionally may exceed this value.

EFFECT OF CHANGES IN THE IONIC COMPOSITION OF BLOOD ON ELECTRICAL ACTIVITY OF HEART

The electrical activity of the heart depends upon the distribution of ions like Na^+, K^+ and Ca^{2+} in the ECF. Therefore, changes in the ECF concentration of these ions will affect the potentials of myocardial fibres and produce changes in the ECG as described.

Plasma level of sodium

Low plasma (ECF) levels of Na^+ may be associated with low-voltage ECG complex.

Plasma levels of potassium

Depending upon the levels of plasma K^+ following ECG changes are seen:

Hyperkalaemia, i.e. increase in the plasma K^+ is very dangerous and potentially lethal condition because of its effects on the heart.

Hyperkalaemia with plasma K^+ ±7.0 mEq/L, the PR and QRS intervals are within normal limits. The T wave become tall and peaked (Fig. 4.2-31), which is a manifestation of altered repolarization.

As the extracellular K^+ concentration increases, the resting membrane potential of the muscle fibres decreases.

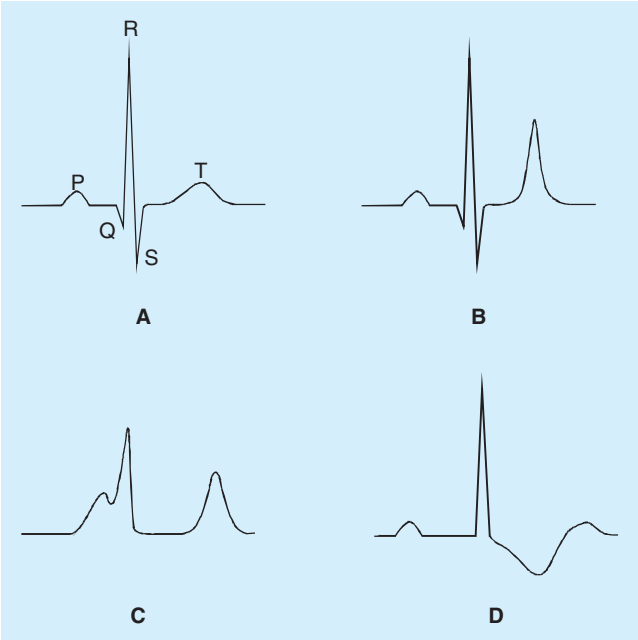


Fig. 4.2-31 Electrocardiographic changes in relation to plasma levels of potassium: A, normal tracing (plasma K^+ 4–5.5 mEq/L); B, hyperkalaemia (plasma K^+ 7.0 mEq/L); C, hyperkalaemia (plasma K^+ 8.0 mEq/L) and D, hypokalaemia (plasma K^+ 2.5–3.5 mEq/L).

Eventually, the fibres become unexcitable and the heart stops in diastole.

Hypokalaemia, i.e. decrease in the plasma levels of potassium produces following changes in ECG:

- PR interval is prolonged,
- U waves become prominent,
- ST segment is depressed and
- Late T wave inversion may occur in the precordial leads.

Plasma levels of calcium

Hypercalcaemia, i.e. increase in the extracellular Ca^{2+}, clinically is rare if ever high enough to affect the heart. However, when large amounts of calcium are infused into the experimental animals, the heart relaxes less during diastole and eventually stops in systole *(calcium rigor).*

Hypocalcaemia, i.e. decreased plasma level of Ca^{2+} produces prolongation of the ST segment and consequently, the QT interval is also increased.

Chapter 4.3

Heart as a Pump: Cardiac Cycle, Cardiac Output and Venous Return

CARDIAC CYCLE

INTRODUCTION

The heart as a pump can be considered actually comprising two separate pumps in the series: a *right heart* that pumps the blood through the lungs and a *left heart* that pumps the blood through the peripheral organs. To act as a pump, the heart contracts and relaxes rhythmically. The terms *systole* (contractile phase) and *diastole* (relaxation phase) usually refer to the ventricular events but may be prefixed by 'atrial' to refer to the atrial contraction and relaxation, respectively. The electrocardiogram records the *electrical events* that precede and initiate the corresponding *mechanical events* as:

- *P wave* is followed by the *atrial contraction,*
- *QRS waves* are caused by depolarization of the ventricles which initiates *contraction of the ventricles* and
- *T wave* occurs slightly before the end of the ventricular contraction.

The cardiac cycle, thus includes both electrical and mechanical events that occur from the beginning of one heart beat to the beginning of the next.

PHASES OF CARDIAC CYCLE

Duration of each cardiac cycle at a normal heart rate of 75 beats/min is 60/75 = 0.8 s.

During each cardiac cycle both atria contract *(atrial systole)* and relax *(atrial diastole),* and both ventricles contract *(ventricular systole)* and relax *(ventricular diastole).*

Therefore, each cardiac cycle can be considered to consist of simultaneously occurring atrial and ventricular cycles with following phases (Figs 4.3-1 and 4.3-2):

Atrial cycle

1. Atrial systole or atrial contraction phase (0.1 s) and
2. Atrial diastole (0.7 s).

Ventricular cycle

Ventricular systole (0.3 s) consisting of:

1. Isovolumic (isometric) contraction phase (0.05 s) and
2. Phase of ventricular ejection which can be further divided into rapid ejection phase (0.1 s) and slow ejection phase (0.15 s).

Ventricular diastole (0.5 s) consisting of:

1. Protodiastole (0.04 s),
2. Isovolumic (isometric) relaxation phase (0.06 s),
3. Rapid passive filling phase (0.11 s),
4. Reduced filling phase or diastasis (0.19 s) and
5. Last rapid filling phase which coincides with the atrial systole (0.1 s).

ATRIAL CYCLE

ATRIAL SYSTOLE

- Atrial systole or the atrial contraction phase lasts for 0.1 s and coincides with the last rapid filling phase of ventricular diastole (Figs 4.3-1 and 4.3-2).
- Before the beginning of atrial systole, the ventricles are relaxing, the atrioventricular (AV) valves are open and the blood is flowing from the great veins into the atria and from the atria into the ventricles. Thus, the atria and ventricles are forming a continuous cavity.
- When the atrial contraction begins, about 75% of the blood has already flown into the ventricles. Thus, atrial contraction usually causes an additional 25% filling of the ventricles.

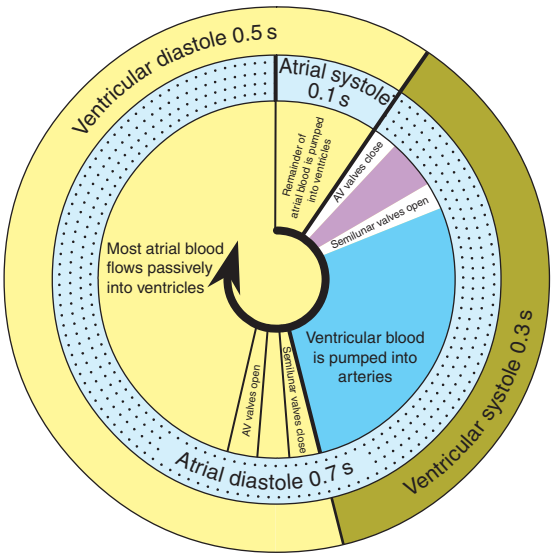


Fig. 4.3-1 Duration of phases of cardiac cycle.

The contraction of atria causes:

- *Increase in the intra-atrial pressure* by 4–6 mm Hg in the right atrium and 7–8 mm Hg in the left atrium. The pressure rise in the right atrium is reflected into the veins and is recorded as *α-wave* from the jugular vein.
- *Increase in the ventricular pressure* occurs slightly due to pumping of blood in the ventricles.
- *Narrowing of origin of great veins* (inferior vena cava and superior vena cava opening in right atrium) and pulmonary veins opening in left atrium) decreasing venous return to the heart. Some regurgitation of the blood occurs into the great veins as no valves are present between them and the atria.

ATRIAL DIASTOLE

After the atrial systole, there occurs atrial diastole (0.7 s). This period coincides with the ventricular systole and most of the ventricular diastole (Fig. 4.3-1).

During the atrial diastole, the atrial muscles relax and there occurs gradual filling of the atria due to continuous venous return and the pressure gradually increases in the atria to drop down to almost zero with the opening of AV valves (Fig. 4.3-3). Then the pressure again rises and follows the ventricular pressure during the rest of atrial diastole.

VENTRICULAR CYCLE

VENTRICULAR SYSTOLE

After the atrial contraction phase is over, the ventricles get excited by the impulse travelling along the conduction system and the ventricles start contracting. The ventricular systole lasts for 0.3 s and has following phases:

1. Phase of isovolumic (isometric) contraction

- With the beginning of ventricular contraction, the ventricular pressure exceeds the atrial pressure very rapidly causing closure of AV valves (this event is responsible for the production of first heart sound).
- Since the AV valves have closed and semilunar valves have not opened, so the ventricles contract as a closed

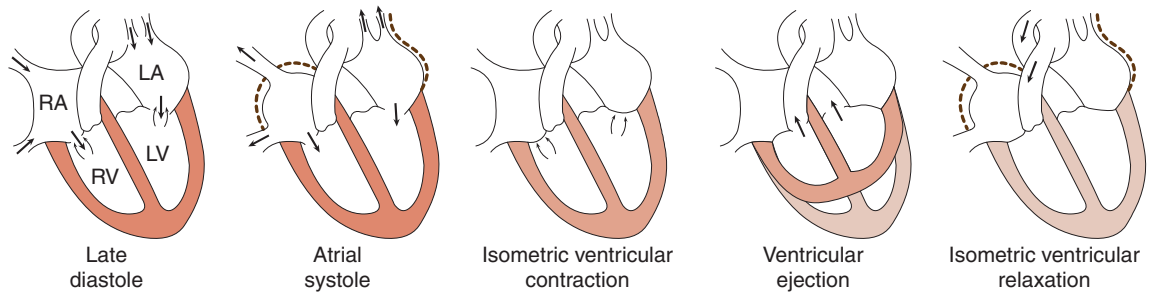


Fig. 4.3-2 Phases of cardiac cycle.

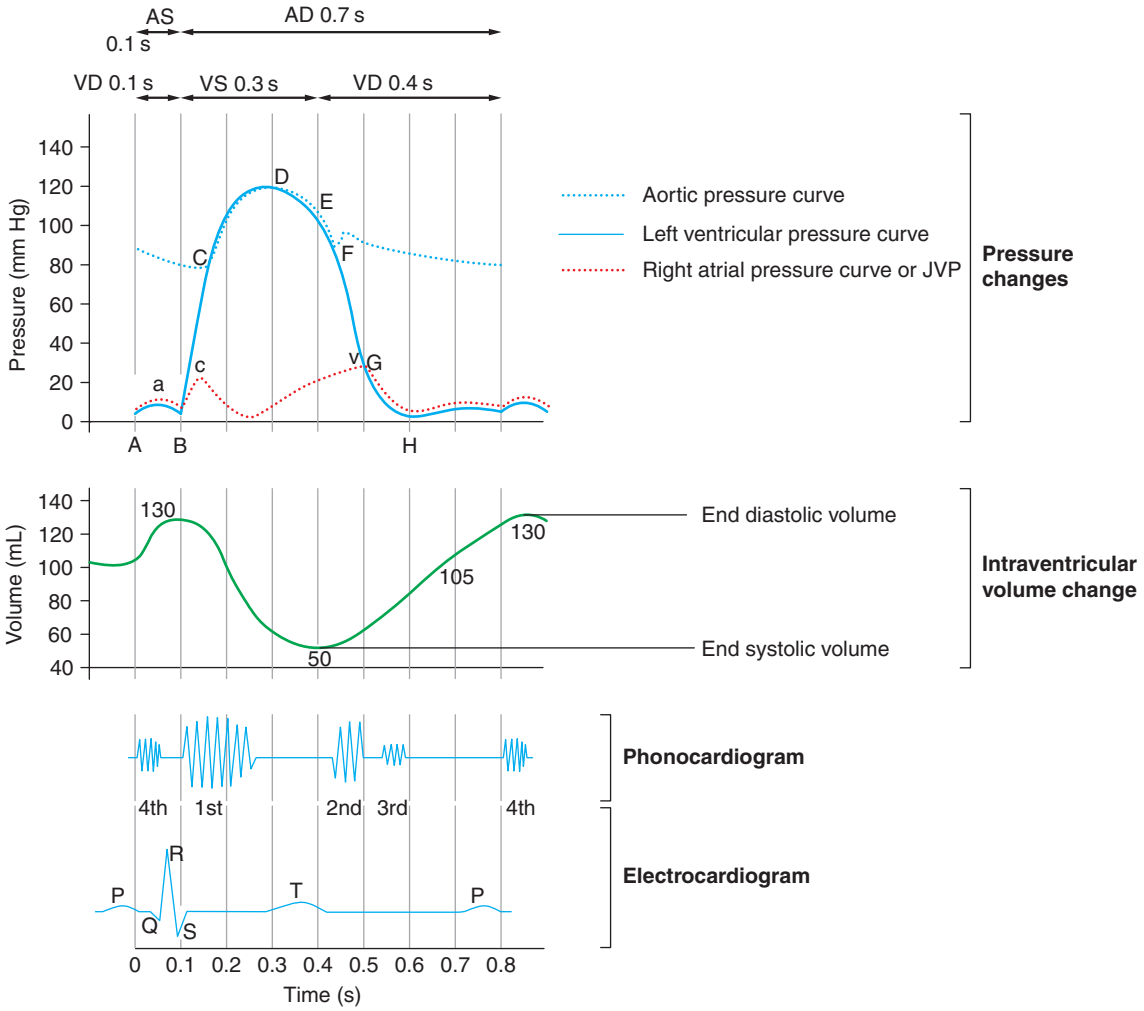


Fig. 4.3-3 Phases and events during cardiac cycle.

chamber and the pressure inside the ventricles rises rapidly to a high level.

- As the ventricles contract, but the volume of blood in the ventricles does not change, so this phase is called *isovolumic contraction phase.*
- During this phase, due to sharp rise in the ventricular pressure, there occurs bulging of AV valves into the atria producing a small but sharp rise in the intra-atrial pressure called *c-wave.*
- This phase lasts for 0.05 s, until the pressure in the left and right ventricles exceeds the pressure in the aorta (80 mm Hg) and pulmonary artery (10 mm Hg) and the aortic and pulmonary valves open.

2. Phase of ventricular ejection

The ventricular ejection phase begins with the opening of semilunar valves and lasts for about 0.25 s. It can be further divided into two phases:

Rapid ejection phase. As soon as the semilunar valves open, the blood is rapidly ejected out for about 0.1 s. About two-thirds of the stroke volume is ejected in this rapid ejection phase. Pressure rises to 120 mm Hg in the left ventricle and to 25 mm Hg in the right ventricle. The right ventricular ejection begins before that of left and continued even after left ventricular ejection is complete. As both the ventricles almost eject same volume of blood, the velocity of right ventricular ejection is less than that of the left ventricle.

Slow ejection phase. It refers to the latter two-third of systole (about 0.15 s) during which the rate of ejection declines. About one-third of the stroke volume is ejected during this phase. The intraventricular pressure starts declining and falls to a value slightly lower than in the aorta, but for a short period momentum keeps the blood flowing forward.

Volume changes. At the end of each diastole, the ventricular volume is about 130 mL. This is called end-diastolic volume. About 80 mL of blood is ejected out by each ventricle during each systole. This is called *stroke volume.* Thus, about 50 mL of the blood is left in each ventricle at the end of systole. This is called *end-systolic volume.*

VENTRICULAR DIASTOLE

1. Protodiastole

When the ventricular systole ends, the ventricles start relaxing and intraventricular pressure falls rapidly. This phase lasts for 0.04 s. During this phase, the elevated pressure in the distended arteries (aorta and pulmonary artery) immediately pushes the blood back towards ventricles which snaps the semilunar valves to close. Closure of semilunar (i.e. aortic and pulmonary) valves prevents the movement of blood back into the ventricles and produces the *second heart sound* (S_2). It also causes a *dicrotic notch* in the down slope of aortic pressure curve called the *incisura.*

2. Isovolumic or isometric relaxation phase

- This phase begins with the closure of the semilunar valves and lasts for about 0.06 s.
- Since semilunar valves have closed and the AV valves have not yet opened, so the ventricles continue to relax as closed chambers in this phase. This causes rapid fall of pressure inside the ventricles (from 80 mm Hg to about 2–3 mm Hg in the left ventricle).
- As in this phase, the ventricular volume remains constant, so this phase is called isovolumic or isometric relaxation phase.
- This phase ends when the AV valves open, as indicated by the peak of *v-wave* on the atrial pressure tracing (Figs 4.3-2 and 4.3-3).

3. Rapid passive filling phase (0.11 s)

- During ventricular systole, the atria are in diastole and venous return continues so that the atrial pressure is high. When the AV valves open, the high atrial pressure causes a rapid, initial flow of blood into the ventricles. The rapid passive filling phase produces the *third heart sound (S3),* which is not normally audible in adults but may be heard in children.
- Once the AV valves open, the atria and ventricles are a common chamber and pressure in both cavities falls as ventricular relaxation continues.

4. Reduced filling and diastasis (0.19 s)

In this phase, pressure in the atria and ventricles reduces slowly and remains little above zero. This decreases the rate of blood flow from the atria to ventricle causing a very slow filling called *diastasis.*

Note. It is important to note that about 75% of blood passes from the atria to the ventricles during rapid filling and reduced filling phases of the ventricular diastole.

5. Last rapid filling phase (0.1 s)

The last rapid filling phase of ventricular diastole coincides with the atrial systole. As described in the beginning, the atrial systole brings about the last rapid filling phase and pushes the additional 25% of the blood in the ventricles. With this phase, the ventricular cycle is completed.

Cardiac cycle: right versus left heart

Both the ventricles pump the same volume of blood over any significant time period; therefore, by and large events on the two sides of the heart are similar. However, there exists a minor asynchronicity between the two sides as:

- Right atrial systole precedes left atrial systole, but the right ventricle starts contracting after the left ventricle. However, the right ventricular ejection begins before the left ventricular ejection, because the pulmonary arterial pressure is lower than the aortic pressure.
- The pulmonary and aortic valves close at the same time during expiration, but the aortic valve closes slightly before the pulmonary valve during inspiration. The slower closure of pulmonary valve during inspiration is because of two factors:
  - Decrease in the resistance of pulmonary vascular tree with prolonged ejection and
  - An increase in systemic venous return which prolongs ejection.

EVENTS DURING CARDIAC CYCLE

The events associated with contraction and relaxation of the heart during a cardiac cycle include pressure changes in the ventricles, atria and aorta; volume changes in the ventricles and valvular events. Most of these have been described during various phases of cardiac cycle; however, they are once again repeated to highlight them.

PRESSURE CHANGES IN THE VENTRICLES

Pressure changes in the ventricles during the cardiac cycle are consistent with the maintenance of systemic and pulmonary circulation. The intraventricular pressure can be measured with the help of cardiac catheterization. Pressure changes observed during various phases of cardiac cycle are as depicted in Figs 4.3-2 and 4.3-3.

During atrial systole

Before the onset of atrial systole, i.e. during diastasis the pressure inside the ventricles is a little above zero. During atrial systole, there occurs a slight increase in the intraventricular pressure (about 6–7 mm Hg in the right ventricle and about 7–8 mm Hg in the left ventricle) due to pumping of blood in the ventricles. In the intra-ventricular pressure curve (Fig. 4.3-3), the segment AB represents the pressure changes during the atrial systole. The point A denotes the onset of atrial systole and the point B denotes the closure of AV valve.

During ventricular systole

Phase of isovolumic (isometric) contraction, since the AV valves have closed and the semilunar valves have not opened, so the ventricles contract as a closed chamber and pressure inside them rises rapidly to a high level. In the intraventricular pressure curve (Fig. 4.3-3), this phase is represented by the segment BC. The point C denotes the opening of semilunar valves and commencement of ventricular ejection phase.

During rapid ejection phase, the ventricles contract at a rate greater than the rate at which blood is ejected so a great rise in the pressure occurs. Pressure rises to maximum of 120 mm Hg in the left ventricle and 25 mm Hg in the right ventricle. The maximum pressure in the left ventricle is four to five times more than in the right ventricle. This is because of the thick wall of the left ventricle. In the intraventricular pressure curve (Fig. 4.3-3), this phase is denoted by the segment CD. The point D denotes the peak point of the intraventricular pressure after which it starts declining.

During slow ejection phase, there is no further ventricular contraction and the pressure starts declining (Fig. 4.3-3, segment DE).

During ventricular diastole

During protodiastole, the intraventricular pressure drops rapidly as the ventricles start relaxing. When the intraventricular pressure falls below that of the aorta and the pulmonary artery, the semilunar valves are closed due to back flow of blood. In the intraventricular pressure curve (Fig. 4.3-3), this phase is represented by the segment EF and the point F denoted the closure of semilunar valves.

During isovolumic (isometric) relaxation phase, since the semilunar valves have closed and AV valves have not yet opened up so the ventricles relax as closed chamber and there occurs a rapid fall in the intraventricular pressure (from 80 mm Hg to about 2–3 mm in the left ventricle). When the pressure inside the ventricles fall below the pressure in the atria, the AV valves open up and the phase of rapid passive filling commences. In the intraventricular pressure curve (Fig. 4.3-3), the segment FG represents this phase and the point G coincides with the opening of the AV valve.

During rapid passive filling phase, the intraventricular pressure further falls since the ventricles are relaxing though blood is being filled in them (segment GH in Fig. 4.3-3).

During reduced passive filling phase, there is no turbulence and the blood flows very slowly and smoothly and virtually there occurs cessation of ventricular filling *(diastasis).* The ventricular pressure remains a little above zero.

PRESSURE CHANGES IN THE ATRIA

The intra-atrial pressure can be recorded with the help of an intracardiac catheterization. The left atrial pressure can also be determined indirectly by measuring the pulmonary capillary wedge pressure. The tracing of the jugular venous pulse is also similar to the intra-atrial pressure curve and it has three positive waves called a, c and v (Fig. 4.3-3). Relationship of intra-atrial pressure changes with the phases of cardiac cycle is:

During atrial systole

Before the onset of atrial systole, the intra-atrial pressure is slightly above zero and is slightly greater than the ventricular pressure. During atrial systole, there occurs a sharp rise in the intra-atrial pressure (by 4–6 mm Hg) in the right atrium and by 7–8 mm Hg in the left atrium) and causes a pressure wave recorded as *a wave* from the jugular vein ('a' stands for atrial systole). Immediately after atrial systole, the intra-atrial pressure falls due to start of atrial relaxation in the atrial diastole (Fig. 4.3-3).

During ventricular systole

During phase of isovolumic (isometric) contraction, due to sharp rise in the intraventricular pressure, AV valves bulge into the atria producing a small but sharp rise in the atrial pressure producing the so-called *c wave* ('c' stands for contraction of the ventricle).

During ventricular ejection phase, the intra-atrial pressure drops sharply in the rapid ejection phase. This happens so, because the papillary muscles (attached to the cusps of AV valves by chordae tendineae) contract when the ventricular walls contract and pull down the fibrous AV ring causing enlargement of the atrial lumen and thus decreasing the intra-atrial pressure. Therefore, as the ventricles contract, the atria get slowly filled with blood flowing in from the great veins and the atrial pressure starts rising.

During ventricular diastole

During isovolumic (isometric) relaxation phase, the atrial pressure continues to rise as long as AV valves remain closed, i.e. till the end of isovolumic relaxation. This results in the third positive wave called *v wave* ('v' stands for venous filling). This shows a gradual increase in the atrial pressure.

During rapid passive filling phase, the AV valves open allowing rapid flow of blood from the atria to the ventricles. So, the atrial pressure drops sharply to a little above zero level and remains so till the beginning of the next atrial systole.

PRESSURE CHANGES IN THE AORTA

Pressure in the aorta varies between 80 and 120 mm Hg during the cardiac cycle and can be recorded by using catheter.

Aortic pressure changes during various phases of the cardiac cycle (Fig. 4.3-3) are:

During atrial systole

During atrial systole, the pressure in the aorta is about 80 mm Hg.

During ventricular systole

During ventricular systole, the intraventricular pressure rises and reaches above that of the aorta during beginning of the *ventricular ejection phase* when the aortic semilunar valve opens and blood starts flowing from the left ventricle into the aorta. Hence, the aortic pressure starts rising along with the intraventricular pressure during the rapid ejection phase and reaches maximum (120 mm Hg) at the end of the rapid ejection phase. It is important to note that during most of the rapid ejection phase, the aortic pressure remains slightly lesser than the ventricular pressure and *during reduced ejection phase,* the aortic pressure starts falling along with the ventricular pressure.

During ventricular diastole

During protodiastole, aortic pressure is slightly higher than that in the left ventricle. This causes backward flow of blood and closure of the aortic semilunar valve. Due to sudden closure of the semilunar valve the back flowing blood collides against the closed aortic valve. This collision causes a small but sharp rise in the aortic pressure. This small rise produces a notch called *incisura.* This sharp pressure rise is recordable even from the peripheral arteries and is called *dicrotic notch.*

During rest of the diastole, the aortic pressure smoothly declines. By the time the aortic pressure declines to about 80 mm Hg, another ventricular systole boosts the aortic pressure again.

PRESSURE CHANGES IN THE PULMONARY ARTERY

Pressure curve in the pulmonary artery is similar to that of the aorta but pressures are low (about one-sixth of that in aorta). Pulmonary artery systolic pressure averages 15–18 mm Hg and its pressure during diastole is 8–10 mm Hg.

VOLUME CHANGES IN THE VENTRICLES DURING CARDIAC CYCLE

During atrial systole

Atrial systole coincides with the last rapid filling phase of the ventricular diastole. When the atrial contraction begins, about 105 mL (75%) of the blood has already flown into the ventricles. The atrial contraction causes additional 25 mL (25%) filling of the ventricles. In the ventricular volume curve (Fig. 4.3-3), this phase is represented by the AB segment. Thus at the end of atrial systole, i.e. at the end of the ventricular diastole, the ventricular volume is about 130 mL. This is called *end-diastolic volume.*

During ventricular systole

During isovolumic contraction phase, as the name suggests, there occurs no change in the ventricular volume (Fig. 4.3-3, segment BC).

During ventricular ejection phase about 80 mL of the blood is ejected out by each ventricle. This is called *stroke volume* (Fig. 4.3-3, segment CD). The percentage of the end-diastolic volume that is ejected out with each stroke during systole (about 65%) is called *ejection fraction.* Thus, about 50 mL of the blood in each ventricle at the end of the ventricular systole is called *end-systolic volume.*

During protodiastole and phase of isovolumic relaxation, there occurs no change in the ventricular volume (Fig. 4.3-3, segments DE and EF).

During rapid filling phase and slow filling phase, the ventricular volume changes rapidly and then slowly, respectively. About 75% of the ventricular filling (105 mL of blood) occurs during these phases (Fig. 4.3-3, segment FG and GH).

VALVULAR EVENTS (HEART SOUNDS)

A total of four heart sounds (1st, 2nd, 3rd and 4th) are produced by certain mechanical activities during each cardiac cycle.

First heart sound (HS_1)

Cause. First heart sound is produced by the vibrations set up by the sudden closure of AV valves at the start of ventricular systole, during phase of isovolumic contraction (Fig. 4.3-3).

Characteristics. The first heart sound is long and soft when heart rate is low and loud when the heart rate is high. Its duration is about 0.15 s and frequency is 25–45 Hz. It sounds like the spoken word 'LUBB'.

Site for auscultation. It can be heard by auscultation of the chest with stethoscope. It is best heard over mitral and tricuspid areas. *Mitral area* is located in the fifth intercostal space just internal to mid clavicular line. *Tricuspid area* is located in the fifth intercostal space near the sternum (Fig. 4.3-4).

In phonocardiogram, the first heart sound is recorded as a single group of 9–13 waves. The amplitude of the waves is small to start with but later rapidly rises to fall to form *crescendo and diminuendo* series of waves (Fig. 4.3-3).

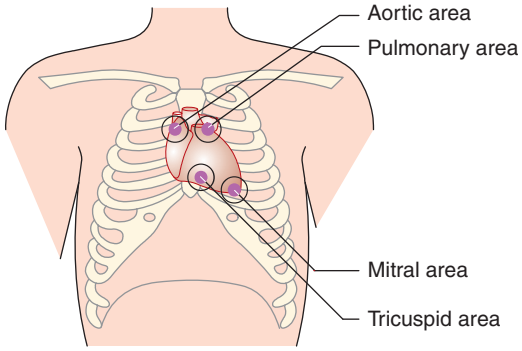


Fig. 4.3-4 Auscultatory areas over the chest.

Correlation with ECG. First heart sound coincides with peak of R wave in ECG.

Second heart sound (HS_2)

Cause. It is caused by the vibrations associated with closure of the semilunar valves just at the onset of ventricular diastole.

Characteristics. The second heart sound is short, loud, high pitched sound. Its duration is 0.12 s and frequency is 50 Hz. It sounds like the spoken word 'DUBB'.

Site for auscultation. It can be heard by auscultation of the chest with stethoscope. It is best heard over the aortic and pulmonary areas. *Aortic area* lies in the right second intercostal space near the sternum, and *pulmonary area* is in the left second intercostal space close to sternum (Fig. 4.3-4).

In phonocardiogram, second heart sound is recorded as a single group of 4–6 waves having same amplitude (Fig. 4.3-3).

Correlation with ECG. Second heart sound usually coincides with the end of T wave in ECG.

Third heart sound (HS_3)

Cause. Third heart sound is caused by the vibrations set up in the cardiac wall by inrush of blood during *rapid filling phase* of the ventricular diastole.

Characteristics. Third heart sound is a short, soft and low pitched sound. Its duration is 0.1 s. Normally, it cannot be heard by auscultation with stethoscope.

In phonocardiogram, the third heart sound is recorded as 1–4 waves grouped together (Fig. 4.3-3).

Correlation with ECG. The third heart sound appears between T and P waves of ECG.

Fourth heart sound (HS_4)

Cause. It is caused by the vibrations set up during the atrial systole which coincides with last rapid filling phase of the ventricular diastole.

Characteristics. It is normally not audible. Sometimes it can be heard immediately before the first sound when atrial pressure is high or when ventricle is stiff in condition such as ventricular hypertrophy. It is a short and low pitched sound. Its duration is about 0.03 s and frequency about 3 Hz.

In phonocardiogram, the fourth heart sound merges with first heart sound many times. When it appears as a separate entity, it has 1–2 waves with very low amplitude (Fig. 4.3-3).

Correlation with ECG. Fourth heart sound coincides with the interval between the end of P wave and onset of Q wave.

Cardiac murmurs

Cardiac murmurs are the abnormal heart sounds produced during the cardiac cycle.

Mechanism of production. Cardiac murmurs are produced by a turbulent blood flow or by change in the direction of blood flow. Normally, the blood flows through the heart and blood vessels as *laminar flow* which is streamlined and silent. The *turbulent flow,* on the other hand, produces vibrations in the tissues that are heard as murmurs.

Causes. Murmurs are caused in the following conditions:

- *Valvular stenosis,* i.e. narrowing of any of the cardiac valve (mitral, tricuspid, aortic or pulmonary valve).
- *Valvular insufficiency,* i.e. regurgitation of any of the cardiac valve.
- *Ventricular septal defect,* i.e. a congenital hole in the ventricular septum.
- *Atrial septal defect,* i.e. a congenital hole in the interatrial septum.
- *Coarctation of aorta,* i.e. congenital narrowing of systemic aorta.
- *Patent ductus arteriosus,* i.e. a congenital disorder in which there is backward flow of blood from the aorta into the pulmonary artery.

Types. Depending upon the timing of appearance these have been classified as (Fig. 4.3-5):

- *Systolic murmur,* which is produced during systole,
- *Diastolic murmur,* which is produced during diastole and
- *Continuous murmur,* which is produced continuously.

The types of murmur produced depend upon the site and type of abnormality (Table 4.3-1).

Site of auscultation. The murmurs are best heard by placing the stethoscope on the chest wall closest to their origin, e.g. aortic area, pulmonary area, mitral area or tricuspid area.

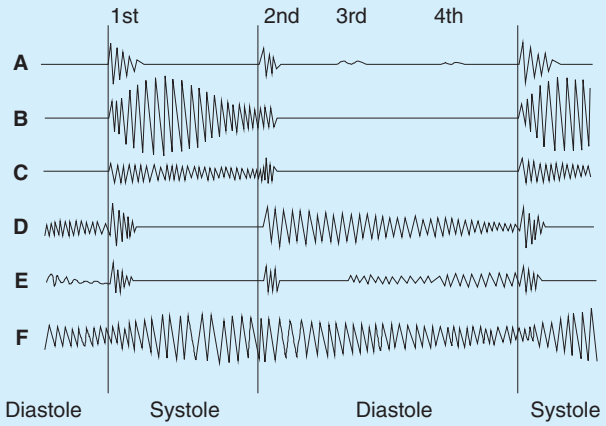


Fig. 4.3-5 Cardiac murmurs on phonocardiography: A, normal phonocardiogram; B, systolic murmur (in aortic stenosis); C, systolic murmur (in mitral regurgitation); D, diastolic murmur (in aortic regurgitation); E, diastolic murmur (in mitral stenosis) and F, continuous murmur (in patent ductus arteriosus).

Table 4.3-1 Types of murmurs depending on site and type of abnormality

Site of abnormality	Type of abnormality	Type of murmur
Aortic or pulmonary valve	Stenosis Insufficiency	Systolic Diastolic
Mitral or tricuspid valve	Stenosis Insufficiency	Diastolic Systolic
Interventricular septum	Cogenital hole	Systolic
Aorta	Coarctation	Systolic
Ductus arteriosus	Patent	Continuous
Blood	Anaemia	Systolic

DURATION OF SYSTOLE AND DIASTOLE VIS-A-VIS HEART RATE

NORMAL DURATION

- Duration of each cardiac cycle at a normal heart rate of 75 beats/min is 60/75 = 0.8 s,
- Duration of ventricular systole is 0.3 s and
- Duration of ventricular diastole is 0.5 s.

Effect of heart rate

- Cardiac muscle has the unique property of contracting and repolarizing faster when the heart rate is high.
- Therefore, when the heart rate increases the total duration of cardiac cycle decreases, e.g. at a heart rate of 200 beats/min the total duration of cardiac cycle is 60/200 = 0.3 s.
- It is important to note that though the duration of all phases of the cardiac cycle decreases at high heart rate, but the *duration of diastole* decreases much more than the duration of systole. For example, when the heart rate increases from 75 to 200 beats/min, the duration of systole decreases from 0.3 to 0.16 s while that of diastole decreases from 0.5 to 0.14 s (Table 4.3-2).

Table 4.3-2 Change in duration of various events of cardiac cycle with increase in the heart rate

Event	Duration (s)	
	At heart rate of 75 beats/min	At heart rate of 200 beats/min
Cardiac cycle	0.8	0.3
Systole	0.3	0.16
Diastole	0.5	0.14

This fact has following important physiologic and *clinical implications:*

- It is during diastole that the heart muscle rests and coronary blood flow to the subendocardial portion of the left ventricle occurs only during diastole. Therefore, at a very high rate there occurs reduction in cardiac perfusion and there are chances of myocardial ischaemia.
- Furthermore, most of the ventricular filling occurs in the diastole. At heart rate up to 180 beats/min, filling is adequate as long as there is ample venous return and cardiac output per minute is increased by an increase in heart rate. However, at very high heart rate, filling may be compromised to such a degree that cardiac output per minute falls and symptoms of the heart failure develop.

ARTERIAL PULSE

Arterial pulse is also an event related to the cardiac cycle. The blood forced into the aorta during the systole not only moves the blood in the vessels forward, but also sets up a pressure wave that is transmitted along the *arteries to the* periphery. The pressure wave expands the arterial walls as it travels and expansion is palpable as the *pulse.*

Velocity of transmission of pulse wave

Velocity of transmission of pulse wave is independent of and much higher than the velocity of blood flow, the maximum value of which in larger arteries is only 50 cm/s. The rate of travel of pulse wave is about:

- 4 m/s in the aorta and its branches,
- 8 m/s in the large arteries and
- 16 m/s in the small arteries of young adults.

Consequently, the pulse is felt at the arteries after short interval of the peak of systolic ejection into the aorta.

Methods of recording arterial pulse

The tracings of arterial pulse can be made by the following techniques of recording:

1. Manometric technique. It is used in animals. In this technique, a cannula is inserted into the dissected artery and is connected to manometer or any other recording device to obtain the arterial pulse tracing.

2. Electronic transducer method. The electronic transducer is placed on the skin overlying any artery. The transducer throws light on the artery and the light reflected from the flowing blood is deducted by the sensor of the transducer. The arterial pulse in the form alterations in the frequency of reflected light rays are amplified and recorded by connecting the transducer to a recording device like polygraph.

Interpretation of arterial pulse tracing

The pulse tracing recorded from the carotid artery shows the following characteristics (Fig. 4.3-6):

- *Ascending limb,* also known as *anacrotic limb* or primary limb, is due to the rise in pressure during systole.
- *Descending limb,* also known as *catacrotic limb* represents the fall in pressure during diastole.
- *Percussion wave (P)* corresponds to the ejection phase of the ventricular systole.
- *Tidal wave (T)* is due to falling blood column during slow ejection phase.
- *Dicrotic notch (N)* is due to the closure of aortic valve and marks end of the ventricular systole.
- *Dicrotic wave (D)* is due to rebound of blood column from the closed aortic valve.

When the record is taken from the peripheral arteries at a distant place from the heart, e.g. femoral or radial arteries the contour or shape of record changes. In arterioles and capillaries, the waves disappear.

Examination of arterial pulse

Examination of the arterial pulse is an essential feature of a clinical examination. Arterial pulse can be palpated from any superficial artery, e.g. radial, femoral, dorsalis pedis and carotid, etc. Most frequently, pulse is examined from the radial artery because it is conveniently approached without exposing the body and can be easily palpated as it is placed superficially against the bone.

Examination of the pulse should include following aspects:

1. Pulse rate refers to the number of pulses per minute. It is a convenient method of determining the heart rate.

- *Normal pulse rate* varies with age being 150–180 min in fetus, 130–140 min at birth, about 90 min at the age of 10 years and about 72/min in adults.

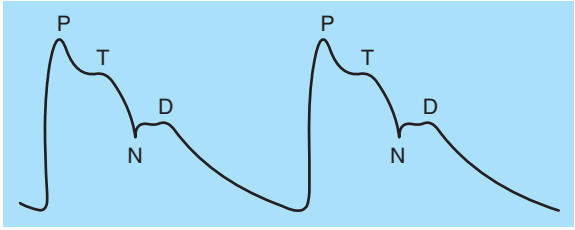


Fig. 4.3-6 Record of arterial pulse from a carotid artery: (P) percussion wave; (T) tidal wave; (N) dicrotic notch and (D) dicrotic wave.

- *Increased pulse rate* represents tachycardia and occurs during exercise, in anxiety, in fever, in hyperthyroidism and in atrial and ventricular tachycardias.
- *Decreased pulse rate* represents bradycardia and is seen in hypothyroidism and incomplete heart blocks.

2. Volume of pulse also known as strength of arterial pulse or amplitude or impact can be felt. It represents stroke volume or the pulse pressure (i.e. systolic–diastolic pressure).

- *Rapid and thready pulse* occurs in hypovolaemia as in severe haemorrhage and there is marked reflex vasoconstriction.
- *Increased volume pulse* is seen during exercise and in ventricular hypertrophy.

3. Rhythm of pulse is noted as regular or irregular. Under normal conditions and during sinus bradycardia or sinus tachycardia pulse appears at regular intervals.

- *Irregular pulse rhythm* is a feature of extra systole, atrial fibrillation and other cardiac arrhythmias, type of arrhythmia is confirmed only on ECG. The irregular pulse rhythm may be regularly irregular or irregularly irregular.

4. Character of pulse is felt on palpation. It denotes the tension and waves in the pulse. Feeling of different characters of the pulse can be learnt by the subjective experience.

- *Normal character* of the pulse is sinuous on examination, i.e. an upstroke is followed by a downstroke (Fig. 4.3-7A). Normally, it is not possible to feel the different waves of the pulse or slight variations in the character. However, in certain heart diseases and valvular defects the normal character is altered and can be easily felt while palpating the peripheral arterial pulse.

A few *abnormal characters* of the pulse are described as:

- *Water hammer pulse* (Fig. 4.3-7B), also known as collapsing pulse, is characterized by a sudden upstroke followed by a sudden downstroke. It is best felt by raising the patient's arm and holding it by grasping the wrist with palm of the observer. Sometimes, the upstroke is so strong

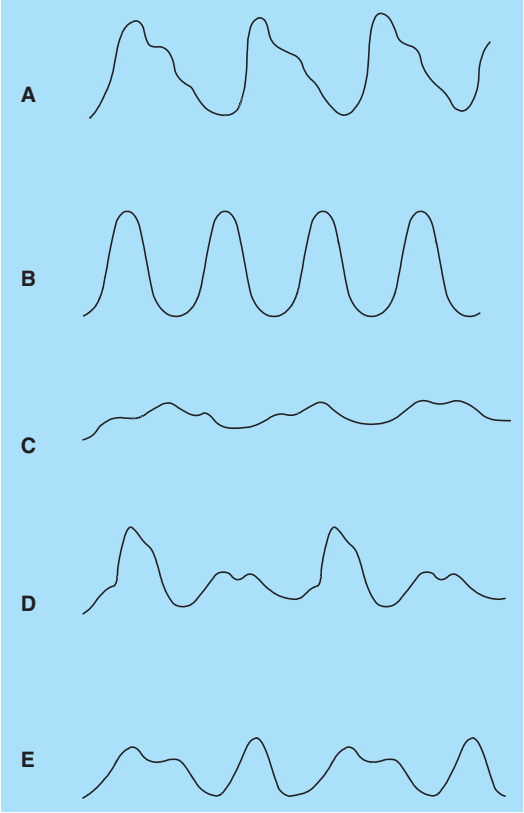


Fig. 4.3-7 Character of arterial pulse: A, normal character; B, water hammer pulse; C, anacrotic pulse; D, pulsus alternans and E, pulsus bisferiens.

that it leads to head nodding with each heart beat called as Corrigan sign or Corrigan pulse.

The cause of water hammer pulse is aortic insufficiency or regurgitation.

- *Anacrotic pulse* (Fig. 4.3-7C). Normally, there is a single upstroke in the arterial pulse. In an anacrotic pulse, there are two upstrokes. It occurs in patients with aortic stenosis.
- *Bisferiens pulse* (Fig. 4.3-7E). This a combination of an anacrotic and collapsing pulses, both can be felt distinctly. It is found in a condition when combined aortic stenosis and incompetency.
- *Pulsus alternans* (Fig. 4.3-7D). In this condition, every normal pulse alternates with a weak pulse.
- *Pulsus paradoxus.* The pulse becomes smaller or even disappears at the end of inspiration when patient breathes deeply.

CARDIAC OUTPUT AND VENOUS RETURN

DEFINITION OF CARDIAC OUTPUT AND RELATED TERMS

The main function of the heart is to pump blood to meet the metabolic needs of the body. The measure of the heart's ability to pump blood is cardiac output. The *cardiac output* refers to the amount of blood ejected by each ventricle per minute. The *stroke volume* is the amount of blood pumped out by each ventricle per beat or per contraction. Therefore, cardiac output (CO) can be calculated by multiplying the stroke volume (SV) by the heart rate (HR):

$$CO = SV \times HR$$

Under normal conditions the average heart rate is about 70 beats/min and stroke volume is about 80 mL and thus cardiac output is $80 \times 70 = 5.6$ L.

The cardiac output is expressed in litres per minute and normally varies from 5–6 L/min. In health, the right and the left ventricular outputs are nearly equal. Thus, each ventricle pumps about 5–6 L of blood into the circulation per minute. This is made possible because of the fact that the right and the left side pumps act in series.

Cardiac index is the cardiac output expressed in relation to the body surface area. The normal cardiac index is about 3.2 L/min/m^2.

Distribution of the cardiac output

Of the total cardiac output, about 75% is distributed to the vital organs of the body and rest of 25% to the skeletal muscle, other organs of the body and skin. The distribution of the cardiac output to various organs of the body is shown in Table 4.3-3.

Table 4.3-3 Distribution of cardiac output to various organs

Body organ	Amount of blood flow (mL/min)		Percentage of total cardiac output	
Liver	1500		25	
Kidney	1300		about 25	
Brain	750			75
Heart	250	1500	25	
Lungs	500			
Skeletal muscles and other body organs	1000	1500	25	
Skin	500			

MEASUREMENT OF CARDIAC OUTPUT

Cardiac output, in experimental animals, can be measured directly with the help of an electromagnetic flowmeter placed on the ascending aorta. However, in human only indirect methods are possible and include:

- Methods based on Fick's principle
- Indicator or dye dilution method
- Thermodilution method

- Method employing inhalation of inert gases
- Physical methods such as
  - Doppler technique echocardiography
  - Ballistocardiography

METHODS BASED ON FICK'S PRINCIPLE

Fick's principle

The Fick's principle states that the amount of a substance taken up by an organ (or by the whole body) per unit of time is equal to the arterial level of the substance (A) minus the venous level (V) times the blood flow (F), i.e.

$$Q = (A - V)\,F$$

or

$$F = \frac{Q}{(A - V)}$$

This principle can be of course, only in situations in which the arterial blood is the sole source of the substance taken up.

In this method (Fig. 4.3-8) cardiac output is determined by measuring the pulmonary blood flow. As we know:

- Pulmonary blood flow/min = right ventricular output.
- Right ventricular output = left ventricular output (cardiac output).

Measurement of pulmonary blood flow can be made by measuring the amount of O_2 taken by the blood from the lungs, O_2 concentration of the venous blood from pulmonary artery (PAO_2) and O_2 concentration of the arterial blood from the pulmonary vein (PVO_2).

- Amount of O_2 uptake/min is determined with the help of a spirometer,
- PAO_2 is measured from the venous blood sample taken from the pulmonary artery directly with the help of a cardiac catheter. The cardiac catheter is inserted into a vein at the forearm and is then guided up under fluoroscopic control through the venous channels into the right atrium, right ventricle and pulmonary artery.
- PVO_2, because of practical difficulty in taking sample from pulmonary vein, is measured from the arterial blood sample taken from any peripheral artery, e.g. brachial artery (the O_2 content of all the major arteries is same as that of pulmonary veins). According to Fick's principle:

$$\text{Pulmonary blood flow} = \frac{\text{Amount of } O_2 \text{ taken by the lungs/min}}{PVO_2 - PAO_2}$$

$$\text{Cardiac output} = \frac{O_2 \text{ taken up by the lungs/min}}{PVO_2 - PAO_2}$$

For example, if O_2 uptake is 250 mL/min. PVO_2 is 19 mL/100 mL and PAO_2 is 14 mL/100 mL, then

$$\text{Cardiac output} = \frac{250 \times 100}{19 - 14}$$

$$= \frac{25{,}000 \text{ mL/min}}{5}$$

$$= 5000 \text{ mL/min}$$

$$= 5 \text{ L/min.}$$

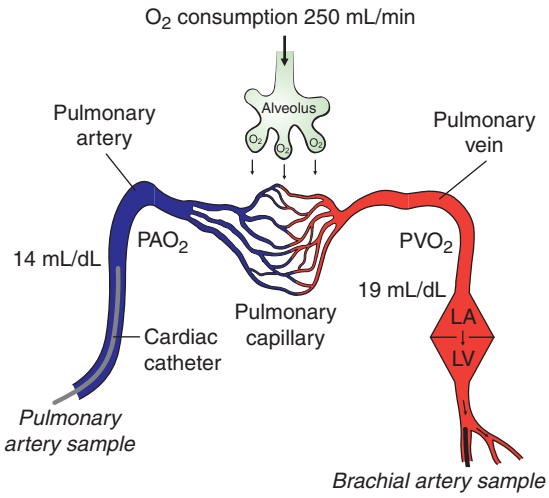


Fig. 4.3-8 Estimation of cardiac output by Fick's principle. (PAO_2 = Oxygen content of pulmonary artery blood; PVO_2 = oxygen content of pulmonary vein blood.)

Disadvantages of Fick's principle

- It is an invasive technique, so there are risks of infection and haemorrhage.
- The cardiac output estimated may be somewhat higher than normal as the patient becomes conscious of the whole technique.
- A fatal complication like ventricular fibrillation may occur if the indwelling catheter irritates the ventricular walls, especially when the cardiac output is being measured during heavy exercise.

INDICATOR OR DYE DILUTION METHOD

Principle

In this method, a known amount of the dye is injected into a large vein or preferably into the right atrium by cardiac catheterization. By its passage through heart and pulmonary circulation it will be evenly distributed in the blood stream. Its mean concentration during the first passage through an artery can be determined from the successive samples of blood taken from the artery. The blood flow in litres/min (F) is given by the following formula:

$$F = \frac{Q}{Ct},$$

where,

F = Blood flow in litres/min,

Q = Quantity of the dye injected,

C = Mean concentration of dye and

t = Time duration in second of the first passage of dye through the artery.

Prerequisites for an ideal indicator

The indicator (dye) used should have following characteristics:

- It should be non-toxic.
- It must mix evenly in the blood.
- It should be relatively easy to measure its concentration.
- It should not alter the cardiac output or haemodynamics of blood flow.
- Either it must not be changed by the body during mixing period or the amount changed must be known.
- The dye commonly used in humans for determining the cardiac output is *Evans blue* (T-1824) or radioactive isotopes.

Procedure

Injection of dye. A few millilitres of venous blood is withdrawn from the antecubital vein and it is mixed with 5 mg Evans blue dye. The blood containing dye is then injected rapidly into the vein.

Estimation of duration of first passage of dye (t) and mean concentration (C) of dye in the arterial blood. Serial samples of the arterial blood from the brachial artery are taken every 2 s and the dye concentration is determined.

- When the dye concentration is plotted as a function of time, a curve shown in Fig. 4.3-9 as A, B, C and D is obtained. The curve shows that the dye concentration reaches a peak and then steadily declines only to rise again (CD part of the curve) owing to recirculation of the dye.
- *Time duration of first passage of dye through the artery (t)* is determined by the extrapolation of the descending limb (BC) of the curve to the time scale axis. The point (E) on the time scale where the extrapolated limb meets it, tells the time (AE) of first circulation of dye in seconds.
- *The mean concentration (C)* of the dye is determined by representing the triangle area ABE as a rectangle AEFG with same area and one of its arm being AE. The height of the rectangle (AG) tells the mean concentration (C) of dye.
- *Calculation of cardiac output* is then made using the formula described above. For example, when:
  - Amount of dye injected (Q) is 5 mg,
  - Time duration for the first circulation is 40 s and
  - Mean concentration of dye (C) = 1.5 mg/L, then

$$\text{Cardiac output} = \frac{Q \times 60}{C \times t} = \frac{5 \times 60}{1.5 \times 40} = \frac{300}{60} = 5\ \text{L/min}$$

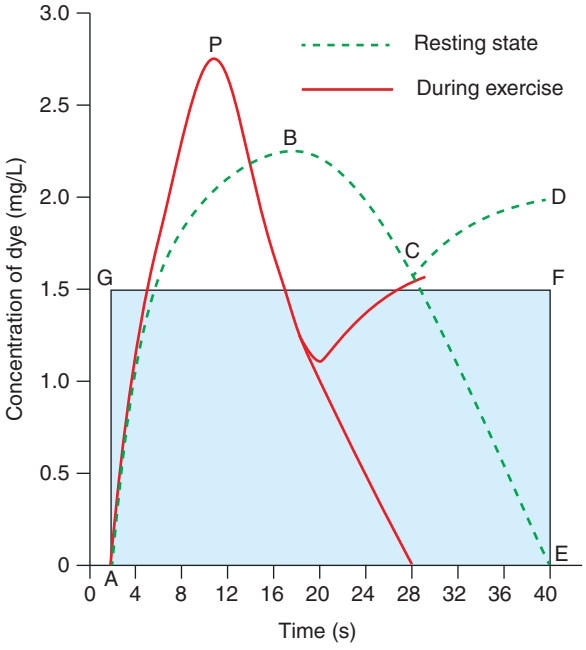


Fig. 4.3-9 Estimation of cardiac output by indicator (dye) dilution method.

THERMODILUTION METHOD

Principle. It is also an indicator dilution technique in which instead of a dye, 'cold saline' is used as an indicator. The cardiac output is measured by determining the resultant change in the blood temperature in the pulmonary artery.

- A known volume of sterile cold saline is then injected into the inferior vena cava.
- Temperature of the blood entering the heart from the inferior vena cava and that of the blood leaving the heart via pulmonary artery is determined by the thermistors.
- The cardiac output is then measured from the values of temperature by applying the principle of indicator dilution technique.

PHYSICAL METHODS

Physical methods developed to measure the cardiac output include the following:

Echocardiography

Echocardiography refers to the ultrasonic evaluation of cardiac functions. It is a noninvasive technique that does not involve injections or insertion of a catheter. It involves B-scan ultrasound at a frequency of 2.25 MHz using a transducer which also acts as a receiver of the reflected waves. The recording of the echoes displayed against time on an oscilloscope provides a record of:

- The movement of the ventricular wall and septum, and valves during the cardiac cycle.
- When combined with the Doppler techniques, echocardiography can be used to measure velocity and volume of flow through the valves.
- Thus, it is particularly useful in evaluating end-diastolic volume (EDV), end-systolic volume, CO and valvular defects.

Ballistocardiography method

This method is not used practically. Ballistocardiography refers to the graphical record of the pulsations created due to ballistic recoil of the pumping heart.

VARIATIONS IN CARDIAC OUTPUT

PHYSIOLOGICAL CAUSES OF VARIATIONS IN CARDIAC OUTPUT

- ***Age.*** Because of less body surface area the children have more cardiac index than adults.
- ***Sex.*** Since the body surface area is less in females so they have more cardiac index than the males.
- ***Diurnal variation.*** In the early morning cardiac output is low which increases in the day time depending upon the basal condition of the individual.
- ***Environmental temperature.*** Moderate change in the environmental temperature does not cause any change in cardiac output. A high environmental temperature is associated with an increase in the cardiac output.
- ***Anxiety and excitement*** are reported to increase the cardiac output by 50–100%.
- ***Eating*** is associated with an increase in cardiac output approximately by 30%.
- ***Exercise*** may increase the cardiac output up to 700% depending upon the vigorousness of exercise.
- ***Pregnancy.*** An increase in cardiac output to the tune of 45–60% is reported during the later months of the pregnancy.
- ***High altitude.*** The cardiac output is increased at a high altitude due to release of adrenaline as a consequence to hypoxia.
- ***Posture change.*** Sitting or standing from lying down position may decrease the cardiac output by 20–30% because of pooling of blood in the lower limbs.

PATHOLOGICAL CAUSES OF VARIATIONS IN CARDIAC OUTPUT

Increase in cardiac output is seen in the following conditions:

- *Fever,* due to increased oxidative processes
- *Anaemia,* due to hypoxia
- *Hyperthyroidism,* due to increased metabolism

Decrease in cardiac output may occur in the following conditions:

- *Rapid arrhythmias,* due to incomplete filling.
- *Congestive cardiac failure,* due to weak contractions of heart.
- *Cardiac shock,* due to poor pumping and circulation.
- *Incomplete heart block,* owing to defective pumping action of the heart.
- *Haemorrhage,* because of decreased blood volume and
- *Hypothyroidism,* due to decreased basal metabolism.

REGULATION OF CARDIAC OUTPUT

The cardiac output increases or decreases in various physiological and pathological conditions as described above. The variations in the cardiac output are brought out by certain factors operating through certain mechanisms by an integrated role.

The CO, as we know, is the product of SV and HR, i.e. CO = SV × HR. Therefore, variations in the cardiac output can be produced by the factors which changes stroke volume or heart rate, or both. The main factors affecting cardiac output are venous return, myocardial contractility, peripheral resistance and heart rate (Fig. 4.3-10).

CARDIAC OUTPUT CONTROL MECHANISMS

The cardiac output is regulated by two mechanisms: intrinsic and extrinsic.

1. Intrinsic autoregulation (Frank–Starling mechanism)

The force of contraction of cardiac muscle fibres like that of the skeletal muscle fibres depends upon its preload. The preload determines the initial length (resting length) of the muscle fibres. According to *Frank–Starling law* of heart, 'within physiological limits' the force of contraction of *cardiac muscle* is proportionate to the initial length of muscle

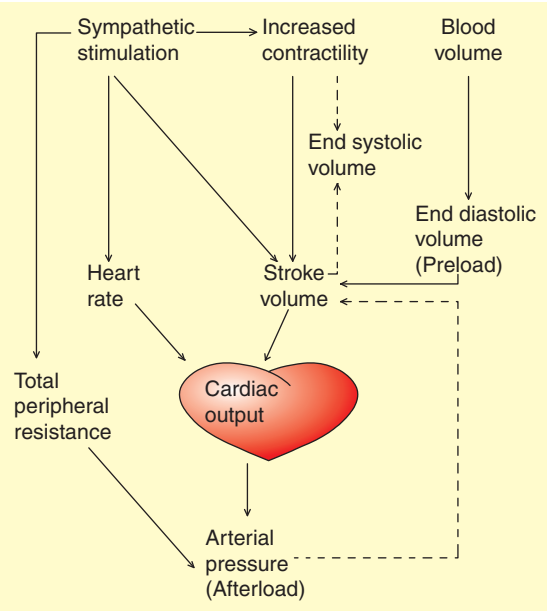


Fig. 4.3-10 Interaction between the factors that regulate cardiac output and arterial pressure. Solid lines indicate increase and dotted lines indicate decrease.

fibres. In the heart, end-diastolic volume forms the preload. Therefore, precisely the Frank–Starling law of heart can be stated as, within physiological limits the force of cardiac contraction is proportional to its EDV. This fact was demonstrated about a century ago by Frank and Starling on the heart–lung preparation in a dog. Since, in this intrinsic regulation mechanism, cardiac muscle fibres are stretched to increase their initial length, it is also termed as *heterometric mechanism.*

The relationship between the ventricular stroke volume and end-diastolic volume is called the *Frank–Starling curve* (Fig. 4.3-11). Details of the effect of preload on the force of myocardial contraction including the *length–tension relationship* are described in detail on page 78.

Factors affecting end-diastolic volume

The end-diastolic volume refers to the *venous return* to the heart during diastole.

Up to physiological limits, the cardiac output is directly proportional to the venous return. Thus, over any significant period of time, venous return must be equal to cardiac output. For individual at rest, the cardiac output and venous return are approximately 5 L/min. A complicated interaction of neural, humoral and physical factors determines the flow rate. Factors affecting venous return (Fig. 4.3-11) are:

1. Respiratory pump. Normally the intrapleural (intrathoracic) pressure at the end of expiration is about –2 mm Hg. During inspiration, the intrathoracic pressure becomes more negative (about –5 mm Hg) due to which the diameter of inferior vena cava is increased and pressure inside it is reduced; and there occurs descent of diaphragm which increases the intra-abdominal pressure. The decreased pressure inside the inferior vena cava coupled with increased intra-abdominal pressure during inspiration results in the increased flow of blood into the right atrium. This mechanism of increased blood flow during inspiration is called *respiratory pump.* This respiratory pump operates strongly in the forced respiration.

2. Cardiac pump. The cardiac pump influences the venous return by two kinds of forces the 'vis-a-tergo' and 'vis-a-fronte'.

- *Vis-a-tergo* refers to the forward push from behind, i.e. the propelling force which pushes the blood from veins into the right atrium. Vis-a-tergo results from the myocardial contraction during systole and is supplemented by the elastic recoil of the arterial wall (windkessel effect).
- *Vis-a-fronte* refers to the suction force acting from the front which basically pulls the blood from the great veins into the right atrium. This suction force is created by a ventricular contraction and has the following two components:
  - *Ventricular systolic suction* results from pulling down of the fibrous AV ring causing enlargement of the atrial lumen and thus decreasing the intra-atrial pressure which sucks blood from the inferior vena cava and the superior vena cava.
  - *Ventricular diastolic suction* results from the opening of AV valves allowing rapid flow of blood from the atria to ventricles. The sudden decrease in the atrial pressure in turn sucks blood from the great veins.

3. Muscle pump. The muscle pump mechanism is responsible for flow of blood from the veins of the limbs to the heart.

Working of muscle pump. Two types of veins are present in the limbs: superficial and deep veins.
Blood flows from the superficial veins into the deep veins through communicating veins. Due to the presence of valves in the limb veins, the blood flows in one direction, i.e. from periphery towards heart and not in the reverse direction.

- *When the skeletal muscles contract,* the deep veins present in between the muscles are compressed and due to increased pressure the valve present proximal to the contracting muscle is opened up while the valve present on distal end is tightly closed and in this way the blood is propelled up towards the heart (Fig. 4.3-12A).

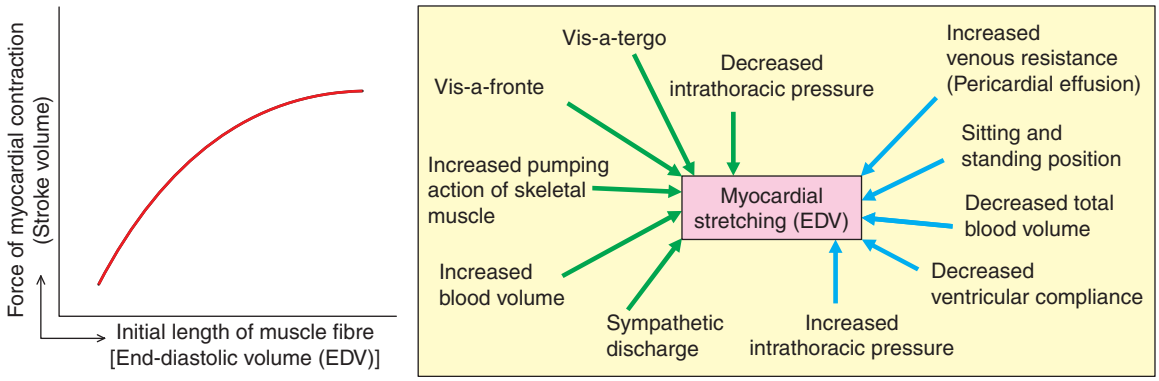


Fig. 4.3-11 Frank–Starling curve and factors affecting end-diastolic volume. Green arrows indicate increase and blue arrows indicate decrease.

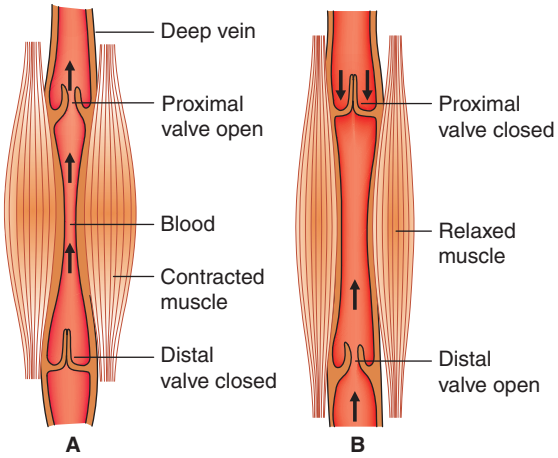


Fig. 4.3-12 Mechanism of muscle pump: A, during contraction of muscles and B, during relaxation of muscles.

- *When the skeletal muscle relaxes,* a negative pressure is created in the segment of veins. So due to back flow the proximal valve is closed and the distal valve is opened and blood is sucked up (Fig. 4.3-12B).
- *With rhythmic contractions of skeletal muscles* in this way, the blood is squeezed out of the limbs towards the heart.

APPLIED ASPECTS

In certain professions (e.g. nurses, traffic police, etc.) the individuals have to keep standing for a long time of a day. In such persons, sometimes the excessive venous pressure stretches the veins of the legs to such an extent that their diameter increases and the venous valves become incompetent. The venous pressure further increases and gradually the veins of lower limbs become large, tortuous and bulbous. This condition is called *varicose veins.*

4. Blood volume. The increased blood volume increases the venous return and a decreased blood volume decreases the venous return.

5. Sympathetic discharge. On sympathetic stimulation there occurs increase in the venous tone which decreases the capacity of the venous system (veins are capacitance vessels).

6. Standing body position is associated with a decreased venous return due to peripheral pooling of the blood.

Clinical significance of Frank–Starling mechanism

For small, momentary adjustments necessary for keeping the outputs of two ventricles equal and adjusting them to match the venous return the intrinsic regulation works continuously. The accuracy with which this adjustment is made can be understood by considering what would happen if the right ventricular output exceeds the left ventricular output by as little as 0.1 mL/beat, i.e. 7 mL/min. Then in a period of 3 h, the pulmonary blood volume will be increased by more than 1 L ($7 \times 60 \times 3 = 1260$ mL). This will prevent the optimal exchange of gases across the lungs and result in a severe pulmonary insufficiency.

Maintenance of constant stroke volume when the peripheral resistance is increased is carried out by intrinsic mechanism. When the peripheral resistance (blood pressure) is increased, initially the heart is unable to pump all the blood it normally does. The accumulated blood in the ventricle stretches the muscle fibres leading to great force of contraction and thus the stroke volume is restored to normal in spite of greater resistance to the outflow.

Intrinsic control mechanism serves as a life-saving device in a cardiac failure. Left ventricular failure causes accumulation of blood within the left ventricle, thereby decreases blood supply to the vital organs. Soon, accumulation of blood in the left ventricle, increases the initial length of muscle fibres leading to greater cardiac output according to Frank–Starling mechanism. However, when accumulation of blood is too great, the Frank–Starling law will fail to operate leading to decrease in the blood supply to the vital organs and ultimately death may occur.

2. Extrinsic regulation (autonomic neural mechanism)

In this mechanism stroke volume increases due to *increased myocardial contractility* without any increase in the initial muscle length. Therefore, it is also called *homometric mechanism.* The homometric regulation is governed by the autonomic neural mechanism as:

Sympathetic activity

Stimulation of the sympathetic nerves to the heart results in increased myocardial contractility and is known as *positive inotropic effect.* The positive inotropic effect of the norepinephrine liberated at the sympathetic nerve endings is augmented by the circulating norepinephrine.

Positive inotropic effect also can be defined as an increase in the *maximal velocity* of shortening (V'max) when it is plotted as a function of after-load. For the ventricles the afterload is the arterial pressure during the ejection phase. During any muscle contraction, the velocity of shortening and the force developed are inversely related. For details of the effect of afterload and the *force–velocity curve* see page 183.

Inhibition of sympathetic system has opposite effects. Under normal conditions, there is continuous slow rate of discharge through sympathetic fibres to the heart which maintains pumping 30% above with no sympathetic stimulation. Therefore, when sympathetic activity is inhibited the ventricular force of contraction decreases.

In intact animals, stimulation of sympathetic nerves produces a marked increase in the heart rate and a moderate increase in the stroke volume leading to a manifold increase in the cardiac output.

Characteristics of increased myocardial contractility

Characteristics of homometric regulation. The increased myocardial contractility achieved by homometric regulation differs from the increase in force of contraction of myocardium achieved by heterometric regulation. Its characteristic features are:

- The ventricles contract more forcefully and more rapidly, i.e. velocity of shortening of muscle is increased. As a result the ventricles are able to do more work per stroke, i.e. ejection fraction increases at the same end-diastolic volume (without increase in venous return).
- Due to more complete emptying of ventricles during each systole the *end-systolic volume is decreased.*
- Due to increased stroke output, *arterial pressure is increased.*

Mechanism of effects of sympathetic stimulation. Sympathetic stimulation increases the myocardial contractility by causing *activation of* β_1 adrenergic receptors, which in turn:

- Increases the concentration of Ca^{2+} within the myocardial cells causing a more rapid and forceful contraction.
- Via protein kinase causes more rapid intake of Ca^{2+} by the sarcoplasmic reticulum, which shortens the duration of both the action potential and contraction.

Parasympathetic activity

There is a negative inotropic effect of parasympathetic (vagal) stimulation. This effect, however, is not much because vagal fibres are mainly distributed to the atria and not much to the ventricles.

ROLE OF HEART RATE IN CONTROL OF CARDIAC OUTPUT

The cardiac output and heart rate both are increased during exercise proportionate to its severity. Since, cardiac output is the product of stroke volume and heart rate, it is tempting to attribute the increase in cardiac output to increase in the heart rate. However, in fact it is not so. It has been seen that when heart rate alone is increased, e.g. by change in the frequency of discharge of an artificial cardiac pacemaker, the cardiac output does not increase at all. As shown in Fig. 4.3-13, the progressive increase in the heart rate is associated with a proportionate decrease in the stroke volume because reduced diastole time and thus reduced end-diastolic volume. Conversely, when the heart rate is reduced the ventricular diastole is prolonged leading to more ventricular filling and thus an increased stroke volume.

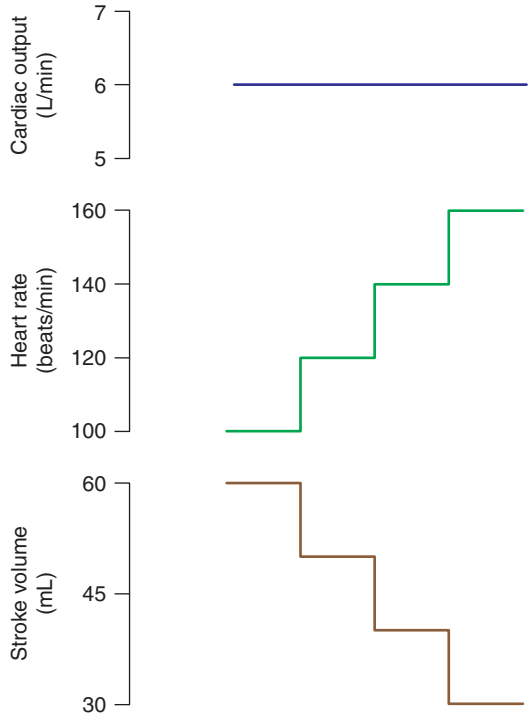


Fig. 4.3-13 Effect of increase in heart rate through an artificial pacemaker on stroke volume and cardiac output.

During exercise, the sympathetic stimulation produces a marked increase in the heart rate (200–300%) due to positive chronotropism and moderate increase (50–60%) in the stroke volume due to positive inotropism leading to manifold increase in the cardiac output.

INTEGRATED CONTROL OF CARDIAC OUTPUT

In intact animals and humans, the intrinsic and extrinsic mechanisms described above operate simultaneously in an integrated way to maintain cardiac output. Therefore, in a given situation, depending upon the status of end-diastolic volume and status of myocardial contractility, the individual will have one of the curves from the 'family of Frank–Starling curves'.

Interaction of Frank–Starling mechanism and myocardial contractility

Factors causing increased myocardial contractility shift the Frank–Starling curve to the left (Fig. 4.3-14). Increased contractility (positive inotropism) refers to the greater contraction force at a given preload or end-diastolic volume. Factors increasing myocardial contractility are:

1. *Sympathetic stimulation increases* myocardial contractility by causing activation of β_1 adrenergic receptors, as discussed above.

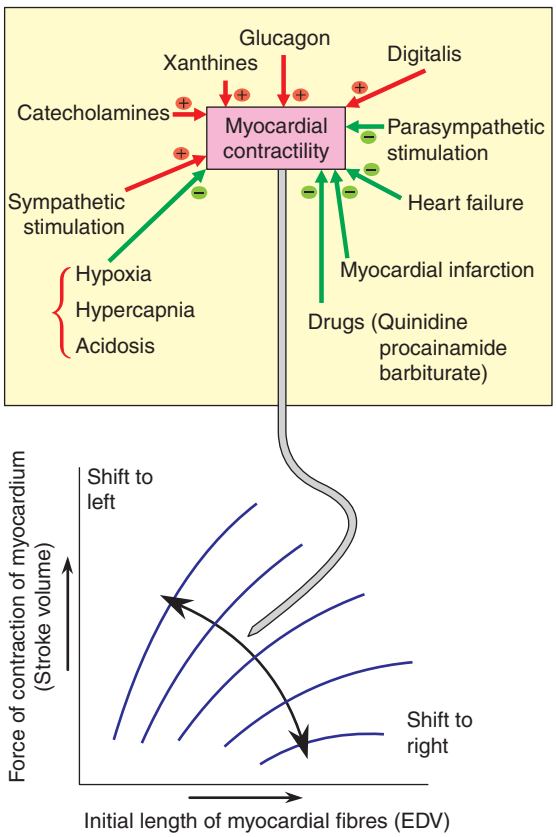


Fig. 4.3-14 Effect of changes in myocardial contractility on the Frank–Starling curve. The factors which increase the contractility shift the curve to left and those decrease the contractility shift the curve to right.

2. *Catecholamines* also exert their positive inotropic effect via their action on cardiac β_1 adrenergic receptors by a mechanism similar to that of sympathetic stimulation.
3. *Xanthines,* such as caffeine and theophylline exert their positive inotropic effect by inhibiting the breakdown of cyclic AMP.
4. *Glucagon* causes positive inotropic effect by increasing the formation of cyclic AMP.
5. *Digitalis and related drugs exert* their positive inotropic effect by their inhibitory effect on Na^+–K^+ ATPase in the myocardium. The inhibition causes an increase in intracellular Na^+, which in turn increases the availability of Ca^{2+} in the cell. Digitalis which was initially prepared from the plant *Digitalis purpurea* has been used for centuries to treat heart failure.

Factors causing decreased myocardial contractility shift the Frank–Starling curve to the right (Fig. 4.3-14). The decreased contractility (negative inotropism) represents a decrease in the force of contraction at any fibre length or ventricular volume. Its causes are:

1. *Parasympathetic (vagal) stimulation.* Since the vagal fibres are distributed mainly to atria and not to ventricles so vagal stimulation causes negative inotropic effect on the atrial muscles and indirectly mild negative inotropic effect on the ventricles reducing strength of heart contraction by 20–30%.
2. *Heart failure* is also associated with reduced myocardial contractility due to intrinsic depression. The cause of this depression is not known.
3. *Myocardial infarction* may result in a fibrotic and nonfunctional area in myocardium resulting in reduction of total ventricular performance.
4. *Hypercapnia,* hypoxia and acidosis produce negative inotropic effect by causing a decrease in the formation of cyclic AMP.
5. *Drugs* such as quinidine, procainamide and barbiturates depress myocardial contractility.

HEART–LUNG PREPARATION

The Frank–Starling's heart–lung preparation is an experimental set up in a dog, devised to demonstrate the effects of various factors on the activities of heart. In this preparation, as the name suggests, blood does not flow to any part of the body except the heart and the lungs. The animal is actually dead and heart is functionally denervated.

Experimental set up

Experimental set up (Fig. 4.3-15) of the heart–lung preparation includes following essential steps:

Trachea is cannulated and lungs are artificially ventilated.

Aorta is ligated beyond the origin of innominate artery so that systemic circulation of the body is blocked.

Innominate artery is cannulated and connected to mercury manometer to measure the arterial blood pressure and also through a series of tubes to:

- *Elastic vessel,* which provides elasticity artificially similar to that of arterial wall,
- *Resistance vessel,* which is used to provide resistance artificially. The resistance applied can be measured through the attached manometer.
- *Warming glass coil,* which is kept inside a water bath with a heater. The temperature of the water bath is controlled, so that the temperature of blood can be maintained.
- *Flowmeter* which determines the amount of blood flowing through it (cardiac output).
- *Venous reservoir* which represents the peripheral venous system. It is connected through a tube to the superior vena cava. A screw-type clamp is fitted to the rubber tube. It is used to adjust the amount of blood returning to heart (venous return).

Thus, the blood ejected from the left ventricle after passing through the above attachment ultimately reaches

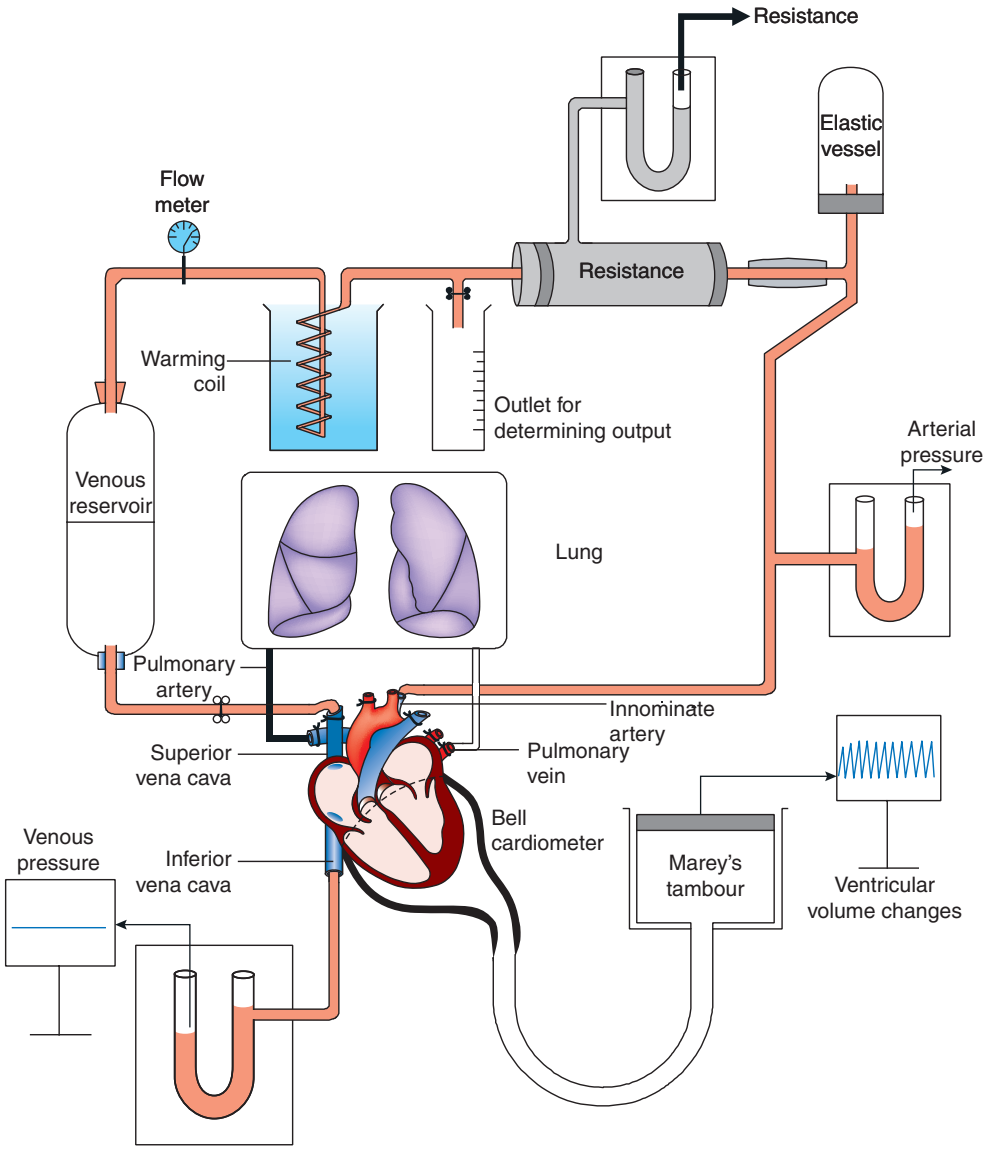


Fig. 4.3-15 Heart-lung preparation.

the right atrium. From there, the blood flows to the right ventricle, pulmonary artery, lungs and back to heart through the pulmonary veins.

Inferior vena cava is attached to manometer to record the right atrial pressure.

Bell's cardiometer is fitted to the ventricle for measuring the stroke volume of the heart. The recording is made through *Marey's tambour* connected to the cardiometer.

Uses of heart–lung preparation

In the heart–lung preparation, heart works as an isolated organ. It can be used to demonstrate the effect of various factors on the heart's activities as:

1. Effect of venous return on stroke volume (Frank–Starling mechanism). Venous return to the heart is changed through the venous reservoir and stroke volume at different levels of venous return is recorded through the cardiometer. The record shows:

- An increase in the stroke volume with increase in venous return (Fig. 4.3-16) and
- A decrease in stroke volume with decrease in the venous return.

These observations demonstrate the Frank–Starling's law (intrinsic mechanism controlling stroke volume).

2. Effect of sympathetic stimulation on stroke volume when the stellate ganglion (cardiosympathetic nerve) is stimulated without any change in the venous return, the stroke volume is increased but with lower end-systolic and end-diastolic volume (Fig. 4.3-17). This activity demonstrates the extrinsic control mechanism of the stroke volume. Right atrial pressure is also reduced. It is due to the increased

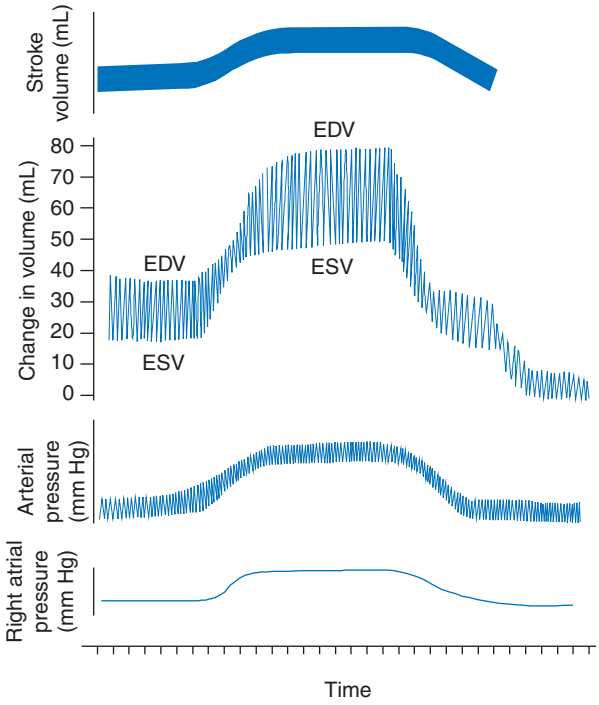


Fig. 4.3-16 Tracings from heart-lung preparation demonstrating the effect of venous return on stroke volume, end-diastolic volume, end-systolic volume, arterial pressure and venous pressure.

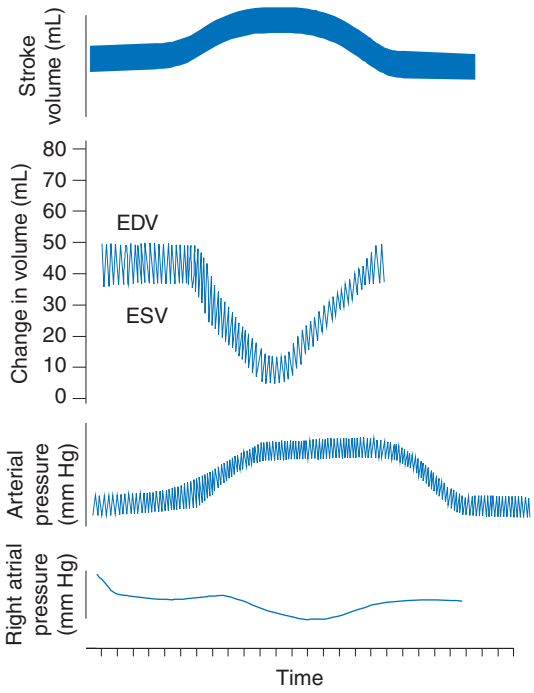


Fig. 4.3-17 Tracings from heart-lung preparation demonstrating the effect of sympathetic nerve stimulation on stroke volume, end-diastolic volume, end-systolic volume, arterial presure and venous pressure.

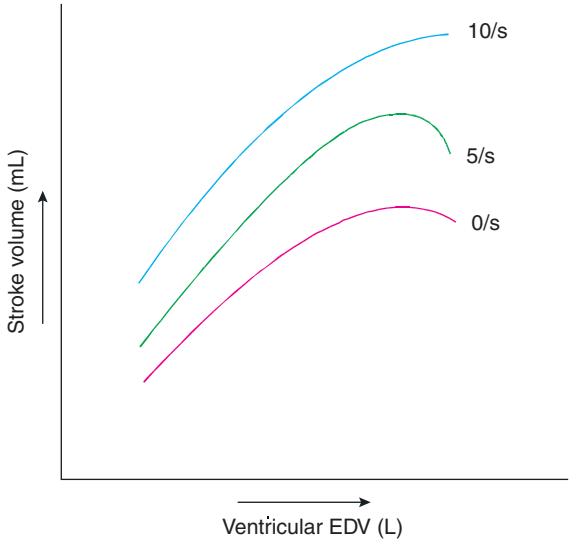


Fig. 4.3-18 Family of Frank–Starling curves obtained by combined effect of increase in end-diastolic volume and sympathetic stimulation on stroke volume on heart-lung preparation.

suction of blood from the atria by the vigorously contracting ventricle.

3. Combined effect of increase in end-diastolic volume and sympathetic stimulation on stroke volume. At a given rate of sympathetic stimulation, the end-diastolic volume is gradually increased and the resulting stroke volume is plotted against the end-diastolic volume, to give a Frank–Starling curve with different rates of sympa thetic stimulation (0/S–10/S) a family of such curves is obtained (Fig. 4.3-18). From these curves, it can be inferred that even during sympathetic stimulation an increase in end-diastolic volume increases the stroke volume of heart.

4. To demonstrate the effect of peripheral resistance on cardiac output. Resistance is increased through the resistance vessel and its effect on cardiac output is recorded:

- With increase in the peripheral resistance, cardiac output is increased and
- With decrease in the peripheral resistance, the cardiac output is decreased.

5. Cardiac function curves can also be recorded using a heart–lung preparation. These include:

- The cardiac output curves,
- The venous return curves and
- The cardiovascular curves.

Chapter 4.4

Dynamics of Circulation: Pressure and Flow of Blood and Lymph

INTRODUCTION

Dynamics of circulation is concerned with flow of blood and lymph and also the pressure in the various segments of the vascular system of the body. For descriptive purposes, the 'dynamics of circulation' is discussed under following headings:

- Functional organization and structure of vascular system,
- Haemodynamics,
- Blood flow and pressure in different segments of circulatory system and
- Blood pressure.

FUNCTIONAL ORGANIZATION AND STRUCTURE OF VASCULAR SYSTEM

ORGANIZATION OF VASCULAR SYSTEM

The vascular system is organized into two separate circulations, systemic and pulmonary, arranged in series. In addition, parallel to the circulation of blood is disposed circulation of lymph which helps the blood circulation to perform its various functions.

Systemic circulation supplies blood to various systemic organs through parallel distribution channels (Fig. 4.4-1). This parallel arrangement of vessels ensures the supply of blood of the same arterial composition (i.e. same O_2 and CO_2 tension, pH, glucose level and essentially the same arterial pressure) to the various body organs. In systemic circulation, from the left ventricle, blood is pumped through the arteries and arterioles to the capillaries, where it equilibrates with the interstitial fluid. The capillaries drain through the venules into the veins and ultimately to the right atrium.

Pulmonary circulation is meant for oxygenation of blood. Since pulmonary circulation is arranged in a series with systemic circulation, so it receives the same amount of blood over any significant time period. In pulmonary circulation, from the right ventricle, blood is pumped through the pulmonary arteries to the pulmonary capillaries. In the pulmonary capillaries, the blood equilibrates with the O_2 and CO_2 of alveolar air. The capillaries then drain the oxygenated

blood through venules and then through pulmonary veins into the left atrium.

Lymphatic circulation, which is disposed in parallel to the circulation of blood can be considered a third type of circulation. Some tissue fluid enters the lymphatic channels as lymph which is ultimately drained into the venous system via the thoracic lymphatic duct and the right lymphatic duct.

Systemic vascular tree

For descriptive purposes and from functional point of view, the systemic vascular tree can be divided into following types of blood vessels:

- *Large elastic arteries* (windkessel vessels) include aorta and its main branches, such as carotid, iliac and axillary arteries.
- *Large muscular* arteries (distribution vessels) which include most of the arteries of the body, e.g. arteries like radial, ulnar, popliteal.
- Arterioles and precapillary sphincters (resistance vessels).

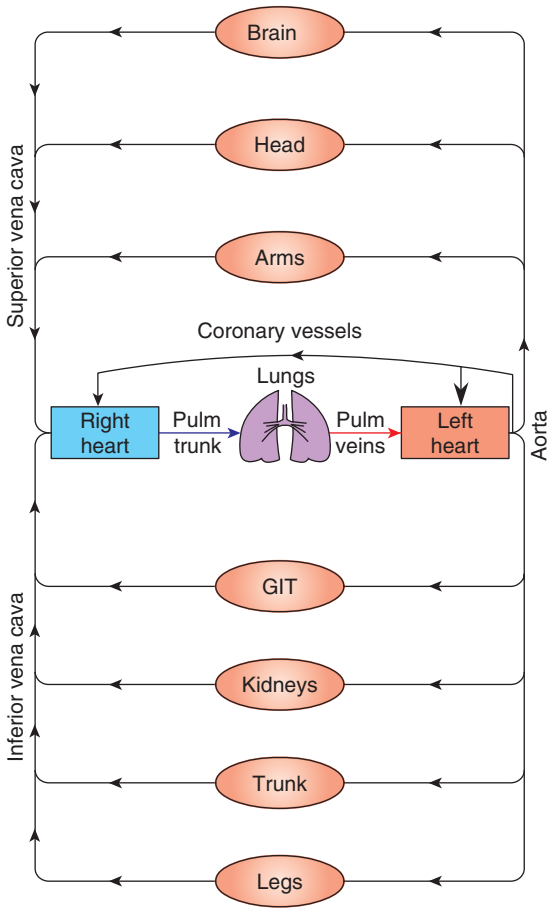


Fig. 4.4-1 A schematic illustration of the organization of cardiovascular system depicting a series arrangement of pulmonary and systemic circulation and parallel arrangement of vessels supplying blood to the organs.

- Meta-arterioles and capillaries (exchange vessels).
- Venules (post-capillary resistance vessels),
- Veins (capacitance vessels) and
- Arteriovenous anastomoses (shunt or thoroughfare vessels).

STRUCTURE OF BLOOD VESSELS

STRUCTURAL CHARACTERISTICS

General structural characteristics

Histologically, walls of most of the blood vessels except the capillaries consist of three coats. General structural characteristics of a large artery are (Fig. 4.4-2):

1. Tunica intima. It is the innermost coat of the vessel wall. In large arteries, from inside-out, it consists of:

- *Endothelial lining,* which is very smooth and silky, and consists of a single layer of cells. It lies in contact with the blood.
- *Basal lamina* is a thin layer of glycoprotein which lines the external aspect of the endothelium.
- *Subendothelial connective tissue* is a delicate layer of connective tissue which lies outside the basal lamina.
- *Internal elastic lamina* is a thin membrane formed by the elastic fibres.

2. Tunica media. It is the middle, thickest coat of the vessel wall. It consists of smooth muscles and elastic tissue. The ratio of these two tissues varies from vessel to vessel. On the outside, tunica media is limited by a membrane formed by the elastic fibres called the *external elastic lamina.*

3. Tunica adventitia. It is the outermost coat of the vessel wall. It is made of connective tissue in which collagen fibres are prominent. This layer prevents undue stretching or distension of the blood vessel.

Specific structural characteristics

Large (elastic) arteries in their tunica media have dominant elastic tissue which provides them property of distensibility and elastic recoil.

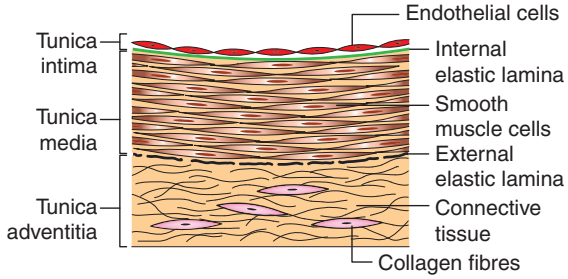


Fig. 4.4-2 Histological structure of an artery.

Medium size arteries. In these arteries, the elastic tissue both in intima and media is much less and thus proportion of smooth muscles increases.

Arterioles have characteristically no elastic tissue. Their media consists of a thick layer of smooth muscles and they have a *relatively narrow lumen.* Because of their structural characteristics they act as resistance vessels.

Meta-arterioles are relatively high-resistance conduits between arterioles and veins.

Capillaries arise directly from the arterioles or meta-arterioles. A cuff of smooth muscle cells called the *precapillary sphincter* surround the origin of capillaries in some region. The capillary wall does not contain any tunica media and adventitia. It is formed essentially by the endothelial cells which are lined on the outside by a basal lamina (glycoproteins), branching perivascular cells called *pericytes.*

Postcapillary venules, which measure 20–60 μm in diameter, are the *most permeable part* of microcirculation.

Veins are relatively thin-walled structures that contain very small amount of elastic tissue and smooth muscles as compared to arteries. These are highly distensible part of the vascular system and that is why also form the so-called *reservoir vessels.*

Essential characteristics

Essential characteristics of blood vessels like lumen diameter, wall thickness, approximate total cross-sectional area and percentage of blood volume contained are shown in Table 4.4-1.

Table 4.4-1 Essential characteristics of various types of blood vessels

Vessel	Lumen diameter	Wall thickness	All vessels of each type total approximate cross-sectional area (cm^2)	Percentage of blood volume contained
Aorta	2.5 cm	02 mm	4.5	2
Artery	0.4 cm	01 mm	20	8
Arteriole	30 μm	20 μm	400	1
Capillary	5 μm	1 μm	4500	5
Venule	20 μm	2 μm	4000	54
Vein	0.5 cm	0.5 mm	40	
Vena cava	3 cm	1.5 mm	18	
Heart	–	–	–	12
Pulmonary, circulation	–	–	–	18

Innervational characteristics

- Smooth muscles of the blood vessels are innervated by the sympathetic fibres. These muscles contain α adrenergic receptors. Therefore, noradrenaline causes contraction of muscle fibres leading to vasoconstriction.
- Sympathetic fibres exert tonic effect even at rest, resulting in the existence of vasomotor tone in the blood vessels.
- Stimulation of sympathetics increases the vasomotor tone, as a result the vessels are constricted and narrowed.
- Inhibition of sympathetic discharge results in a decreased vasomotor tone and hence vasodilatation.
- Skeletal muscle arterioles also contain β_2 receptors in addition to the α receptors therefore, adrenaline causes dilatation of these vessels.
- Besides α adrenergic receptors, smooth muscles of the blood vessels are also stimulated by other agents like O_2 tension, lactic acid, etc.

HAEMODYNAMICS

Haemodynamics, which refers to the study of blood flow in various segments of the vascular system, can be discussed under following headings:

- General principles governing (factors affecting) blood flow,
- Types of blood flow,
- Measurement of blood flow,
- Distribution of blood flow to various regions of the body and
- Regulation of blood flow in different situations.

GENERAL PRINCIPLES GOVERNING (FACTORS AFFECTING) BLOOD FLOW

FLOW–PRESSURE–RESISTANCE RELATIONSHIP

Relationship between the flow of a fluid with the pressure and resistance offered to it through a rigid tube was studied by a French physiologist Poiseuille (in 1842) and Hagen. This relationship is known as Poiseuille's law or Poiseuille–Hagen law.

POISEUILLE'S LAW

Poiseuille's law expressing the relation between the flow (Q) and pressure gradient (ΔP) in a long narrow tube of length (L), the viscosity of fluid (η) and the radius (r) of the tube is as:

$$Q = \frac{\Delta P \pi r^4}{8 \eta 1}$$

Thus, according to the Poiseuille's law, the flow (Q) of a Newtonian fluid through a rigid tube is determined by:

1. Pressure gradient (ΔP), i.e. difference in the pressure between the two ends of the tube. In other words, fluid always flows from an area of high pressure (P_1) to one of

lower pressure (P_2), and rate of flow (Q) is determined by the pressure gradient ($P_1 - P_2$), i.e. $Q \propto \Delta P$ or ($P_1 - P_2$).

2. Radius of tube. The flow of fluid varies directly as the fourth power of radius (r^4). Thus if the radius is halved, the flow will decrease by 16 times and vice versa. Thus this factor is very important for the flow of blood through the blood vessels.

3. Viscosity of fluid (η). The flow of fluid varies inversely with the viscosity of fluid, i.e. greater the viscosity, lesser the flow and vice versa.

4. Length of the tube (L). The flow is inversely proportional to the length of the tube. This is easily understandable, as every segment of the tube is offering resistance to the flow; therefore, longer the length greater will be the total resistance offered.

Resistance (R). According to mathematical calculation in principles of physics, resistance (R) is represented by $8\,Ln/\pi r^4$. By replacing $8\,Ln/\pi r^4$ with R in the Poiseuille's law, it becomes:

$$Q = \frac{\Delta P}{R}$$

Thus, the Poiseuille's law can be considered analogous to the Ohm's law of current in which:

$$\text{Current (flow)} = \frac{\text{Voltage (Pressure gradient)}}{\text{Resistance}}$$

Hence, the rate of flow (Q) is inversely proportional to the resistance (R).

The Poiseuille's law is valid for straight rigid tubes with Newtonian fluid flowing through them. Since, the blood vessels are not rigid and the blood is not a Newtonian fluid; therefore, strictly speaking the Poiseuille's law does not apply to flow of blood through the vascular system. Nevertheless, the important principles relating flow, pressure gradient and resistance remain applicable, so they are discussed in relation to blood flow as:

- Blood flow and pressure gradient relationship, and
- Blood flow and resistance relationship.

BLOOD FLOW AND PRESSURE GRADIENT RELATIONSHIP

According to the Poiseuille's law, fluid always moves from an area of higher pressure to one of lower pressure. This downhill movement of blood (i.e. along the pressure gradient) occurs in the vascular system. In the systemic circulation, pressure at the beginning of aorta is about 100 mm Hg and near the terminal portion of inferior vena cava it is nearly zero (Fig. 4.4-3). Therefore, $P_1 - P_2 = (100 - 0) = 100$ mm Hg. Note the progressive decrease in pressure from the left ventricle through the systemic circulation until the blood enters the right ventricle. It is important to note that greatest pressure drop occurs in the arterioles, which represent the highest resistance segment of the systemic circulation. Further, according to the Poiseuille's law, at constant length and radius of a tube and viscosity of the fluid, the relationship between the pressure and flow through a rigid tube is linear, i.e. as pressure increases, flow of fluid also increases (Fig. 4.4-4A). The relationship between the flow of blood and pressure in the blood vessels is however not linear because blood vessels are distensible (e.g. aorta) (Fig. 4.4-4B) and show active myogenic contractile response to stretch (e.g. arterioles).

This myogenic contractile response to stretch offsets the elastic effect and so the flow in these vessels becomes less than the rigid tubes, i.e. the curve becomes concave to the pressure axis (Fig. 4.4-4C).

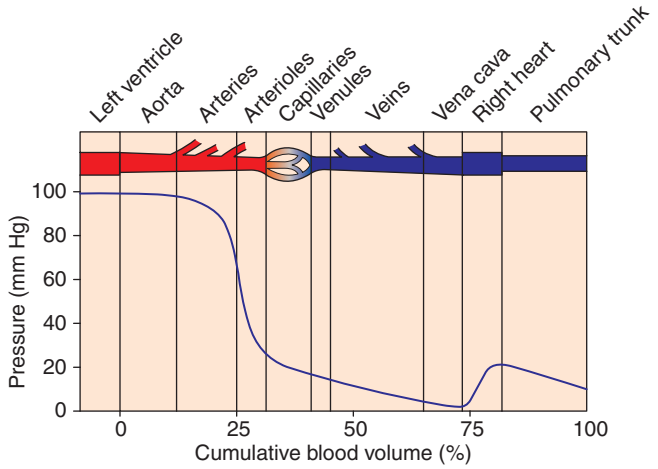


Fig. 4.4-3 Mean lateral pressure in various components of vascular system and cumulative blood volume.

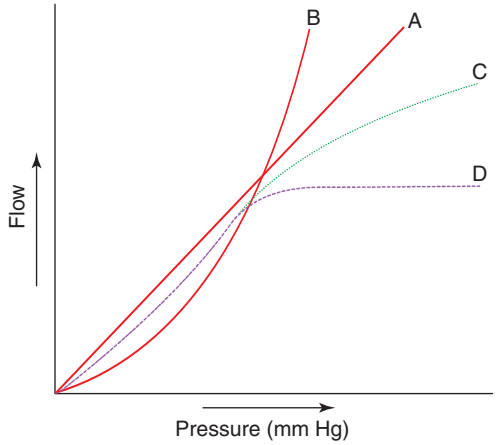


Fig. 4.4-4 Relationship between pressure and flow: A, in a rigid tube; B, in a distensible blood vessel (aorta); C, in a distensible blood vessel containing active myogenic contractile element, whose contraction affects the distensible effects of raised pressure and D, in a distensible vessel containing myogenic contractile element, which serves to stabilize flow over a wide range of pressure (80–200 mm Hg), such vessels would show autoregulation.

In the arterioles where the myogenic contractile response even exceeds the elastic effect of raised pressure, the blood flow is maintained and does not increase with the further increase in the pressure (Fig. 4.4-4D). Such arterioles serve to stabilize the blood flow over a wide range of pressure (80–200 mm Hg). This phenomenon is called *autoregulation.*

Critical closing pressure

Since, the flow–pressure relationship in a rigid tube is linear, so flow will cease only if the pressure is zero (Fig. 4.4-5A). However, in a blood vessel the flow ceases when the blood pressure is 20 mm Hg or even more (Fig. 4.4-5B). The pressure value at which the vessel collapses, its lumen closes and flow ceases is called *critical closing pressure (CCP).* The blood flow ceases when the blood pressure falls below the CCP because:

- Certain amount of intramural pressure is essentially required to push the RBCs (with average diameter of 7.5 mm) through the capillaries (average diameter 5 µm).
- Further, the tissue pressure exerted over the vessels also causes their collapse. So, certain amount of intramural pressure is must to counteract the tissue pressure effect to keep the vessels patent for the blood to flow.

Values of critical closing pressure

- When whole blood is flowing through the vessels the average value of critical closing pressure (CCP) is 20 mm Hg (Fig. 4.4-5B).
- When plasma is flowing the value of CCP is about 5–10 mm Hg.
- On sympathetic stimulation, value of CCP increase to 60 mm Hg (Fig. 4.4-5C)
- On sympathetic inhibition, value of CCP falls to 0 mm Hg.

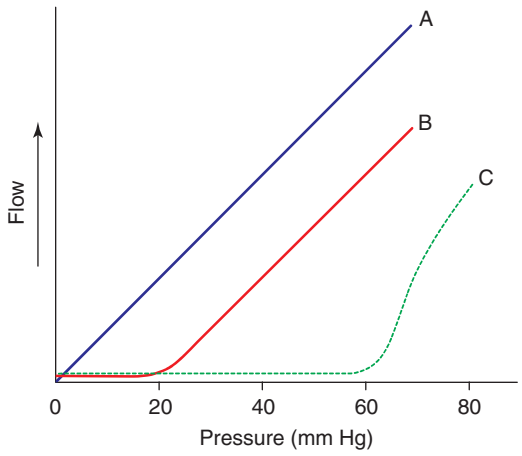


Fig. 4.4-5 Relationship between pressure and flow in a rigid tube A, in a blood vessel; B, and on sympathetic stimulation; C, depicting critical closing pressure (CCP).

Equilibrium of factors at critical closing pressure

- In general, vasomotor tone of the blood vessels tries to constrict the vessels. This tone is increased by the sympathetic stimulation and decreased by the sympathetic inhibition.
- Intramural pressure tries to dilate the blood vessels up to a certain limit. Then a point is reached where the intramural pressure (P) is not able to maintain equilibrium with tension (T) in the vessel wall. At this point, the blood vessel closes and the pressure at which it occurs is called critical closing pressure. Laplace law described below helps to explain this equilibrium relationship.

Law of Laplace

Law of Laplace governs the relationship between the distending pressure and tension in the wall of a distensible viscus including blood vessels (Fig. 4.4-6). According to this law, the distending pressure (P) in a distensible hollow object is equal at equilibrium to the tension in wall (T) divided by the two principal radii of curvature of the object (r_1 and r_2).

$$P = T\left(\frac{1}{r_1} + \frac{1}{r_2}\right), \text{ where}$$

T is expressed in dynes/cm,
r_1 and r_2 are in cm and so
P is expressed in dynes/cm^2

- In a sphere, $r_1 = r_2$. Therefore,

$$P = \frac{T}{r}$$

- *In a cylinder* such as a blood vessel, one radius is infinite. Therefore,

$$P = \frac{T}{r}$$

Physiological applications of law of Laplace

This law applies to all the hollow viscous structures in the body. Some of its important applications are:

In vascular system. As described above, for the blood vessels Laplace equation is P = T/r. This equation shows that

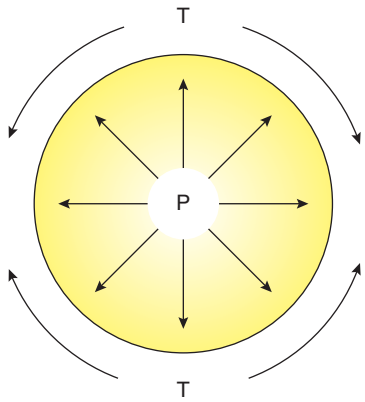


Fig. 4.4-6 Relationship between distending pressure (P) and wall tension (T) in a hollow viscus.

smaller is the radius of a blood vessel, lesser is the tension (T) on the walls of the blood vessel required to balance the distending pressure or force (P). For example,

- In aorta, the tension at normal pressure is 170,000 dynes/cm;
- In inferior vena cava, it is about 21,000 dynes/cm;
- But in the capillaries, it is approximately 16 dynes/cm. This explains, why capillaries being so thin walled and delicate are not prone to rupture.

In heart. *Law of Laplace also explains* the disadvantage faced by a dilated heart. When the radius of a cardiac chamber is increased, a greater tension must be developed in the myocardium to produce any given pressure; consequently, a dilated heart must do more work than a non-dilated heart.

FLOW AND RESISTANCE RELATIONSHIP

From the Poiseuille's law, it can be derived that the flow is inversely proportional to the resistance, i.e. $Q \propto 1/R$; and that resistance is represented by $8L\eta/\pi r^4$, i.e. resistance depends upon the length (L) and fourth power of radius (r^4) of the tube and viscosity of the fluid (η). In vascular system, resistance to flow is represented by the total peripheral resistance and is expressed as peripheral resistance unit.

Peripheral resistance unit (PRU)

As discussed above, the Poiseuille's law can be considered analogous to the Ohm's law, so:

$$\text{Flow, i.e. } Q\ (\text{mL/s}) = \frac{P_1 - P_2\ (\text{mm Hg})}{\text{Resistance (R)}}$$

$$\text{Thus, } R = \frac{P_1 - P_2\ (\text{mm Hg})}{Q\ (\text{mL/s})}$$

Therefore, in vascular system, *PRU* is mm Hg/mL/s.

Total peripheral resistance (TPR)

At rest, the resistance (R) of the entire systemic circulation is called 'total peripheral resistance'. Its values are:

- At rest, TPR is 1 PRU (1 mm Hg/mL/s),
- TPR, during maximum vasoconstriction may increase to 4 PRU,
- TPR during maximum vasodilation may decrease to 0.2 PRU and
- Pulmonary vascular resistance is about 0.1–0.2 PRU.

Factors that affect resistance to blood flow

According to the Poiseuille's law, resistance (R) depends upon the length of tube (L), fourth power of radius of tube (r^4) and viscosity of the fluid (η). Since in the intact body length of the blood vessels does not change, i.e. remains constant, so the major factors that determine the resistance to flood flow are:

I. The viscosity of blood and
II. The radius of vessels.

I. Blood viscosity and resistance

Resistance (R) to blood flow is directly proportional to viscosity (η) of blood.

Definition and unit of viscosity

Viscosity was described by Isaac Newton in 1713 as an internal friction to flow in a fluid or lack of slipperiness. These terms emphasize that when fluid moves along a tube, laminae in the fluid slip on one another and move at different speeds thereby causing a velocity gradient in a direction perpendicular to the wall of the tube. Thus, resistance met by the fluid moving through a tube in a streamlined flow is due to the friction between adjacent laminae and not due to the friction between vessel wall and fluid. Thus, greater the internal friction, greater is difference in velocity (shear rate) between two laminae and greater the coefficient of viscosity.

Unit of viscosity is *Poise* (after Poiseuille). A fluid of 1 poise viscosity has a force of 1 dyne/cm^2 of contact between layers when flowing with a velocity gradient of 1 cm/s. One Poise is considered to be consisting of 100 centipoise (CP). *Viscosity of water* at 21°C is 0.01 poise or 1 centipoise.

Relative viscosity is a more often used term and refers to the viscosity of a fluid relative to the viscosity of water at body temperature (37°C).

- Water has a viscosity of 0.695 centipoise at body temperature.
- Plasma, which has a viscosity of 1.2 CP at 37°C, thus has a relative viscosity of 1.7.
- Blood (plasma plus cells) which has a viscosity of 2.8–3 CP at 37°C, thus has a relative viscosity of about 4–5.

Factors affecting blood viscosity

1. Shear rate or velocity gradient. Viscosity of the blood decreases as the shear rate or velocity gradient increases and vice versa.

- At high shear rate, the RBCs occupy the central axis of the tube and move with their long axis parallel to the direction of flow where flow rate is fastest leaving cell free zone of plasma at periphery. This process is called plasma skimming (Fig. 4.4-7A). This causes least friction between the cells and plasma and thus viscosity decreases. At high rates of flow, blood behaves almost like a Newtonian fluid with constant viscosity.

  Note. The phenomenon of plasma skimming is responsible for low value of haematocrit in capillary blood. The

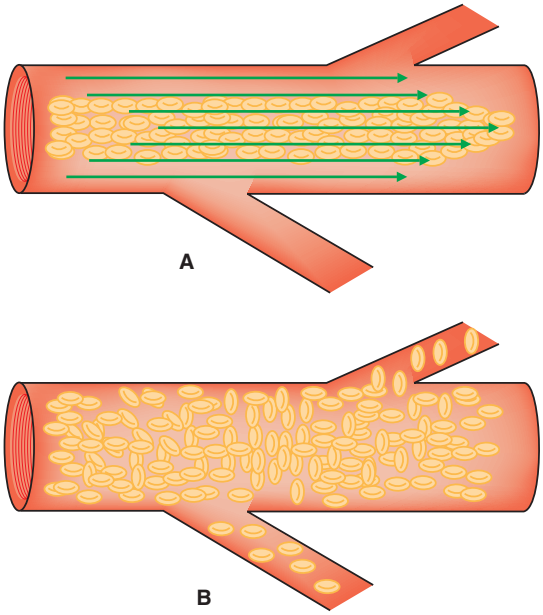


Fig. 4.4-7 Features of streamlined flow at a high flow rate (A) and at a slow flow rate (B).

haematocrit of capillary blood is about 25% lower than that of venous blood.

- *When shear rate is low* (at low flow rate) tendency of RBCs to occupy central axis decreases (Fig. 4.4-7B). The suspended RBCs in the plasma increase internal friction due to collision among these suspended particles. So, with slow shear rate the viscosity increases.

2. Haematocrit. In general, variation in the haematocrit is the major factor that changes the viscosity of blood. An increase in haematocrit tends to reduce flow rate because of increased viscosity.

3. Temperature. Cooling increases the viscosity of blood. In intact body though there is no variation in the body temperature normally, the cutaneous and subcutaneous vessels and even those of the deeper regions in the limbs are subjected to considerable alteration of temperature due to atmospheric temperature. Thus, when the hand is kept in ice water, regional viscosity of the blood shows a threefold increase.

Pathological conditions associated with the high blood viscosity are:

- Severe polycythaemia.
- Abnormal-shaped RBCs (congenital spherocytosis).
- Abnormally high immunoglobulin level.
- Lowered body temperature (in frost bite).

II. Radius of blood vessels and resistance

As mentioned earlier, the rate of flow is proportional to the fourth power of the radius (r^4) of the blood vessels. Thus, even a small change in the calibre of the blood vessels will produce a marked change in the blood flow. For example, at a pressure head of 100 mm Hg, the change in the flow rate with change in the radius:

Radius	*Flow rate*
1 mm	1 mL/min
2 mm	16 mL/min
4 mm	256 mL/min.

Among the blood vessels, in aorta and large arteries, there is little resistance. Arterioles are the major seat of vascular resistance.

VELOCITY OF BLOOD FLOW

The velocity of blood flow refers to the displacement of blood per unit time, i.e. cm/s while, the rate of blood flow is the amount of blood flowing per unit time, i.e. cm^3/s. The physiologically important aspects to be considered in relation to velocity of blood flow are:

- Velocity and cross-sectional area relationship and
- Velocity–pressure relationship.

VELOCITY AND CROSS-SECTIONAL AREA RELATIONSHIP

- The relationship between velocity (V), quantity of blood flow (Q) and cross-sectional area (A) of the blood vessel is:

$$V = \frac{Q}{A}$$

- Thus, if quantity of blood flow (Q) remains constant and the cross-sectional area (A) increases, then the velocity of blood flow will decrease.
- Since, the cross-sectional area of capillaries is 1000 times as that of the aorta, so the velocity of blood flow in the capillaries is approximately 1 mm/s as compared to 40 cm/s in the aorta.
- The total cross-sectional area of various types of blood vessels, the velocity of blood flow and pressure in the various segments of the cardiovascular system plotted as a function of the cumulative blood volume is shown in Fig. 4.4-8 and Table 4.4-1.

VELOCITY–PRESSURE RELATIONSHIP

To understand the effect of velocity of blood flow on the pressure, it is important to understand the concept of total energy (total pressure), potential energy and kinetic energy put forward by Burton in 1965.

Total energy (total pressure) of a fluid in a tube at any point equals the sum of the potential energy and kinetic energy.

Potential energy of a fluid in a tube comprises hydrostatic pressure and lateral (static) pressure.

1. *Hydrostatic pressure (Ph)* results from a difference in vertical height in a fluid-filled system. Because of gravity, fluid

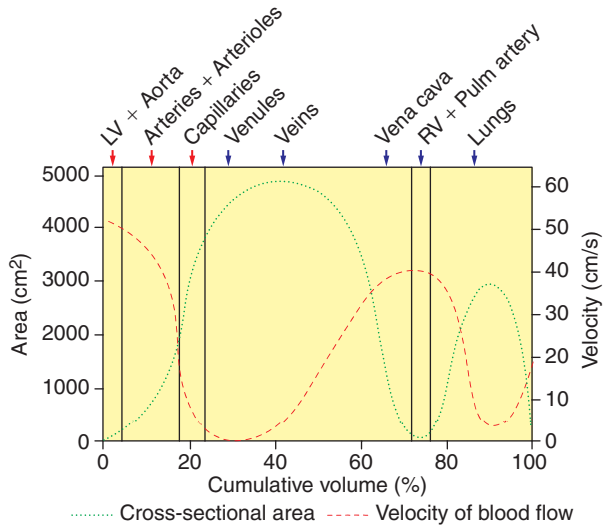


Fig. 4.4-8 Relationship between the cross-sectional area, velocity of blood flow and pressure as a function of cumulative blood volume.

has weight that generates force, which is proportional to its vertical height

$$Ph = \delta \cdot h \cdot g,$$

where δ = fluid density, h = height of fluid column above or below a reference level and g = gravitational constant.

Some facts about the effects of hydrostatic pressure on human vascular system are:

- The zero reference level is at the right atrium, so vascular segments higher than the reference point will have negative gravity effect and vascular segments below reference point will have positive gravitational effect.
- Both arteries and veins at any given horizontal level are affected by the same hydrostatic pressure of blood so that the pressure gradient between arteries and veins is same (about 100 mm Hg), i.e. it is not altered (Fig. 4.4-9).
- As shown in Fig. 4.4-9, in an upright person, assuming that foot is 100 cm below the heart, the pressure in the vessels (arteries as well as veins) of foot is 100 cm H_2O (77 mm Hg) higher than the pressure at the root of aorta.
- In a supine position, the hydrostatic effect is eliminated because the entire cardiovascular system is essentially at the same horizontal level.

2. *Lateral (static) pressure* represents the pressure in the cardiovascular system that is usually measured with a strain gauge or transducer after eliminating the hydrostatic pressure effect. To eliminate the hydrostatic pressure effect, it is essential to place the gauge or the sphygmomanometer cuff at the *zero reference* (phlebostatic) level, which is equivalent to the level of right atrium.

Kinetic energy is the momentum that blood gains because of its mass and velocity:

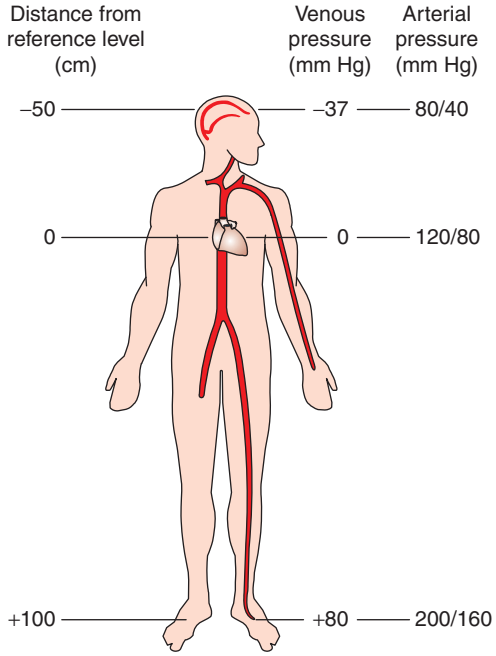


Fig. 4.4-9 Effect of hydrostatic pressure of blood on arterial and venous pressure in an upright position.

$$\text{Kinetic energy} = \frac{mV^2}{2}$$

where m = mass, and V = velocity.

Thus, when velocity of flow is very low kinetic energy is negligible.

Total energy (total pressure), as stated above according to Burton, at any point equals the sum of potential energy and the kinetic energy. Therefore, when velocity of flow is very low (i.e. with negligible kinetic energy), the magnitude of total energy (i.e. total pressure or perfusion pressure) is almost equal to the lateral pressure. However, when the velocity of flow is high, the lateral pressure exerted by the flowing fluid is much less than the total pressure exerted. This is because of the fact that at high velocity of flow some potential energy is *converted into kinetic energy.* The effect of velocity of flow on total pressure exerted is expressed mathematically by *Bernoulli's principle* which states that in a supine position (i.e. when the effect of gravity or hydrostatic pressure effect is removed) the total energy or pressure (E) of flowing blood in a vessel is:

$E = P + \frac{1}{2}\rho V^2$, where

P = lateral pressure, ρ (Rho) = the density of blood and

V = Velocity of blood in the vessel.

Experimental demonstration of velocity–pressure relationship. The velocity–pressure relationship expressed above in *Bernoulli principle* can be demonstrated experimentally by a system of pilot tubes (Fig. 4.4-10) as:

- AB is a tube through which fluid is flowing from direction A to B.
- The tube AB has central narrow segment S_2 where velocity of flow is higher. On each side of the narrow segment

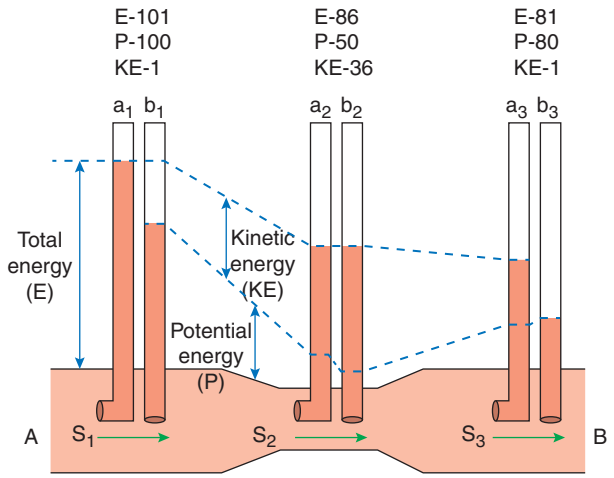


Fig. 4.4-10 Experimental set up demonstrating the effect of velocity on pressure and also effect of resistance on pressure (for explanation see text).

are wide segments S_1 and S_3, where velocity of flow is low. The length of arrows represents relative velocity of flow.

- Six tubes (a_1, a_2, a_3, b_1, b_2, and b_3) are inserted in the tube AB. Lower end of tubes a_1, a_2 and a_3 are so constructed that they face upstream and hence record the total energy (total pressure). While tubes b_1, b_2 and b_3 record lateral pressure only.
- *In segment S_1 of the tube AB*, where velocity of the flow is low (because the diameter is big), there is little difference in the height of fluid level in a_1 (representing total energy) and b_1 (representing lateral pressure).
- *In segment S_2 of the tube AB,* where velocity of the flow is very high (due to smaller diameter) there is marked difference in the height of fluid level in a_2 (representing total energy) and b_2 (representing lateral pressure). This shows that when velocity of flow is high, the potential energy is converted into the kinetic energy and the potential energy (lateral pressure) becomes much less than the total pressure (potential energy plus kinetic energy).
- *In segment S_3 of the tube AB,* where velocity of the flow slows again (due to bigger diameter), part of the kinetic energy is transformed into the potential energy and consequently, lateral pressure is increased and the difference between the height of fluid level in a_3 (representing total energy and b_3 (representing lateral pressure) decreases.

Conversion of energy and loss of energy. If on considering the values of total pressure (E), lateral pressure (P) and kinetic energy (KE) in tubes a_1, a_2, a_3 and b_1, b_2, b_3, following conclusions can be drawn:

- *Energy can be converted* between the potential and the kinetic energy (P ⇔ KE) as:
  - in segment S_1, P is 100 and KE is 1;
  - in segment S_2, P falls to 50 and KE increases to 36 and
  - in segment S_3, P increases to 80 and KE falls to 1.
- *Energy in a moving fluid is progressively lost* due to the resistance offered by the walls of the tubes as E which in segments S_1 is 101, falls to 86 in S_2 and in S_3 is 81. Further, the energy loss (pressure drops) occurs more where the resistance is high (narrow S_2 segment).

BLOOD FLOW: TYPES, MEASUREMENT AND DISTRIBUTION

TYPES OF BLOOD FLOW

Blood flow in the vascular system is of two types:

- Laminar blood flow and
- Turbulent blood flow.

Laminar blood flow

Blood flow in the blood vessels is normally in streamline, like the flow of liquids in narrow rigid tubes. Such a blood flow is called laminar blood flow and is considered to consist of a series of thin laminae slipping over one another.

- The outermost lamina, i.e. an infinitely thin layer of blood in contact with the wall of the blood vessel does not move and the next lamina has some velocity. The subsequent inner layers have progressively increasing velocity and thus the innermost lamina, i.e. the core of the blood stream has the maximum velocity (Fig. 4.4-11).
- The laminar blood flow being streamlined is noiseless and within physiological limits shows a linear relationship with the pressure.

Turbulent blood flow

The above described laminar blood flow occurs up to a certain velocity, at or above which the blood flow becomes turbulent. The velocity of flow at which blood flow becomes turbulent is called critical velocity.

- In turbulent blood flow, the blood moves in an irregular varying paths continuously mixing within the vessel and colliding with the vessel wall. This causes a greater energy loss as compared to the laminar flow.
- The turbulent blood flow is noisy and does not show a typical linear relationship with the pressure.
- Normally, none of the small vessels show turbulent flow. In humans, the critical velocity sometimes exceeds in the aorta at the peak of systole.

Probability of turbulence

The chances of blood flow becoming turbulent are determined by the *probability of turbulence,* which is denoted as Re (Reynold number), named for the man who described it. According to Reynold, the probability of turbulence (Re) is:

- Directly proportional to the:
  - density of blood (ρ, i.e. rho) equal to 1,

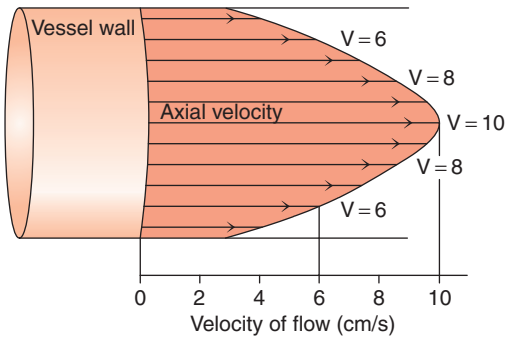


Fig. 4.4-11 Laminar blood flow showing different velocities of the different laminae resulting in a parabolic distribution of velocities. Note that the central core of blood stream has greatest velocity.

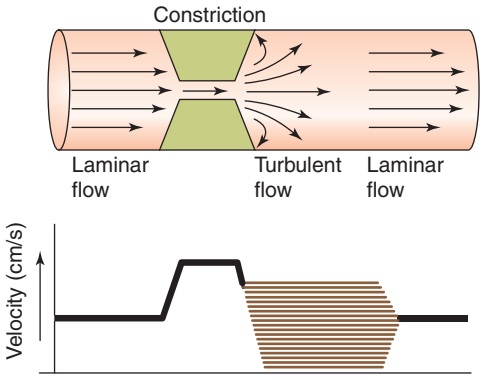


Fig. 4.4-12 Turbulent blood flow caused by the constriction of the lumen of blood vessel.

- diameter of the vessel (D) in cm,
- velocity of blood flow (V) in cm/sec; and is

- Indirectly proportional to:
  - viscosity of the blood (η, i.e. eta) in poises.

$$\text{thus Re} = \frac{\rho DV}{\eta}$$

- Blood flow is usually not turbulent when Re is less than 2000.
- Chances of blood flow becoming turbulent are increased when Re exceeds 2000.

When Re is more than 3000, turbulence is almost always present.

Conditions associated with turbulent blood flow

- Constriction of the artery by an atherosclerotic plaque (Fig. 4.4-12) or by any other cause, e.g. application of external pressure while measuring the blood pressure with a sphygmomanometer is associated with a blood flow velocity which exceeds critical velocity and thus causes the turbulent blood flow. The turbulent flow generates vibrations (sounds) which can be heard over the artery by a stethoscope, e.g. *Korotkoff sounds* heard while recording the blood pressure or the *murmur* heard over a constricted artery.
- *Anaemia.*

MEASUREMENT OF BLOOD FLOW

Following methods are known for the measurement of blood flow:

1. Methods based on the Fick's principle

Methods based on the Fick's principle can be used to measure the blood flow through some of the organs like the lungs, the heart and the brain. The methods are similar to those described for the measurement of cardiac output (page 216).

2. Para-amino-hippuric (PAH) acid clearance method

Para-amino-hippuric acid clearance method is used to determine the renal blood flow.

3. Venous occlusion plethysmography

Venous occlusion plethysmography is a simple but crude method of measuring the blood flow through the limbs.

Principle. It is based on the principle that if venous return of a region (part) is suddenly obstructed, that part increases in size due to arterial flow. The increase in size is equivalent to the blood flow to that part.

4. Electromagnetic flowmeter

The electromagnetic flowmeter is based on the principle that when a vessel containing blood (a conductor) is placed between the two poles of a magnet, voltage is generated in the blood flowing through the magnetic field. The magnitude of the voltage is proportionate to the volume of flow and can be measured with an appropriately placed electrode on the surface of the vessel (Fig. 4.4-13).

5. Doppler flowmeter (Ultrasonic flowmeter)

Ultrasonic flowmeter is based on the principle of Doppler effect. In this instrument, ultrasonic waves are sent into a vessel diagonally from one crystal and the waves reflected from the red and white blood cells are picked up by a second downstream crystal (Fig. 4.4-14). The frequency of the reflected waves is higher by an amount that is proportionate to the rate of flow towards the second crystal because of the Doppler effect.

DISTRIBUTION OF BLOOD FLOW TO VARIOUS REGIONS OF THE BODY

At rest, about 5 L of blood enters aorta per minute. In terms of tissue weight, blood flow to liver, brain and heart is very high. Kidney has a high blood flow because it is related to

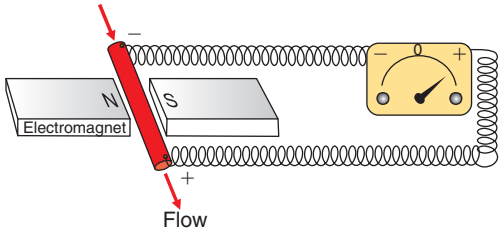


Fig. 4.4-13 Principle of electromagnetic flowmeter.

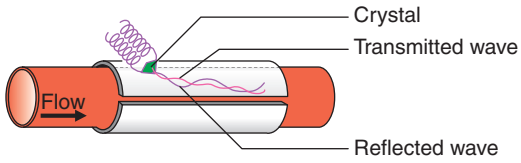


Fig. 4.4-14 Principle of ultrasonic flowmeter.

Table 4.4-2 Distribution of blood flow to various organs of the body

Organs	Blood flow per organ (mL/min)		Blood flow (mL/100 g/min)	
	At rest	During maximum activity	At rest	During maximum activity
Heart	250	1200	80	400
Brain	750	2100	55	150
Liver	1500	3000	58	120
Skeletal muscles	150	1800	04	70
Kidney	1200	1400	400	450
Skin	200	3500	08	150

the excretory function rather than the metabolic requirement. The distribution of blood flow to various organs and regions of the body during resting conditions and during maximum activity conditions is shown in Table 4.4-2.

- At rest, at least 50% of the circulating blood volume is in the systemic veins.
- Twelve percent is in the heart cavities, and 18% in the lower pressure pulmonary circulation. Only 2% in the aorta, 8% in the arteries, 1% in the arterioles and 5% in the capillaries.
- When extra blood is administered by transfusion, less than 1% of it is distributed in the arterial system (the high pressure system) and all the rest in the systemic veins, pulmonary circulation and heart chambers other than the left ventricle (the 'low pressure system').

PRESSURE AND FLOW-IN VARIOUS FUNCTIONAL SEGMENTS OF SYSTEMIC VASCULAR TREE

PRESSURE AND FLOW FUNCTIONS OF ELASTIC ARTERIES

The large elastic arteries (windkessel vessels) include aorta and its main branches, such as carotid, iliac and axillary arteries. These vessels contain elastic tissue in their walls in abundance which provides them two properties of distensibility and elastic recoil. The effect of these two properties of the elastic arteries on pressure and flow of blood is:

DISTENSIBILITY

As we know, the heart acts as a pump and ejects about 70 mL of blood into the aorta with each systole. The distensibility (compliance) of the elastic arteries allows them to accommodate the stroke volume of heart with only a moderate increase in pressure (from 80 mm Hg to 120 mm Hg) (Fig. 4.4-15A). Due to distension of these vessels, a part of energy released from the heart is stored as the potential energy in the wall of aorta.

ELASTIC RECOIL

During diastole, the stretched elastic wall of the aorta recoils and the potential energy stored in the wall is released onto the blood. This causes the blood to flow during diastole also, in this way the pressure in the aorta does not fall below 80 mm Hg (Fig. 4.4-15B). In other words, the elastic recoil of big arteries acts as a subsidiary pump for a continuous blood flow. This recoil effect is called *windkessel effect.* Windkessel is a German word meaning elastic reservoir.

Functions of elastic vessels

1. They reduce the velocity of blood flow to some extent during the ventricular contraction (systole) due to property of distensibility.
2. They cause increase in the velocity of blood flow to some extent during the ventricular diastole by elastic recoil. Thus, the windkessel effect reduces the energy expenditure of heart.
3. Pumping action of the heart along with the elastic recoil of aorta together constitutes a driving force for blood to move forward (towards periphery). This force is called *vis-a-tergo force* and is an important determinant for the venous return.
4. Conversion of pulsatile blood flow from the heart to a steady continuous flow. The elastic vessels act together with arterioles (resistance vessels) to convert this pulsatile flow into a steady continuous flow in the tissue capillaries, which allows maximum exchange between the blood and tissues.

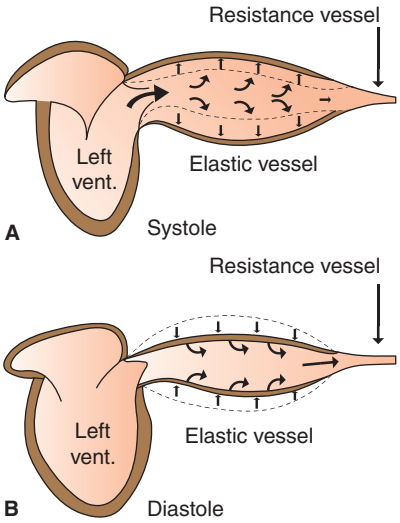


Fig. 4.4-15 Distensibility (A) and elastic recoil (B) seen in aorta, and its main branches maintain arterial pressure and flow during diastole.

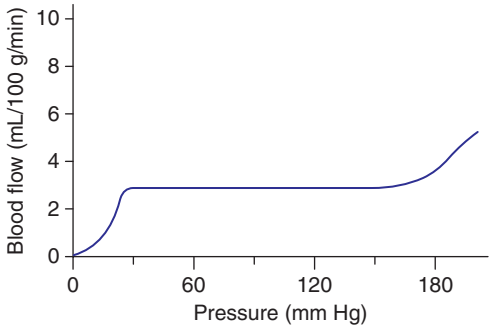


Fig. 4.4-16 Autoregulation of blood flow. Note that the blood flow remains relatively constant over a wide range of arterial pressure. This is accomplished by change in resistance proportionate to change in arterial pressure.

APPLIED ASPECTS

1. ***Due to age-related degenerative changes,*** the elasticity of large vessels is decreased and so is the windkessel effect. Therefore, in old age systolic blood pressure increases due to loss of distensibility and diastolic blood pressure decreases (due to loss of elastic recoil). Thus, in normal healthy individual aged about 70 years, typical blood pressure is 160/70 mm Hg. That is, there occurs *systolic hypertension* with an *increased pulse pressure* (SBP-DBP).
2. ***Atherosclerotic changes in small blood vessels*** are also common in old age. These produce *essential hypertension, i.e.* an increase in the systolic as well as the diastolic blood pressure.

PRESSURE AND FLOW FUNCTIONS OF MUSCULAR ARTERIES

The muscular arteries comprise most of the named arteries in the body, such as radial artery, facial artery, ophthalmic artery and so on. These arteries serve as the *distributing channels* to the organs.

PRESSURE AND FLOW FUNCTIONS OF ARTERIOLES

STRUCTURAL CHARACTERISTICS

Each arteriole is only a few millimeter long and branches many a times to supply about 10–100 capillaries. The characteristic features of arterioles are:

- A thick muscular wall having a profuse vasomotor (sympathetic) innervations and
- A relatively narrow lumen (30 µm), because of which these vessels are considered the major site of peripheral resistance. The arterial pressure drops by about 50 mm Hg while passing through a few millimetre long arterioles.

FUNCTIONS OF ARTERIOLES

1. Control of blood flow to the organs

The arterioles play a major role in the control of blood flow to the organs or tissues. So, they are considered *stopcocks (valves) of circulation.* The constriction of the arterioles increases the resistance and decreases the blood flow while dilation of arterioles decreases the resistance and increases the blood flow. The arterioles control blood flow to the organs by the following two mechanisms:

Autoregulation. It is the ability of an organ or tissue to adjust its vascular resistance and maintain a relatively constant blood flow over a wide range of arterial pressure (Fig. 4.4-16). Autoregulation is well developed in the kidney, brain, heart, skeletal muscle and mesentery. Two theories have been put forward to explain the phenomenon of autoregulation.

Metabolic theory proposed that an increased arterial blood pressure initially increases the blood flow to a tissue or organ. This increased blood flow washes out the vasodilator substances such as CO_2, H^+, nitric oxide, adenosine, prostaglandins, K^+, phosphate ions and low oxygen levels in the area. As a result, the arterioles constrict, the vascular resistance increases and the blood flow returns to normal.

Myogenic theory. According to this theory, the vascular smooth muscle (VSM) responds to wall tension. An increase in the arterial pressure initially stretches the smooth muscle fibres; in response to which the VSM contracts and increases the resistance and compensates for the higher arterial pressure, returning the blood flow to control levels.

2. Conversion of pulsatile flow from the heart to a steady continuous flow

As described in the function of the elastic vessels, the arterioles along with the elastic vessels convert the pulsatile flow in the arteries to a steady flow in the capillaries.

MICROCIRCULATION

ARCHITECTURE OF MICROCIRCULATION

The microcirculation involves a meshwork of vessels less than 100 µm in diameter. These include small arterioles, meta-arterioles, capillaries, post-capillary venules and arteriovenous shunts (Fig. 4.4-17):

Meta-arterioles. The arterioles divide into smaller muscles walled vessels, sometimes called meta-arterioles and these in turn feed into capillaries.

Precapillary sphincters refer to a cuff of smooth muscle cells that surround the origin of capillaries. These determine the size of the capillary exchange area at one particular moment in the tissue. For example, increase in the sphincter patency increases number of open capillaries. Precapillary sphincters respond to local or circulating vasoconstrictor substances.

Capillaries arise directly from the arterioles or meta-arterioles. These vessels allow easy exchange of gases and nutritive substances across them and so are also called as *exchange vessels.* Capillaries constitute the most important segment of the circulatory system. Their structure and functions will be discussed in detail.

Postcapillary venules, which measure 20–60 µm in diameter, are the *most permeable part* of the microcirculation.

Arteriovenous anastomosis (shunt or thoroughfare vessels). These are short, low-resistance connections between the arterioles and veins, bypassing the capillaries. These are abundantly innervated by the vasomotor sympathetic fibres. These vessels are especially found in the skin of fingers, toes and ear-lobes, where they are involved in the regulation of body temperature.

STRUCTURAL CHARACTERISTICS OF CAPILLARIES

Each capillary has an average diameter of 5 µm, length of 50 µm, wall thickness of 1 µm and cross-sectional area of 40 μm^2.

The capillary wall essentially consists of a single layer of endothelial cells which are lined on the outside by a basal lamina (glycoprotein), overlying the basal lamina there may be isolated branching perivascular cells called *pericytes.*

The endothelial structure of capillaries varies in different organs depending on the function of the particular tissue.

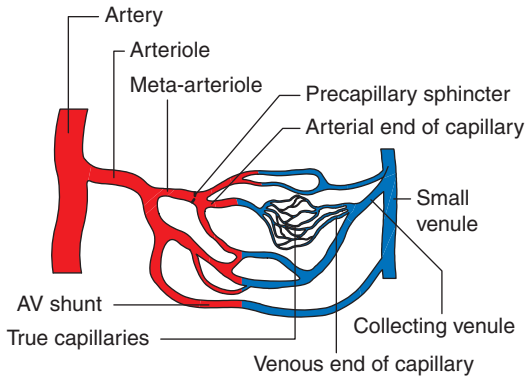


Fig. 4.4-17 Architecture of microcirculation.

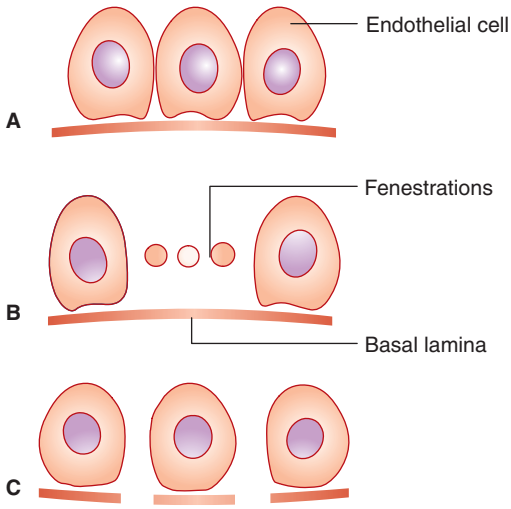


Fig. 4.4-18 Structure of different types of capillaries: A, continuous; B, fenestrated and C, discontinuous.

Under the electron microscope, three types of capillaries have been identified:

1. Continuous capillaries are characterized by a single layer of endothelial cells which are almost continuous, except for small clefts of 6–7 nm in size in between the cells. It is believed that most of the water soluble ions (Fig. 4.4.18A) and molecules pass across the capillary through these clefts (or *slit-pores).* These are the most common type of capillaries and are found in most of the body tissues viz. skeletal muscle, adipose tissue, connective tissue, pulmonary circulation and so on.

2. Fenestrated capillaries consist of thin endothelial cells with large fenestrations (20–100 nm in diameter) in between which are bridged by a thin basement membrane which surrounds the endothelial cells (Fig. 4.4-18B). The fenestrations permit the passage of relatively large molecules and make the capillaries porous. Fenestrated capillaries are found in the renal glomeruli, intestinal villi and most endocrinal glands.

3. Discontinuous capillaries are characterized by large gaps (600–3000 nm in diameter) between endothelial cells

that are not closed by the basement membrane (Fig. 4.4-18C). Through these gaps even formed elements of blood can pass freely. Such capillaries are also called sinusoids and are found in the bone marrow, liver and spleen.

FUNCTIONAL CHARACTERISTICS OF CAPILLARIES

The primary function of circulation is to transport nutrients to the tissues and remove waste products that occur in the capillaries. About 10 billion capillaries which have a total surface area of 500–700 m^2 provide this function for the body. The cross-sectional area of capillary bed when fully patent is 2800 times that of aorta.

Active and inactive capillaries

In resting tissues, most of the capillaries (75%) are collapsed (inactive capillaries) and blood bypasses them to flow through the *thoroughfare vessels* connecting the arterioles to the venules.

In active tissues, the meta-arterioles and the precapillary sphincters dilate and the blood flows through all the capillaries (active capillaries).

The opening and closing of the precapillary sphincters is controlled mostly by the local metabolic vasodilators and possibly also through sympathetic innervation.

Blood flows into the capillaries intermittently because of phenomenon of vasomotion, i.e. intermittent contraction of meta-arterioles and precapillary sphincters. This in turn is mainly controlled by the concentration of oxygen and waste products of tissue metabolism.

TRANSCAPILLARY EXCHANGE

The capillary blood brings oxygen, electrolytes and nutrients to the tissues and removes the waste products of cellular metabolism. The exchange of these substances occurs across the thin membrane formed by the endothelial cells.

Mechanisms of transcapillary exchange

Exchange of substances occurs primarily in the capillaries and post-capillary venules. The major mechanisms of exchange are diffusion and filtration (bulk flow). Some substances also pass through the cells by the *vesicular transport.*

Diffusion across microvascular endothelium

Diffusion is the principal mechanism of the microvascular exchange of materials between the plasma and the interstitial fluid against their concentration gradient.

- *Lipid soluble substances* like O_2 and CO_2 diffuse most freely across the cell membrane.
- *Water and water soluble micromolecules* (molecular weight less than 69,000) like Na^+, Cl^-, K^+, glucose, urea, etc. diffuse almost freely through the intercellular clefts and intracellular pores in the capillary membrane.
- *Large molecules* (molecular weight more than 69,000) such as albumin and other plasma proteins cannot cross the endothelial barrier. However, some amount of albumin does enter the interstitial fluid.

Filtration and reabsorption across microvascular endothelium

The rate of filtration and absorption at any point along the capillary wall depends on the balance of forces known as *Starling forces.* According to the Starling hypothesis, the filtration absorption is expressed as

$$K(Pc + \pi i) - (Pi + \pi c),$$

K = the permeability-surface area coefficient,
Pc = the hydrostatic capillary pressure,
Pi = the hydrostatic interstitial pressure,
π_c = oncotic pressure of blood and
π_i = oncotic pressure of the interstitium.

thus, Pc–Pi, represents the hydrostatic pressure gradient and π_c–π_i, represents the oncotic pressure gradient.
Starling forces are defined:

1. Hydrostatic capillary pressure (Pc) tends to force the fluid out through the capillary membrane. The values of hydrostatic capillary pressure in most of the tissues are:

- At the arterial end = 30–40 mm Hg,
- At the venous end = 10–15 mm Hg and
- In the middle = 25 mm Hg.

2. Hydrostatic interstitial pressure (Pi) tends to force fluid inward through the capillary membrane. It is about –2 mm Hg in the subcutaneous tissue.

3. Oncotic pressure of blood or plasma colloid osmotic pressure (pc) results from the osmotic pressure of plasma proteins. It tends to pull fluid inward through the capillary membrane. It is about 25–27 mm Hg.

4. Oncotic pressure of the interstitium (πi) is due to the presence of proteins in the interstitial space. It tends to pull fluid out of the capillary membrane. The effective oncotic pressure in the interstitium is estimated to range between 2 and 10 mm Hg (average 3 mm Hg).

Calculation of net filtration at the capillaries

From the above description, the net forces acting on the fluid at the arteriolar and venous end of a typical muscle *capillary* can be calculated (Table 4.4-3):

- Thus at arteriolar end, the filtering force of 15 mm Hg moves the fluid out of capillary (filtration) into the interstitium, while at venous end the filtering force of –10 mm Hg

moves the fluid into the capillary (reabsorption) from the interstitium.

- The net filtering force for the whole capillary is 5 mm Hg.

In the example given above we have studied the balance of the Starling forces at the arteriolar and venous end of the capillary only. However, over the length of the capillary, the hydrostatic pressure gradually declines to zero near the middle of capillary. From here the inward forces become dominant and reabsorption process starts to reach its maximum at the venular end (Fig. 4.4-19).

Capillary filtration and reabsorption in different tissues

In above example, we have calculated *net filtration forces.* However, the net filtration in the capillaries varies in the different tissues not only on the balance of Starling forces (net filtration force) but also by the capillary filtration coefficient (K).

Table 4.4-3 Calculation of net filtration force at the capillaries

	At the arteriolar end (mm Hg)	At the venous end (mm Hg)
• **Forces tending to move fluid outward**		
– Capillary hydrostatic pressure (Pc)	35	10
– Interstitial oncotic pressure (πi)	3	3
Total outward force (Pc + πi)	**38**	**13**
• **Forces tending to move fluid inward**		
– Oncotic pressure of blood (πc)	25	25
– Interstitial hydrostatic pressure (Pi)	–2	–2
Total inward force (πc + Pi)	**23**	**23**
• **Filtering force**	(38 – 23) = 15	(13–23) = 10
• **Net filtering force of the whole capillary**	= (15 – 10) = 5	

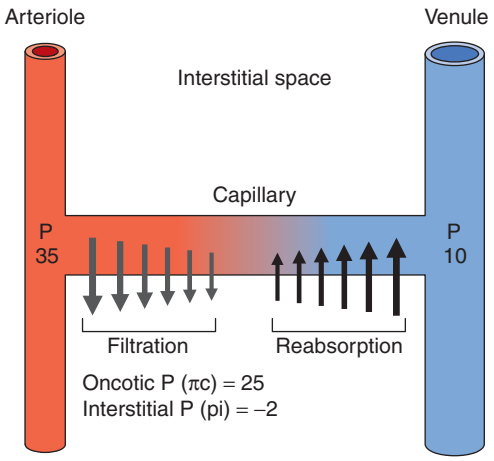


Fig. 4.4-19 Filtration and reabsorption in a capillary due to balance of the Starling forces.

Capillary filtration coefficient or the so-called permeability-surface area coefficient (K) in different tissues varies. For example,

- In subcutaneous tissues K is –0.01 mL/min/mm Hg/100 g tissue and
- In kidney K is –4.2 mL/min/mm Hg/100 g tissues, which is almost 400 times as great as K of many other tissues. This obviously causes a much greater rate of filtration in the glomerular capillaries of the kidney.

Thus, depending upon the balance of Starling forces and capillary filtration coefficient the capillary exchange in some important tissues is:

- *In renal glomerular capillaries,* fluid moves out of almost the entire length of the capillaries.
- *In interstitial capillaries,* fluid moves out of the capillaries from the arteriolar end up to the middle part and then moves into the capillaries.
- *In pulmonary capillaries,* filtration does not occur at all.

LYMPHATIC CIRCULATION

LYMPH: FORMATION AND COMPOSITION

Formation. As discussed in capillary exchange, most (90%) of the fluid filtered at the arterial end of the capillary is reabsorbed at its venous end and the remaining 10% enters the circulation through the lymphatics and is called lymph. Thus, the lymph is a transudate formed from the blood in the tissue spaces, i.e. it is derived from the interstitial fluid.

Composition of lymph is similar to the plasma except that its protein content is usually lower than that of the plasma (2–5 g/dL).

Protein content of the lymph varies with the region it drains are as below:

- Liver 6 g/dL
- Intestine and Thoracic duct 4 g/dL
- Skeletal muscle and Skin 1.5–2 g/dL
- Choroid plexus Zero

Fat content. Since, the lymphatic system also provides a route of absorption of long-chained fatty acids and cholesterol from the intestine (in the form of *chylomicrons);* therefore, after a fatty meal these fat globules may be so numerous that lymph becomes milky and is then called *chyle.*

Cellular content. Suspended in the lymph are cells that are chiefly lymphocytes. Most of these lymphocytes are added to the lymph as it passes through the lymph nodes.

LYMPHATIC VESSELS

The lymphatic system constitutes an accessory route for the removal of interstitial fluid. The small lymph vessels are called lymph capillaries and the large lymph vessels are called lymphatic trunks and the largest lymph vessel is thoracic duct.

Lymph capillaries

The lymph capillaries *originate* as closed endothelial tubes that are permeable to fluid and high-molecular weight compounds.

The *structure* of lymph capillaries (Fig. 4.4-20) is basically similar to that of the blood capillaries with following differences:

- The basal lamina around the endothelial cells is absent or poorly developed.
- Pericytes or connective tissues are not present around the lymph capillaries.
- There are no visible fenestrations in the endothelium.
- The junctions between endothelial cells are open. In fact, the edges of the endothelial cells overlap in such a way that they form minute flap valves.

Larger lymph vessels

The lymphatic capillaries join to form larger lymph vessels which ultimately form lymphatic trunks and lymphatic ducts as:

Thoracic duct is the largest lymph vessel in the body. It carries lymph from both sides of the body below the diaphragm and from the left side above the diaphragm. Near its termination it receives the *left subclavian lymphatic trunks* carrying lymph from left upper limb, the *left jugular lymphatic trunk* carrying lymph from the left half of head and neck and sometimes the *left bronchomediastinal lymphatic trunk* carrying lymph from the left half of thorax (usually this trunk enters the subclavian vein independently).

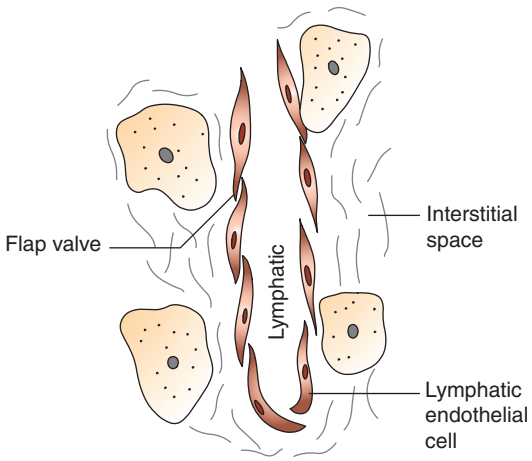


Fig. 4.4-20 Structure of a lymphatic capillary.

The thoracic duct ends by opening into the junction of the left subclavian vein and the internal jugular vein.

Right lymphatic duct drains lymph from the right half of the body above the diaphragm. It is formed by the *right bronchomediastinal trunk* carrying lymph from the right half of the thorax, *right jugular trunk* draining lymph from the right half of head and neck and *right subclavian trunk* carrying lymph from the right upper limb. The right lymphatic duct ends by opening into the right subclavian vein.

Structure of larger lymph vessels is similar to that of the veins:

- *Three coats,* i.e. tunica intima, tunica media and tunica adventitia can be distinguished.
- *Valves* similar to those in veins are present in abundance in small as well as in large lymphatic vessels. The valves often give lymph vessels a beaded appearance.

LYMPH FLOW

Functions of lymph flow

1. Returns proteins from tissue spaces to blood. The lymphatic system recovers approximately 200 g of protein daily that has been lost from the microcirculation.

2. Absorption of nutrients, especially fats from the gastrointestinal tract.

3. Acts as a transport mechanism to remove red blood cells that have lost into the tissues as a result of haemorrhage.

4. Supplies nutrients and oxygen to those parts where blood cannot reach.

5. Role in defence mechanism. Lymph nodes associated with lymphatic system act as efficient filters. They have sinuses lined with phagocytic cells that engulf bacteria, red cells and other particulate material.

Mechanism of lymph flow: Factors affecting are:

1. Intrinsic lymphatic pump. Lymph is pumped out of the tissues by the lymphatic vessels which have valves and smooth muscles in their walls. They contract in a peristaltic fashion, propelling the lymph along the vessels. The extensive system of one-way valves present in the lymphatics maintains lymph flow towards the heart.

2. Pumping by external compression of the lymphatics. Though the contractions of lymphatics are the principal factor propelling the lymph, the lymph is also pumped by the external compression of the lymphatics by:

- Contraction of the skeletal muscles,
- Movements of different body parts,
- Arterial pulsations and
- Compression of tissue by objects outside the body.

3. Negative intrathoracic pressure during inspiration increases the rate of lymph flow.

4. Suction effect of high-velocity blood flow in the veins in which the lymphatics terminate also promotes lymph flow.

5. Interstitial fluid pressure. An increase in the interstitial fluid pressure increases the lymph flow up to a certain limit.

6. Increase in capillary surface area by capillary distension is associated with increased lymph flow under following conditions:

- Increased capillary pressure,
- Increase in local temperature and
- Infusion of fluid.

7. Increase in capillary permeability under following conditions is also associated with the increased lymph flow:

- Increase in temperature,
- Effect of toxins and
- Decreased oxygen (hypoxia).

8. Increase in functional activity of the tissue also increases the lymph flow.

Normal lymph flow

Normal lymph flow is 2–4 L/day (80–150 mL/h) for the entire body.

Rate of lymph flow varies in different organs and is highest in the gastrointestinal tract and the liver.

In lymphatics rate of lymph flow is 100 mL/h through thoracic duct and about 20 mL/h through other lymphatic channels.

VENOUS CIRCULATION

STRUCTURAL CHARACTERISTICS OF VEINS

Walls of veins, as compared to arteries, at equivalent levels in the vascular tree are thin walled and contain small amount of elastic tissue and smooth muscle, and have larger lumen. Because of their structural characteristics, they are more distensible and collapsible.

Lumen of the veins is larger than the equivalent arteries.

Valves are present in the veins of the dependent parts of the body (Fig. 4.3-12) that prevent the back-flow of venous blood.

FUNCTIONS OF THE VEINS

1. Blood reservoirs

Because of their feature of *distensibility and collapsibility,* the veins serve as blood reservoirs. About 60–70% of the circulating blood is present in the venous system. When their blood content decreases, the veins assume an elliptical profile because of their collapsibility. As the venous blood content increases, they assume more and more circular profile to accommodate progressively greater amount of blood per unit length. Further, increase in the volume of blood is accommodated by distension of the walls without any significant increase in the venous pressure (Fig. 4.4-21). Due to this property, veins are also called the *capacitance vessels.*

2. Conduits

The systemic veins carry blood from the tissues to the right atrium and the pulmonary veins collect blood from the lungs and return it to the left atrium.

3. Maintenance of cardiac output

Veins help to maintain the cardiac output whenever there is loss of blood. After blood loss from an external or internal injury, reflex increases in the sympathetic discharge produces contraction of the smooth muscles in the walls of the veins. As a result of venular contraction, there occurs a decrease in their capacity leading to an increased venous return to the heart. In this way, veins help to maintain cardiac output by maintaining normal venous return in spite of the blood loss.

VENOUS PRESSURE AND FLOW

Central venous pressure

Central venous pressure refers to the pressure in the right atrium because all the systemic veins open into the right atrium. The normal right atrial pressure is about 0 mm Hg (i.e. equal to atmospheric pressure) but it can rise to as high as 20–30 mm Hg under abnormal conditions, such as heart

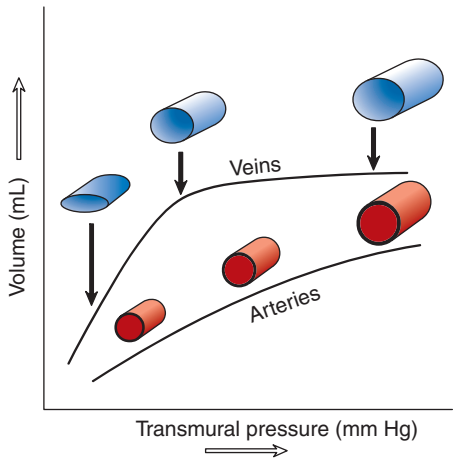


Fig. 4.4-21 Pressure–volume relationship of arteries and veins. The slope of each line at any point represents compliance (distensibility). Note that the veins are much more compliant than the arteries at low pressure because veins are not completely distended at these pressures.

failure and massive blood transfusion. The right atrial pressure can decrease to as low as –3 to –5 mm Hg when the heart (right atrium) is pumping with vigour or when venous return is greatly depressed.

Peripheral venous pressure

The pressure in the venules is about 10 mm Hg. As the veins approach the heart, there is a gradual decrease in the venous pressure. In the great veins near the heart, venous pressure is approximately 5 mm Hg.

Measurement of peripheral and central venous pressure

Clinical assessment of venous pressure is made by observing the degree of distension of neck veins. When the right atrial pressure is increased up to 10 mm Hg, the lower neck veins begin to protrude in a sitting position (in normal person, in this position neck veins are never distended).

Peripheral venous pressure measurement with a manometer can be made easily by connecting it to cannula inserted into a superficial vein.

Central venous pressure measurement using a cardiac catheter whose tip is led up to the superior vena cava through a superficial vein can be made accurately. The other end of the catheter is connected to a pressure transducer.

VENOUS FLOW AND VENOUS RETURN

As we know, the blood flows in the veins towards the heart due to a pressure gradient which exists between the right atrial pressure (0 mm Hg) and the peripheral veins (6–7 mm Hg).

Venous return has been discussed in detail on page 221.

BLOOD PRESSURE

DEFINITIONS (TERMINOLOGY)

BLOOD PRESSURE

Blood pressure is the lateral pressure exerted by the flowing blood on the walls of the vessels. It is usually measured in mm Hg. Without any further qualification the term blood pressure denotes the *arterial pressure.* While describing the pressure exerted by the blood column in other types of blood vessels, the type of vessels is also mentioned, e.g. *capillary pressure* and *venous pressure.*

Systolic blood pressure

- The maximum arterial pressure during the systole is called *systolic blood pressure* and occurs during the ventricular ejection.
- The systolic blood pressure is a function of the cardiac output (CO), i.e. it represents the extent of work done by the heart.
- Normal systolic blood pressure in a young adult is 120 mm Hg (range: 105–135 mm Hg).
- Systolic blood pressure undergoes considerable fluctuations, e.g. it is increased during excitement, exercise and meals, and is decreased during sleep and rest.

Diastolic blood pressure

- Diastolic blood pressure refers to the minimum arterial pressure during diastole and occurs just before the onset of the ventricular ejection.
- Normal diastolic blood pressure in a young adult is 80 mm Hg (range 60–90 mm Hg).
- The diastolic pressure is the function of TPR and indicates the constant load against which the heart has to work. It undergoes much less fluctuations.

Conventional expression of blood pressure

Conventionally, systolic and diastolic blood pressures are denoted as numerator and denominator, respectively. For example, blood pressure of a normal person is written as 120/80 mm Hg.

Pulse pressure

- Pulse pressure (PP) is the arithmetic difference between the systolic and the diastolic blood pressures.
- Normally, average pulse pressure is 40 mm Hg.
- Pulse pressure determines the pulse volume and
- The high pulse pressure is an indicative of the systolic hypertension and indirectly determines a decrease in elasticity of blood vessels.

Mean arterial pressure

- Mean arterial pressure (MAP) is the average of all pressure measured millisecond by millisecond throughout the cardiac cycle.
- Since, the duration of cardiac systole is shorter than the duration of diastole, so the MAP is not equal to the algebraic mean of the systolic and diastolic blood pressures, i.e. it is not equal to

$$(\text{Systolic pressure} + \text{diastolic pressure}) \times \frac{1}{2}$$

- Practically, MAP is roughly equal to the diastolic pressure (DP) plus one third of pulse pressure (PP), i.e.

$$\text{MAP} = \text{DP} + \frac{1}{3}\,\text{PP}$$

- MAP is same for each organ and determines the pressure head. Thus, the regional blood flow through an organ depends on it.

- All cardiovascular reflexes are sensitive to mean arterial pressure.
- Normal value of MAP is 93 mm Hg (range: 90–100 mm Hg).

DETERMINANTS OF ARTERIAL BLOOD PRESSURE

- The arterial blood pressure (BP) is a function of the product of cardiac output (CO) and total peripheral resistance (PR), i.e.

$$\text{Arterial BP} = \text{CO} \times \text{PR}$$

- Therefore, the arterial blood pressure is affected by conditions that affect either cardiac output or peripheral resistance.
- Changes in the cardiac output affect the systolic pressure more than the diastolic pressure while changes in the peripheral resistance affect diastolic pressure more than the systolic pressure.
- As discussed in the section on cardiac output (page 215), the cardiac output is a function of *heart rate* and *stroke volume,* so these two are important determinants of the blood pressure.
- The peripheral resistance (page 230) depends upon the viscosity of blood, elasticity of the vessel wall and velocity of the blood flow. Thus, these factors are also important determinants of the blood pressure.

Effect of some of the important determinants of the blood pressure is shown in Fig. 4.3-10 and described here briefly:

1. Heart rate. An increase in the heart rate usually increases the cardiac output and decreases the duration of cardiac cycle and thus raised the blood pressure.

Conversely, reduction in heart rate decreases the arterial blood pressure.

2. Stroke volume. An increase in the stroke volume increases the cardiac output and raises the arterial pressure and the reverse effect occurs due to decrease in the stroke volume.

3. Arterial elastic constant. It refers to the stiffness of the arterial system which progressively increases from birth until death. An increase in the arterial elastic constant (or loss of elasticity of vessel wall) with advancing age results in a decreased stretching of the elastic vessels during systole. This results in an increased pressure during systole *(systolic hypertension)* with normal diastolic blood pressure. It is characterized by high pulse pressure, when stiffness occurs in small vessels also, the diastolic blood pressure is also increased.

4. Arterial blood volume. An increase in total blood volume increases both systolic and diastolic blood pressures by increased quantity of blood in the arterial system and greater stretching of the vessel wall.

Conversely, *haemorrhage* and *blood pooling* reduces the arterial pressure by reducing the circulating blood volume.

5. Peripheral resistance (PR). Peripheral resistance is an important determinant of arterial pressure. An increase in PR increases and a decrease in PR decreases the arterial pressure.

Peripheral resistance in turn is determined by the radius of vessels (arterioles), velocity of blood flow and viscosity of blood.

VARIATIONS IN BLOOD PRESSURE

PHYSIOLOGICAL FACTORS AFFECTING BLOOD PRESSURE

1. Age. In healthy humans, both systolic and diastolic pressures rise with age.

- *At birth,* systolic blood pressure is 40 mm Hg (range: 20–60 mm Hg). It then rises rapidly up to 1 month of age.
- *At 1 month* of age, the systolic blood pressure becomes about 80 mm Hg and then rises slowly.
- *At about 17 years* of age, normal adult level of blood pressure 120/80 mm Hg.
- *At about 70 years of age,* the normal value of blood pressure is 160/90 mm Hg. The increase in blood pressure associated with advancing age is due to increase in rigidity of vessel wall.

2. Sex. *Before menopause,* females have little lower (4–6 mm Hg) systolic blood pressure than males of corresponding age. *After menopause,* systolic blood pressure in females is little higher (4–6 mm Hg) than males of same age group.

3. Effect of meals. *Systolic blood pressure* increases by 4–6 mm Hg after meals and this effect lasts for about 1 h. *Diastolic blood pressure* either remains unchanged or decreases slightly due to the vasodilatation in splanchnic vessels.

4. Emotions. Increased sympathetic activity during emotional situations leads to increase in the systolic blood pressure.

5. Climatic temperature. *Exposure to cold* produces rise in the blood pressure, while *exposure to hot temperature* lowers the blood pressure.

6. Diurnal variation. Systolic blood pressure shows a diurnal variation of about 6–10 mm Hg; the values being lower in the morning and higher in the afternoon. In night workers however, a reverse rhythm is observed.

7. Exercise. In muscular exercise, generally systolic blood pressure rises and diastolic blood pressure falls. For details see page 370.

8. Effect of gravity. In standing position, due to hydrostatic (gravitational) effect of the blood column, the pressure in

the vessels below heart level is increased and in the vessels above heart level it is decreased (Fig. 4.4-9). For every centimetre below or above the heart level the pressure increases or decreases by 0.77 mm Hg, respectively. Therefore, for clinical recording blood pressure should always be checked at the heart level.

9. Effect of change in posture. Sudden change in posture from lying down to standing initiates some momentary changes in blood pressure which in normal humans are immediately rectified by baroreceptor reflexes and practically such changes are not experienced. However, in patients with autonomic disturbances these changes become symptomatic. Chain of physiological changes in blood pressure during change of posture is discussed below.

Immediately on standing, there occurs peripheral pooling of blood in dependent parts leading to decreased venous return and decreased cardiac output and momentary fall in the systolic blood pressure. Fall in systolic pressure immediately decreases baroreceptor discharge via vasomotor centre leading to an increased *diastolic blood pressure.* On standing, there also occurs an increase in the peripheral resistance and momentary increase in diastolic blood pressure.

Thus, immediately after standing from lying down posture a rise in diastolic blood pressure can be recorded for about 30–60 s. Later on due to decrease in baroreceptor discharge, the blood pressure comes back to normal and no symptoms are experienced by the normal individuals.

10. Sleep. In complete relaxed state during early hours of sleep there occurs fall in blood pressure up to 15–20 mm Hg. However, in disturbed sleep blood pressure increases due to increased sympathetic discharge.

11. Body built. Systolic blood pressure is slightly higher in obese individuals as compared to thin built individuals. This usually occurs because there is more tissue between the cuff and artery and so, some of the cuff pressure is dissipated. Therefore, use of a cuff that is wider than the standard arm cuff is recommended in the obese individuals.

PATHOLOGICAL VARIATIONS IN BLOOD PRESSURE

Hypertension

Definition. Hypertension (HT) refers to a condition in which value of systolic blood pressure is persistently more than 140 mm Hg and/or that of diastolic blood pressure is above 90 mm Hg. If there is increase only in the systolic blood pressure, it is called *systolic hypertension* in which pulse pressure is raised.

Types of hypertension

It is of two types: primary and secondary hypertension.

1. Primary hypertension, also known as *essential hypertension,* is characterized by a raised blood pressure without any underlying disease. Risk factors for primary HT include: heredity, obesity, mental tension and smoking.

2. Secondary hypertension refers to a condition in which blood pressure is raised due to some other underlying disease. Common causes of secondary hypertension are:

- *Cardiovascular diseases,* e.g. atherosclerosis.
- *Renal diseases,* e.g. glomerulonephritis and tumour of juxtaglomerular cells leading to formation of excess of angiotensin II.

Goldblatt's hypertension also known as renovascular hypertension refers to the hypertension due to compression of renal artery or its branches. It can be of two types:

- *One kidney Goldblatt hypertension*. This happens when one kidney is already removed and the renal artery of other kidney is constricted due to any reason.
- *Two kidney Goldblatt hypertension*. This occurs when the artery to one kidney is constricted while the artery to other kidney is normal.

Mechanism. Due to occlusion of renal artery there occurs renal ischaemia, which triggers release of renin causing rapid elevation of blood pressure for the first hour or so. This is followed by a slower additional rise in blood pressure during next several days. This happens because the hyperreninemia increases angiotensin II levels (for details see page 261) causing severe vasoconstriction and aldosterone release leading on to sodium and water retention.

- *Endocrinal disorders* like hyperaldosteronism (excessive secretion of aldosterone from adrenal cortex), phaeochromocytoma (tumour of adrenal medulla) and Cushing's syndrome (excessive secretion of glucocorticoids from adrenal cortex).
- *Neurologic disorders* which may produce hypertension include raised intracranial pressure.
- *Pregnancy-induced hypertension* is noticed in some of the pregnant women. Its exact cause is not known.

IMPORTANT NOTE

Malignant hypertension or accelerated hypertension refers to a sudden marked rise in blood pressure (e.g. systolic up to 250 mm Hg and diastolic up to 150 mm Hg). Malignant hypertension is an emergency and may sometimes be even fatal.

Hypotension

Hypotension refers to a condition in which values of blood pressure are below the normal range. Clinically, when the systolic blood pressure is less than 90 mm Hg, it is considered hypotension. It is of following types:

- *Primary hypotension,* also known as essential hypotension, is a disorder of unknown aetiology.

- *Secondary hypotension,* occurs secondary to some other underlying diseases, such as myocardial infarction, neurogenic shock, haemorrhagic shock, hypoactivity of pituitary gland and hypoactivity of adrenal glands.
- *Postural hypotension* refers to the sudden fall in blood pressure when patients stand up from lying down posture. It occurs due to some dysfunction of autonomic nervous system.

MEASUREMENT OF BLOOD PRESSURE

DIRECT METHOD

Direct method of measuring blood pressure is used in the experimental studies. In it, a cannula or T-tube is inserted into an artery and connected to either

- ***Mercury manometer*** and pressure is recorded on the kymograph, or
- ***Pressure transducer*** (strain gauge) which in turn is connected to Polyrite for recording.

The record will show fluctuations as depicted in Fig. 4.4-22, upper level of which indicates systolic blood pressure and lower level indicates diastolic blood pressure.

INDIRECT METHOD

In human beings, blood pressure is measured indirectly by using sphygmomanometer.

Sphygmomanometer

Commonly called blood pressure apparatus, is the instrument used to measure blood pressure. It consists of following parts:

1. Manometer. *Mercury manometer* is commonly used in classical sphygmomanometer (Fig. 4.4-23). It consists of a *graduated narrow glass tube* having markings 0–300. Upper end of the tube is closed and lower end is connected to the lower end of a wide lumen *mercury reservoir.* Upper end of the mercury reservoir is connected to an inflatable rubber bag through a rubber tube.

2. The cuff. The blood pressure apparatus cuff also known as 'armlet' or 'Riva-Rocci cuff' (after the name of discoverer) consists of an inflatable rubber bag which is enclosed in a cotton bag having a long strip of inelastic cloth. The dimension of the commonly used rubber bag is 24 cm × 12 cm.

3. Air pump. It is a rubber bulb with a one-way valve at its free end, and a 'leaky valve' and a knurled screw at the other end where the rubber tube leading to the cuff is attached. The cuff can be inflated by turning the leak-valve screw clockwise and alternately compressing and releasing the bulb. Deflation is achieved by turning the screw anticlockwise.

Procedure

The blood pressure may be tested with subject lying supine or sitting, but should be physically and mentally relaxed and free from excitation.

The blood pressure can be measured by an auscultatory method.

Auscultatory method. Auscultatory method, described by Korotkoff in 1905, is the most useful technique.

- The cuff of the blood pressure apparatus is applied on the upper arm with the centre of the rubber bag lying over the brachial artery which lies medially, and its lower edge should be about 3 cm above the elbow.
- Raise the pressure in the cuff by 30–40 mm above the level, when the radial artery pulse disappears.
- Then diaphragm of stethoscope is placed on the brachial artery in the cubital space and is kept in a position with the help of thumb and fingers of left hand.
- Pressure in the cuff is lowered slowly by opening the leak valve of the air pump with right hand. While doing so, initially no sound is heard.

However, when mercury column is lowered further a tap sound is heard. The character and quality of sound goes on changing while further lowering the mercury column by deflating the cuff, and ultimately the sound disappears. These sounds are called Korotkoff sounds and from these the levels of systolic and diastolic blood pressure are noted as described below.

Korotkoff sounds

Phases of Korotkoff sounds. Depending upon the characteristic features, the Korotkoff sounds have been described in five phases (Fig. 4.4-24):

Phase I sounds start with a *clear tap* which indicates the *systolic blood pressure.* The clear, tapping and sharp sounds last for 10–12 mm Hg fall in mercury column.

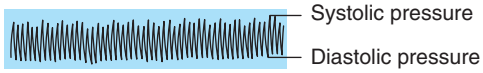


Fig. 4.4-22 Record of arterial blood pressure obtained by direct arterial cannulation method.

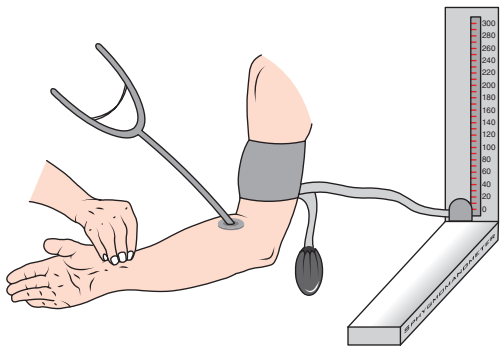


Fig. 4.4-23 Procedure of recording blood pressure by an auscultatory method.

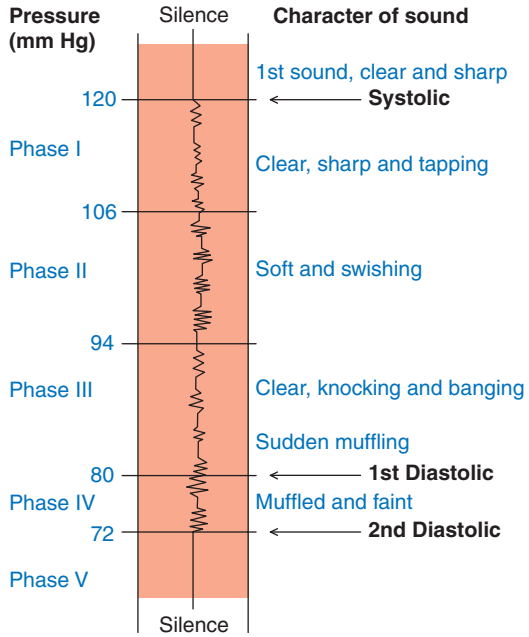


Fig. 4.4-24 Phases of Korotkoff's sounds.

Phase II sounds are murmurish, i.e. soft and swishing and last for next 14–15 mm Hg fall in mercury column.

Phase III sounds are clear, knocking and banging in character and last for next 14–15 mm Hg fall in mercury column.

Phase IV sounds start with *sudden muffling* and mark the *diastolic blood pressure.* The muffled sounds are indistinct, dull and faint, as if coming from a distance and last for next 4–5 mm Hg fall in mercury column.

Phase V is labelled when no sound is heard. Since the beginners may not appreciate beginning of muffling of sounds and therefore, disappearance of the sound may be considered as a mark of diastolic pressure. However, in some clinical situations such as hyperthyroidism and aortic valve insufficiency where the sounds continue to be heard even when the pressure is low, the level at which muffling of sounds starts is to be taken as diastolic blood pressure.

Mechanism of Korotkoff's sounds

- Normally, the blood flow through the arteries is streamlined or laminar and no sounds are heard over them when auscultated.
- As shown in Fig. 4.4-25 while testing blood pressure, when the cuff pressure is raised above the expected systolic pressure level, the blood flow in the brachial artery completely ceases and no sounds are heard. When the cuff pressure is reduced gradually, a time comes when, at the peak of each systole, the intra-arterial pressure just exceeds the cuff (extra-arterial) pressure. The small amounts of blood that are ejected at a high velocity (exceeding the critical velocity) through the partially narrowed artery result in *intermittent turbulence* which produces the sounds. Also the blood column in the distal part of the artery, i.e. below the cuff, is set into vibrations by the jets of blood striking against it, which contributes to the sounds. These sounds are called *Korotkoff sounds.*
- As the cuff pressure falls further, the blood flow becomes laminar once again and the sounds disappear (Phase V).
- The change in character of sounds from phase I to phase IV is related to the degree of turbulence.

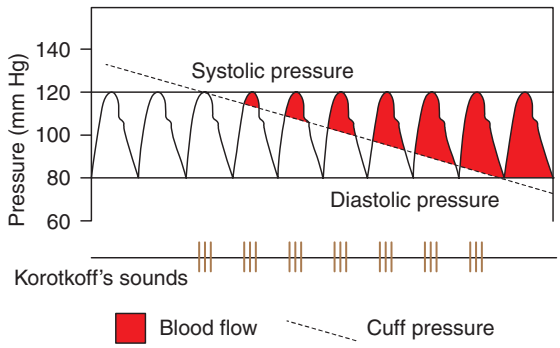


Fig. 4.4-25 Change in the blood flow and production of Korotkoff's sounds with change in the cuff pressure.

REGULATION OF BLOOD PRESSURE

Arterial blood pressure is controlled by several mechanisms which under physiological conditions maintain the normal MAP within a narrow range of 95–100 mm Hg. The different mechanisms concerned with regulation of blood pressure have been discussed in detail in Chapter 4.5 on Cardiovascular Regulation and have been briefed here just for orientation about control of blood pressure. Each mechanism performs a specific function. Various mechanisms controlling arterial pressure can be grouped as:

A. Rapid blood pressure control mechanism,
B. Intermediate blood pressure control mechanisms and
C. Long-term blood pressure control mechanisms.

RAPID BLOOD PRESSURE CONTROL MECHANISM (NERVOUS REGULATING MECHANISM)

Rapid blood pressure control mechanism or the so-called short-term control mechanism primarily includes the following three nervous reflexes:

1. Baroreceptor reflexes (see page 255).
2. Central nervous system ischaemic response (see page 259).
3. Chemoreceptor reflexes (see page 259).

Salient features. The nervous reflexes which rapidly control the blood pressure are described in detail in Chapter 4.5. Their salient features are:

- These *act very rapidly,* i.e. within seconds to few minutes of alterations in the blood pressure.

- These are *short-term mechanisms,* i.e. these act for few hours to few days and are thus insignificant in the long-term regulation of blood pressure.
- These are useful in preventing *acute decreases in blood pressure* (e.g. during severe haemorrhage) as well as in preventing *excessive increases in blood pressure* (e.g. as might occur in response to excessive blood transfusion).

INTERMEDIATE BLOOD PRESSURE CONTROL MECHANISMS

The intermediate blood pressure control mechanisms that are important in the blood pressure control after several minutes of acute pressure changes are:

- Renin–angiotensin vasoconstrictor mechanism, (for details see page 261).
- Stress relaxation and reverse stress relaxation mechanism.
- Capillary fluid shift mechanism.

Salient features of intermediate blood pressure control mechanisms are:

- These mechanisms come into play after several minutes of acute pressure changes and reach full function within a few hours.
- These mechanisms play their role from few days to few weeks.
- All these mechanisms basically try to control the alterations in the blood pressure by altering the blood volume.

Stress relaxation and reverse stress relaxation mechanisms

Stress relaxation mechanism refers to vasodilatation occurring due to stress on the vascular smooth muscles. When pressure in the vessels become too high (e.g. following massive slow intravenous transfusion), the vessels become stretched and continue to stretch for minutes or hours. This causes relaxation of blood vessels simply by a vascular tone adjustment. This leads to an increase in the capacity of the arterial system with a concomitant fall in blood pressure.

Reverse stress relaxation mechanism operates when the blood pressure is low due to less stress on the vessels walls and tries to restore it back to normal. For example, when blood pressure falls due to prolonged slow bleeding, there occurs tightening of blood vessel walls by vascular tone adjustment secondary to less stress on the vessel wall (reverse stress relaxation mechanism). This mechanism tries to restore the blood pressure back to normal. This mechanism can correct up to 15% change in blood volume below normal.

Capillary fluid shift mechanism

Capillary fluid shift mechanism helps in restoring both low and high blood pressure back to normal:

When blood pressure is raised, the mean capillary pressure is also high resulting in shift of fluid from circulation to the interstitial fluid compartments. This reduces the blood volume to restore the arterial pressure.

When blood pressure is lowered, the mean capillary pressure is also low, resulting in absorption of fluid from the interstitial compartments to circulation. Thus the blood volume is increased which helps to return the blood pressure back to normal.

The capillary fluid shift mechanism is about two times more effective than baroreceptor reflex mechanism in controlling the blood pressure, but it acts much more slowly (intermediate acting mechanism) than baroreceptor mechanism (rapid acting mechanism).

LONG-TERM BLOOD PRESSURE CONTROL MECHANISMS

Kidneys play main role in the long-term control of blood pressure by the following mechanisms:

1. Direct mechanism, i.e. Renal body fluid feedback mechanism.

2. Indirect mechanisms control kidney functions indirectly via following hormonal mechanisms:
(i) Aldosterone system and
(ii) Renin–angiotensin system.

Renal-body fluid system for arterial pressure control

The most important mechanism for the long-term control of blood pressure is linked to control of circulatory volume by the kidney, a mechanism known as the *renal-body fluid feedback system.* In fact, it is similar to the capillary fluid shift mechanism except that only the renal glomerular capillaries are involved in the process.

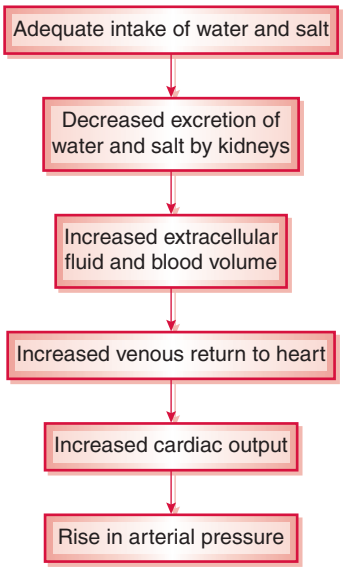
Modes of operation of renal-body fluid feedback system

The renal-body fluid system corrects the blood pressure by causing appropriate changes in the blood volume through *diuresis* and *natriuresis.*

When blood pressure rises too high, the kidneys excrete increased quantities of sodium and water because of *pressure natriuresis* and *pressure diuresis,* respectively. As a result of increased renal excretion, the extracellular fluid volume and blood volume both decrease until blood pressure returns to normal and the kidneys excrete normal amounts of sodium and water.

When the blood pressure falls too low, the kidneys reduces the rate of sodium and water excretion, and over a period of hours to days, if the person drinks enough water and eats enough salt to increase blood volume, the blood pressure will return to its previous level. This mechanism being very slow to act, is not of major importance in the acute control of arterial pressure. However, it is by far the most potent of all long-term arterial pressure controllers.

The sequence of events in order of occurrence during control of blood pressure by this mechanism is summarized as:



Determinants of the renal-body fluid feedback mechanism

The two factors that determine the long-term control of arterial pressure by renal-body fluid mechanism are: renal output curve for salt and water and level of salt and water intake. As long as these two factors remain constant, the mean arterial pressure will also remain exactly at the normal level of 100 mm Hg. For the arterial pressure to deviate from the normal level for long periods of time, one of these two factors must be altered.

Figure 4.4-26 shows the effect of different arterial pressures on urine volume output by an isolated kidney. The figure demonstrates that:

- As arterial pressure rises there occurs a marked increase in the output of volume (pressure diuresis) and sodium (pressure natriuresis).
- As long as the arterial pressure will remain above the normal equilibrium point, renal output will exceed the intake of salt and water resulting in a progressive decline in an extracellular fluid volume.
- When blood pressure falls below the equilibrium point, the renal output of water and salt will be lower than the intake resulting in a progressive increase in an extracellular volume.

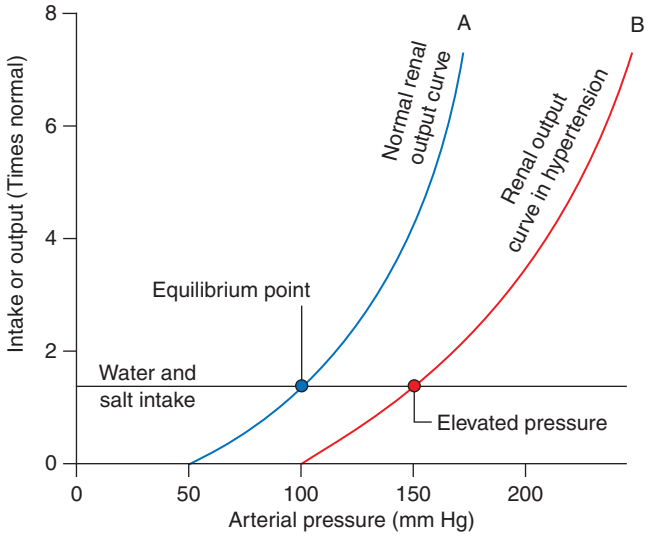


Fig. 4.4-26 Analysis of arterial blood pressure regulation by equating the renal output curve with salt and water intake. The equilibrium point describes the level at which arterial pressure will be regulated.

- At the normal arterial pressure, a balance between renal output and intake of salt and water occurs at so-called the *equilibrium point.*
- As shown in Fig. 4.4-26B, due to an abnormality of the kidney the renal output curve is shifted and the equilibrium point is obtained at a level of high blood pressure (150 mm Hg).
- The renal output and salt and water intake demonstrate that theoretically arterial pressure will be raised with increase in salt and water intake. However, in reality, the blood pressure does not rise every time the sodium intake is increased. This is accomplished mainly by decreasing the formation of angiotensin II and aldosterone, which increases the ability of kidney, to excrete salt and water and results in a compensatory *left shift of the pressure natriuresis curve.*

Salient features of renal-body fluid feedback mechanism

The salient features of the renal-body fluid feedback mechanism can be summarized as:

- The renal-body fluid feedback mechanism takes several hours to show any significant response.
- These mechanisms operate very powerfully to control arterial pressure over days, weeks and months.
- The effectiveness of these mechanisms becomes steadily greater with time.
- These mechanisms, if given sufficient time, control arterial pressure at the level that provides normal output of salt and water by the kidneys.
- As long as kidney function is unaltered, these mechanisms overcome the disturbances that tend to alter arterial pressure such as increased total peripheral resistance over a long period and thus are able to control the blood pressure.

Chapter 4.5

Cardiovascular Regulation

INTRODUCTION

Cardiovascular control makes *circulatory adjustments* which are essential to cope up with the timely needs of each and every organ of the body and is thus of fundamental importance for survival.

NEED FOR CARDIOVASCULAR CONTROL

Functions served by cardiovascular control are:

- *To increase the blood supply* to active tissues, e.g. during exercise to skeletal and cardiac muscles.
- *Redistribution of blood* to increase or decrease the heat loss from the body as per requirements.
- *Circulatory adjustments* during routine cardiovascular stresses like change in posture, hours of excitement, fear, anxiety, meals and sleep, etc.
- *Maintenance of adequate flow to vital organs,* such as brain, heart and kidney, all the times including emergencies such as shock and haemorrhage, even at the expense of the circulation to the rest of the body, as the vital organs may develop irreversible changes in no time. For example, the brain is irreversibly damaged within three minutes of ischaemia, while skin, skeletal muscle and gastrointestinal tract can tolerate reduction of blood for a longer duration.

CIRCULATORY ADJUSTMENTS

Circulatory adjustments which ensure that all of the organs receive sufficient blood flow are: (1) control of blood volume and (2) control of arterial pressure. These circulatory adjustments are made by the cardiovascular control mechanisms primarily by regulating following parameters:

A. Regulation of cardiac performance, i.e. alterations in the activities of heart which include:

1. *Chronotropic action,* i.e. effect on heart rate which may be in the form of:
   - Increased heart rate (tachycardia) or positive chronotropic effect.
   - Decreased heart rate (bradycardia) or negative chronotropic effect.
2. *Inotropic action,* i.e. effect on force of contraction which may be in the form of:
   - Increase in the force of contraction (positive inotropic effect) or
   - Decrease in the force of contraction (negative inotropic effect).
3. *Dromotropic action,* i.e. effect on conduction of impulse through the heart, which may be in the form of:
   - Increase in the velocity of impulse conduction (positive dromotropic effect) or
   - Decrease in the velocity of impulse conduction (negative dromotropic effect).

4. *Bathmotropic action,* i.e. effect on the excitability of the cardiac muscle, which may be in the form of:
   - Increased excitability of cardiac muscle (positive bathmotropic effect) or
   - Decreased excitability of cardiac muscle (negative bathmotropic effect).

B. Regulation of performance of blood vessels, which primarily includes:
- *Alterations in diameter of arterioles,* which change the peripheral resistance and also the hydrostatic pressure in the capillaries.
- *Alterations in diameter of veins,* which changes the venous pressure and thus the venous return and the cardiac output.

CARDIOVASCULAR CONTROL MECHANISMS

The cardiovascular control mechanisms which play their role in making circulatory adjustments during the routine and emergency cardiovascular stresses can be grouped as:

- Neural control mechanism,
- Humoral control mechanism and
- Local control mechanisms.

NEURAL CONTROL MECHANISMS

Neural regulation of circulation is of fundamental importance since it responds within seconds. The nervous regulation mainly controls systemic functions of circulatory system (whenever required) such as:

- Redistribution of blood flow to different parts of the body.
- Increasing pumping activity of heart.
- Rapid control of arterial pressure.

Components of neural control mechanism

The neural cardiovascular regulating mechanism consists of:

A. Medullary cardiovascular control centres. These are the prime centres concerned with neural control of circulation. These include:

- Medullary sympathetic centre (vasomotor centre),
- Medullary parasympathetic centre (nucleus ambiguus),
- Medullary relay centre for cardiorespiratory and afferents (nucleus of tractus solitarius, i.e. NTS).

B. Autonomic nervous system supplying the heart and blood vessels. The regulation of circulation by medullary control centres is exerted almost entirely through the autonomic nervous system (ANS). The sympathetic component of ANS is most important for controlling circulation and the parasympathetic component mainly contributes to the regulation of heart functions.

C. Afferent impulses to medullary centres. The vasomotor centre is influenced by afferent impulses from the higher centres and a large number of other areas.

D. Role of skeletal nerves and muscles in controlling blood pressure.

MEDULLARY CARDIOVASCULAR CONTROL CENTRES

1. VASOMOTOR CENTRE

Though popularly known as vasomotor centre (VMC), more appropriately it should be called as *medullary sympathetic centre.* It is the primary cardiovascular regulatory centre located in the medulla oblongata of brainstem. It consists of groups of neurons situated bilaterally in the reticular substance of medulla at the floor of fourth ventricle. The medullary cardiovascular centre is constituted by following different areas (Fig. 4.5-1):

Pressor area

Pressor area is located in the rostral ventrolateral medulla (RVLM). It contains glutaminergic neurons which exert *excitatory* effect on the thoracolumbar spinal sympathetic neurons.

Continuous sympathetic vasoconstrictor tone. Normally, the neurons forming pressor area of the VMC show *inherent tonic activity,* i.e. they discharge rhythmically (at a rate of about 1 impulse per second) in a tonic fashion to excite sympathetic preganglionic neurons present in the intermediolateral grey column of the spinal cord. In this way, the continuous signals are passed to the sympathetic vasoconstrictor nerves fibres over the entire body. When this tone is blocked, for example, by spinal anaesthesia, the blood vessels throughout the body dilate and arterial pressure may fall to as low as 50 mm Hg.

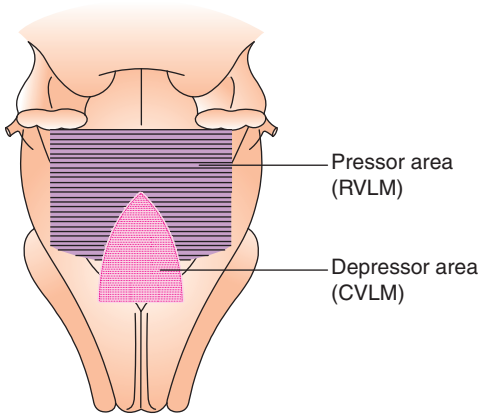


Fig. 4.5-1 Medullary cardiovascular control centres.

Stimulation of pressor area produces:

- *Arteriolar constriction,* which increases the systemic blood pressure.
- *Venoconstriction,* which decreases blood stored in the venous reservoir and increases venous return.
- *Increase in heart rate* or positive chronotropic effect.
- *Increase in force of contraction* or positive inotropic effect.

Depressor area

Depressor area is situated bilaterally in the *caudal ventrolateral medulla* (CVLM). Stimulation of neurons forming depressor area produces decrease in the sympathetic activity due to inhibition of the tonically discharging impulses of the pressor area causing:

- *Arteriolar dilation,* which decreases systemic blood pressure.
- *Venodilatation* which increases storage of blood in the venous reservoir and decreases venous return and cardiac output.
- *Decrease in heart rate* or negative chronotropic effect.
- *Decrease in force contraction* or negative inotropic effect.

2. MEDULLARY PARASYMPATHETIC CENTRE

Medullary parasympathetic centre or *cardiac vagal centre* (earlier also called cardioinhibitory centre) is now called by its specific name, i.e. the *nucleus ambiguus.* The neurons located in this centre are not tonically active. Nucleus ambiguus receives afferents via NTS and in turn sends inhibitory pathway in the form of vagal fibres to: *heart* to decrease the heart rate and force of cardiac contraction.

3. MEDULLARY RELAY STATION FOR CARDIORESPIRATORY AFFERENTS

NTS of the vagus nerve forms the so-called medullary relay station for the cardiorespiratory afferents. It receives afferents from most of the baroreceptors and chemoreceptors. Cells of the NTS, in turn, relay the information to vasomotor centre and cardiac vagal centre (nucleus ambiguus) that control sympathetic and parasympathetic outputs, respectively.

AUTONOMIC NERVE SUPPLY TO HEART AND BLOOD VESSELS

AUTONOMIC NERVE SUPPLY TO HEART

Sympathetic supply

Spinal sympathetic centre is formed by the neurons located in the intermediolateral horns of the spinal cord extending from T_1 to L_2 spinal segments.

- *Pre-ganglionic sympathetic fibres* (small, myelinated) supplying the heart arise from the neurons lying in the intermediolateral horns of the T_1–T_5 *spinal segments* and pass into the sympathetic trunk to superior, middle and inferior cervical ganglia and upper thoracic ganglia where they synapse (Fig. 4.5-2).
- *Post-ganglionic fibres* (long, unmyelinated) leave the ganglia and pass via superior, middle and inferior *cardiac sympathetic nerves* and supply to the nodal tissues [sinoatrial (SA) node and atrioventricular (AV) node] and the cardiac muscles (of atria and ventricles Fig. 4.5-2). It is important to note that:
  - Sympathetics from the right side are primarily distributed to the SA node.
  - Sympathetics from the left side are primarily distributed to the AV node.

Stimulation of cardiac sympathetic nerves causes:

- *Increased heart rate* (positive chronotropic effect),
- *Increase in the conduction of impulse* through heart (positive dromotropic action),
- *Increase in the excitability* of myocardium (positive bathmotropic effect) and
- *Increase in the force of contraction* of myocardium (positive inotropic effect).

Parasympathetic supply

Parasympathetic fibres to the heart are carried through two vagii (Fig. 4.5-3):

Preganglionic fibres (long, myelinated) arise from the *nucleus ambiguus* located in the medulla and travel along the vagus to reach the heart through their cardiac branches to synapse in the *ganglia* located within the superficial and deep cardiac plexuses and also in the walls of atria.

Postganglionic fibres (small, unmyelinated) are distributed to the atria, SA node, AV node and AV bundle. It is important to note that:

- The right vagus is distributed mainly to SA node.
- The left vagus is distributed mainly to AV node.
- No vagal motor fibres are distributed to the ventricles and
- Parasympathetic fibres to the heart are *endocardiac.*

Stimulation of parasympathetic fibres to heart causes:

- *Decrease in heart rate* (negative chronotropic effect).
- *Decrease in conduction of impulse* through the conduction tissue (negative dromotropic effect).
- *Decrease in the excitability of atria* only (negative bathmotropic effect).
- *Decrease in the force of contraction* of atria only (negative inotropic effect). There is no effect on the force of contraction of ventricles.

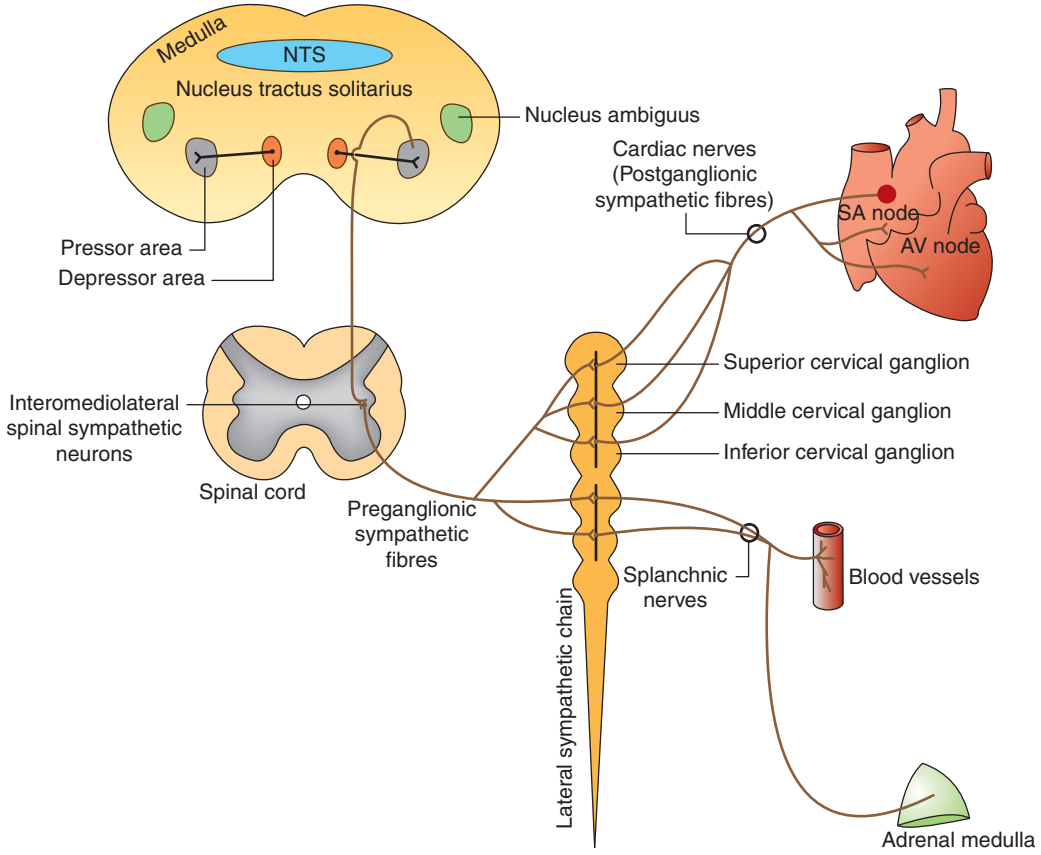


Fig. 4.5-2 Sympathetic innervation to heart and blood vessels.

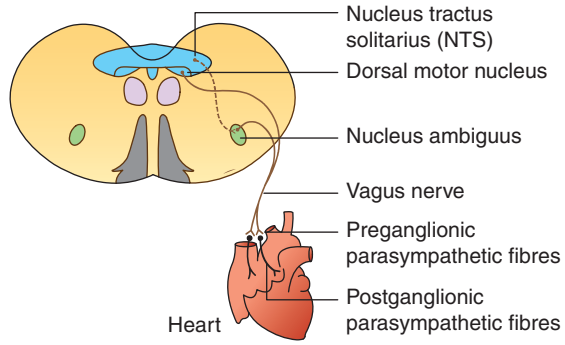


Fig. 4.5-3 Parasympathetic innervation of the heart.

Vagal tone. There is a good deal of tonic vagal discharge, called the vagal tone, in humans and other large animals. Therefore, when vagii are cut in the experimental animals, the heart rate rises. Similarly, in adult humans, the resting heart rate which is about 72 beats/min rises to 150–180 beats/min after the administration of vagolytic drugs, such as atropine because of the unopposed sympathetic tone. When both adrenergic and cholinergic systems are blocked in humans, the heart rate is approximately 100 beats/min. Since the resting heart rate is about 72/min, it confirms that at rest, the vagal tone is greater than the sympathetic tone.

AUTONOMIC NERVE SUPPLY TO BLOOD VESSELS

The autonomic efferents supplying the blood vessels produce two types of effects:

- Vasoconstriction and
- Vasodilation.

Vasoconstriction effect

Vasoconstriction effect is produced by the sympathetic fibres supplying the blood vessels which originate from the intermediolateral horns in T_1–L_2 spinal segments.

Vasoconstrictor fibres have norepinephrine and sometimes neuropeptide Y as neurotransmitter and are called *noradrenergic fibres.*

Sympathetic vasoconstrictor fibres show tonic (i.e. continuous) discharge at the rate of about 1 impulse/s. Therefore, when the sympathetic nerves are cut *(sympathectomy)*, there occurs:

- *Vasodilation* – which leads to decreased peripheral resistance → decreased diastolic blood pressure.
- *Venodilatation* – increased venous capacity → decreased venous return → decreased end-diastolic volume → decreased stroke volume → decreased cardiac output → decreased systolic blood pressure.

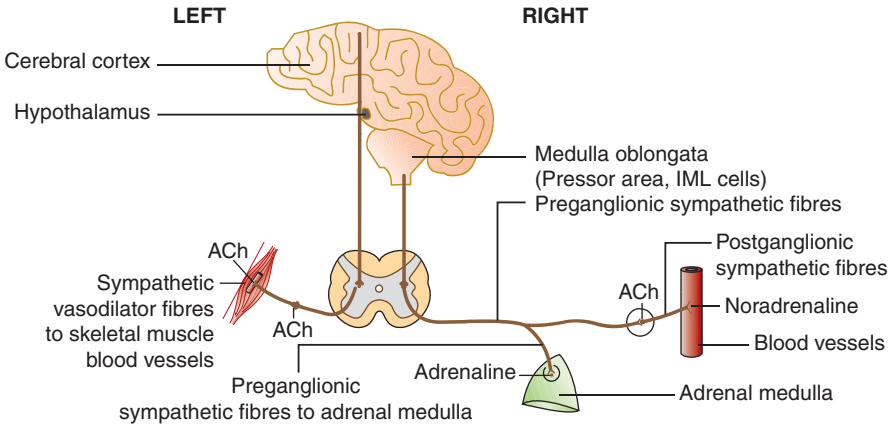


Fig. 4.5-4 Pathway of sympathetic vasodilator fibres (on left side) and sympathetic vasoconstrictor system (on right side).

Stimulation of sympathetic fibres produces:

- *Constriction of arterioles* – increased peripheral resistance → increased diastolic blood pressure and
- *Venoconstriction* – decreased venous capacity → increased venous return → increased end-diastolic volume → increased stroke volume → increased cardiac output → increased systolic blood pressure.

Both these mechanisms are responsible for the regional redistribution of blood and at the time of need the blood is diverted from the skin, skeletal muscles and splanchnic area to the heart and brain.

Vasodilation effect

Neural vasodilation effect on the blood vessels is produced by following mechanisms:

1. Decrease in discharge of noradrenergic vasoconstrictor nerves. In most tissues, vasodilation is produced by decreasing the rate of tonic discharge in the vasoconstrictor nerves.

2. Sympathetic cholinergic vasodilator nerves. Some of the organs of the body, such as skeletal muscles, heart, lungs, liver, kidney and uterus in addition to adrenergic vasoconstrictor sympathetic fibres also receive innervation by *cholinergic vasodilator sympathetic* fibres having acetylcholine and vasoinhibitory peptide (VIP) as neurotransmitter. These fibres originate from the cerebral cortex, relay in the hypothalamus and mid brain, and pass through the medulla (without relay in the VMC) to the sympathetic neurons located in the intermediolateral grey column of the spinal cord (Fig. 4.5-4). These fibres are not *tonically active* and get activated only in biological stresses, for example during exercise, child birth, etc. and help in increasing the blood flow.

3. Parasympathetic vasodilator nerves. Blood vessels, in general, do not have parasympathetic innervation with following exceptions:

- *Sacral outflow parasympathetic fibres* represented by *nervi erigentes,* which supplies sexual erectile tissue and is responsible for vasodilation in external genitalia during sexual excitement.
- *Cranial outflow of parasympathetic fibres* along chorda tympani branch of facial nerve to salivary glands.
- The post-ganglionic cholinergic neurons on the blood vessels contain acetylcholine, VIP and PHM-27 as neurotransmitters.

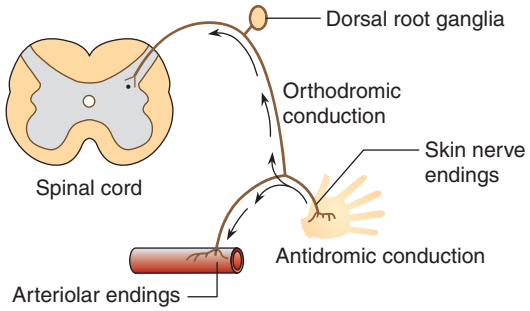


Fig. 4.5-5 Pathway of axon reflex.

Note. It is important to note that parasympathetic vasodilator fibres play little role in the control of general circulation. Activation of such nerves only contributes to pleasure and fulfilling important biological functions.

4. Vasodilation by axon reflex. Conduction of normal sensory afferent impulses from the skin to spinal cord is called *orthodromic conduction.* However, in certain situations, for example, when a firm stroke is applied across the skin, the afferent impulses in the sensory nerves from the skin are relayed *antidromically* down branches of the sensory nerves that innervate blood vessels (Fig. 4.5-5). The antidromic conduction of impulses causes release of substance P from the nerve endings which produces vasodilation and increases capillary permeability. This local neural mechanism (which does not involve the CNS) is called *axon reflex.* It is responsible for the local vasodilation and does not contribute in systemic control of circulation.

AFFERENT IMPULSES TO MEDULLARY CARDIOVASCULAR CONTROL CENTRES

The medullary control centres are influenced by afferent control impulses from the higher centres and a large number of other areas (Fig. 4.5-6). These include:

- Afferent impulses from higher centres controlling vasomotor centre.
- Afferent impulses from respiratory centres.
- Cardiovascular reflex mechanisms operating through medullary control centres
  - Baroreceptor reflex
  - Chemoreceptor reflexes
- Direct effects on vasomotor area
  - Central nervous system ischaemic response
  - Cushing reflex
- Afferents from nociceptive stimuli.

AFFERENT IMPULSES FROM HIGHER CENTRES CONTROLLING VASOMOTOR CENTRE AND CARDIAC VAGAL CENTRE

CEREBRAL CORTEX

There are descending tracts to the vasomotor area from the cerebral cortex (particularly the limbic cortex) that relay in the hypothalamus. Some examples of the influence of limbic system on the VMC are:

- *Tachycardia and hypertension* produced by emotions.
- *Bradycardia and fainting* occurring during sudden emotional shock.
- *Fight or flight response* is a complex set of response which increases cardiac output and raises blood pressure in anticipation of flight or physical defence.

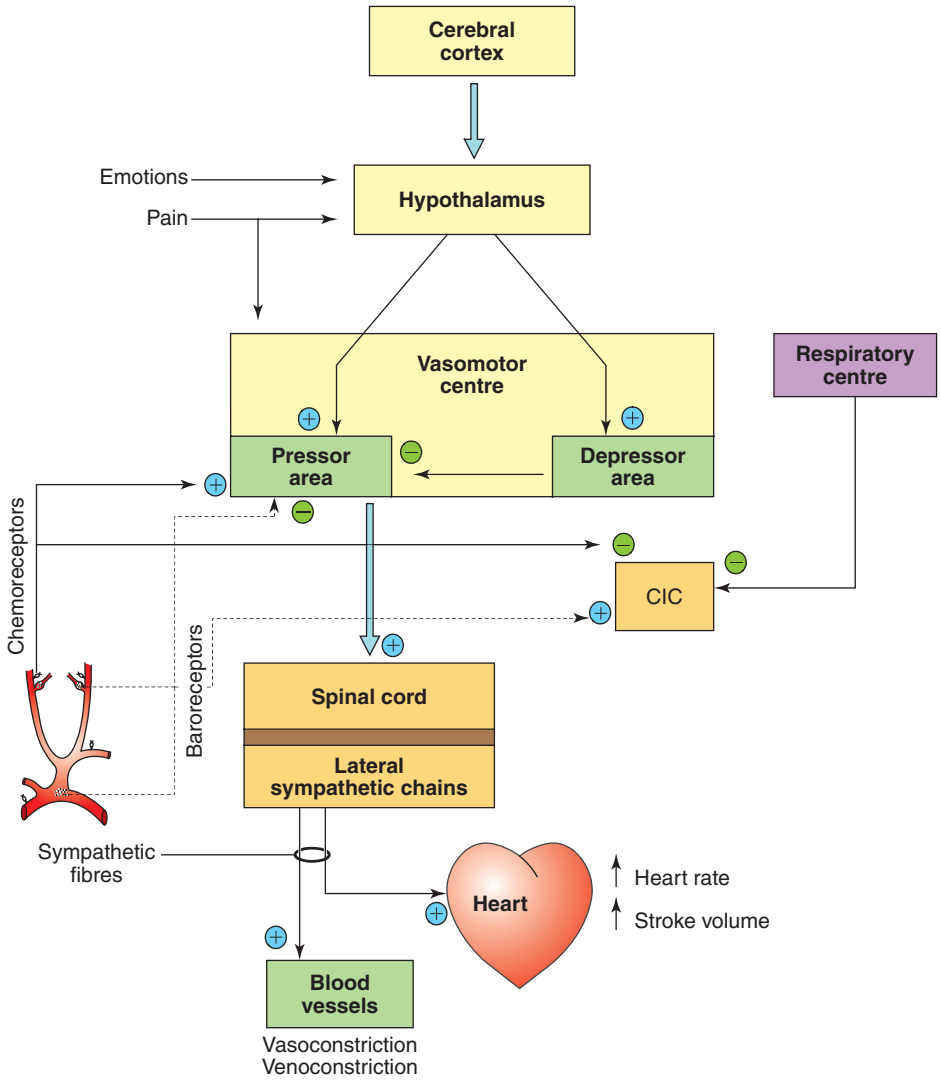


Fig. 4.5-6 Scheme to show afferent impulses affecting medullary cardiovascular control centres.

HYPOTHALAMUS

The hypothalamus serves to integrate many somatic and autonomic responses. Examples are:

- *Temperature regulation.* The effect of temperature changes on the hypothalamic centres is relayed to the medulla, which causes the vessels of skin to constrict (heat conservation) or to dilate (heat dissipation).
- *Emotional stresses* influence heart rate and blood pressure by impulses relayed from the hypothalamus to stimulate or inhibit the medullary centres.

RETICULAR FORMATION

Reticular formation of pons, mesencephalon and diencephalon also influences the vasomotor area, for example:

- *Pain* usually causes rise in the blood pressure via afferent impulses in the reticular formation converging on the vasomotor area. However, prolonged severe pain may cause vasodilation and fainting.

AFFERENT IMPULSES FROM RESPIRATORY CENTRES

Impulses arising from the respiratory centres affect the heart rate by changing the vagal tone and the alterations produced are known as *sinus arrhythmia* which occurs during forced breathing. Sinus arrhythmia is common in some children and in some adults even during quiet breathing.

- *During inspiration,* the impulses arising from the respiratory centres inhibit the cardiac vagal centre causing reduced vagal tone and *sinus tachycardia.*
- *During expiration,* the respiratory centres stop sending inhibitory impulses to the cardiac vagal centre causing increased vagal tone and sinus bradycardia.

CARDIOVASCULAR REFLEX MECHANISMS AFFECTING MEDULLARY CONTROL CENTRES

Cardiovascular reflex mechanisms are multiple subconscious special nervous control mechanisms that operate through medullary control centres all the time to maintain the arterial pressure within normal range include:

- Baroreceptor reflex mechanisms and
- Chemoreceptor reflex mechanism.

BARORECEPTOR REFLEX MECHANISMS

Baroreceptors, also known as mechanoreceptors or pressure receptors, are the stretch receptors located in the walls of heart and large blood vessels. These are spray-type nerve endings, i.e. they are extensively branched, knobby, coiled and intertwined ends of myelinated nerve fibres. These are stimulated by distension of the structures in which they are located and so they discharge at an increased rate when the pressure in these structures rises. The increased baroreceptor discharge leads to *inhibition* of tonic discharge of vasoconstrictor nerves and *excitation* of vagal innervation of heart and thereby produces vasodilation, venodilation, bradycardia, decrease in cardiac output and decrease in blood pressure.

With this definition of the baroreceptor reflex mechanism, the baroreceptors will be discussed in detail as under following headings:

CLASSIFICATION AND LOCATION OF BARORECEPTORS

Functional classification

Functionally, baroreceptors can be grouped as:

1. ***High-pressure baroreceptors,*** which monitor the arterial circulation. These include the baroreceptors located at:
   - Carotid sinus,
   - Aortic arch,
   - Wall of left ventricle,
   - Root of right subclavian artery and
   - Junction of the thyroid artery with common carotid artery.
2. ***Low-pressure baroreceptors*** are located in the low-pressure area of circulation and are collectively referred to as *cardiopulmonary receptors.* These include:
   - *Atrial receptors* scattered in the wall of right and left atrium.
   - *Baroreceptors located in the right atrium at the entrance* of the superior and inferior vena cavae and in the left atrium at the entrance of pulmonary veins.
   - *Pulmonary receptors located* in the wall of pulmonary trunk and its divisions into the right and left pulmonary artery.

Anatomical classification

Anatomically, baroreceptors can be grouped as:

1. ***Arterial baroreceptors,*** which are located in the walls of the arteries, distributed mainly in the adventitial layer.
2. ***Cardiac baroreceptors*** are located in the walls of heart subendocardially which include:
   (i) *Atrial receptors*
      - Atrial stretch receptors which are scattered throughout the wall of atria and interatrial septum.
      - Pulmonary venoatrial receptors, which are located in the left atrium just at the entrance of pulmonary veins.

   (ii) *Ventricular receptors,* which are scattered throughout the left ventricle and interventricular septum.

CAROTID AND AORTIC ARCH BARORECEPTORS

Location of carotid and aortic arch baroreceptors

Carotid baroreceptors are located in the carotid sinus which is a small dilatation of the internal carotid artery just above the bifurcation of the common carotid artery into external and internal carotid branches (Fig. 4.5-7).

Aortic arch baroreceptors are located in the wall of arch of aorta (Fig. 4.5-7).

Other systemic arterial baroreceptors (similar to carotid and aortic baroreceptors) are also found at the root of right subclavian artery and junction of thyroid artery and in the common carotid artery.

Innervation of baroreceptors (Fig. 4.5-8)

Carotid sinus baroreceptors are innervated by the carotid sinus nerve (Hering's nerve), which is a branch of glossopharyngeal nerve.

All other baroreceptors are supplied by the vagus nerve.

Afferent fibres from the baroreceptors pass via the glossopharyngeal and vagus nerves to the medulla. Most of them end in the NTS, where they secrete an excitatory transmitter, presumably, glutamate.

Buffer nerves. The carotid sinus nerve and vagal fibres from the carotid sinus and aortic arch baroreceptors, respectively are commonly called buffer nerves, as these are involved in buffering the blood pressure, i.e. preventing sudden rise and fall in the blood pressure.

Projections from NTS (excitatory glutaminergic projections) terminate on to the:

- *Depressor area of VMC,* where they stimulate GABA (gamma-aminobutyric acid)-secreting inhibitory neurons which decrease sympathetic activity by inhibiting the tonically discharging impulses from pressure area of VMC to sympathetic neurons of the spinal cord.

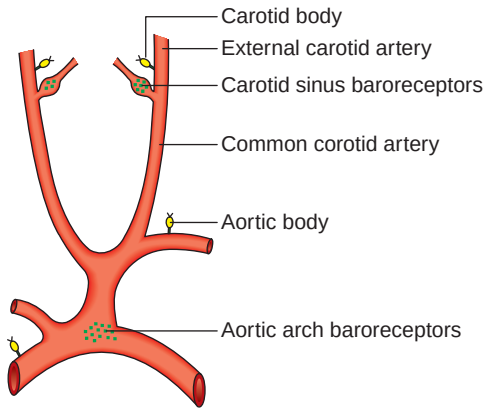


Fig. 4.5-7 Location of baroreceptors (carotid sinus and aortic arch) and chemoreceptors (carotid bodies and aortic bodies).

- *Cardiac vagal centre* (nucleus ambiguus), after receiving the impulses from NTS, *sends inhibitory* pathway along the vagus nerve to:
  - *Heart* (through cardiac branches of the vagus nerve to decrease heart rate and force of contraction).

Response of carotid and aortic baroreceptors to pressure

Response from carotid baroreceptors have been studied in detail. Salient features of these receptors responses to pressure are:

Baroreceptor response. At normal blood pressure levels, the fibres of the buffer nerves discharge at a low rate which increases when the pressure in the carotid sinus and aortic arch rises, and declines when the pressure falls (Fig. 4.5-9).

The effect of different arterial pressure levels on the discharge rate in carotid sinus nerve shown in Fig. 4.5-9 depicts that:

- The minimum pressure about 60 mm Hg at which carotid baroreceptors are stimulated is called *threshold of baroreceptor reflex.*
- Above threshold level, the baroreceptors respond progressively more rapidly till the discharge rate reaches *a plateau,* at 150–160 mm Hg, i.e. there is no further

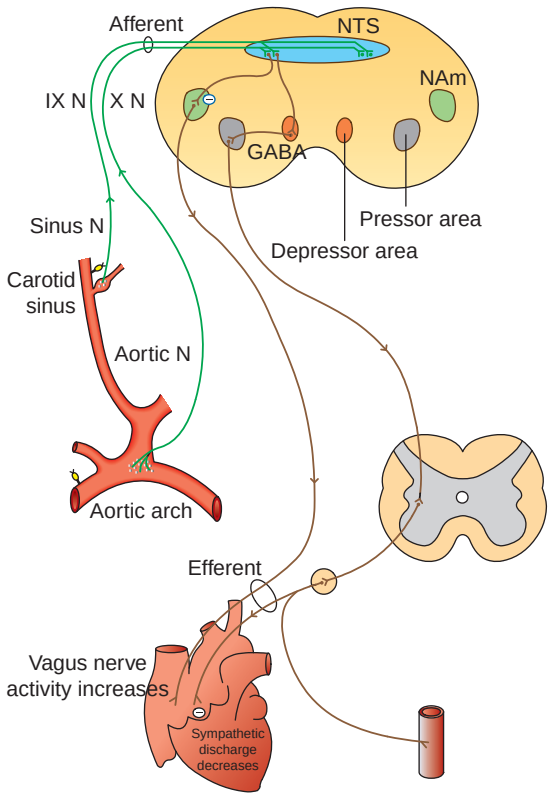


Fig. 4.5-8 Neural pathway of baroreceptor reflex.

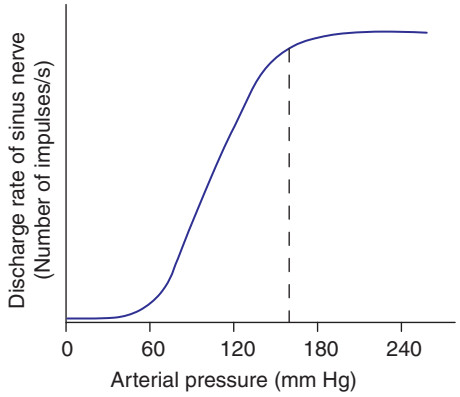


Fig. 4.5-9 Response of carotid baroreceptors at different levels of arterial pressure.

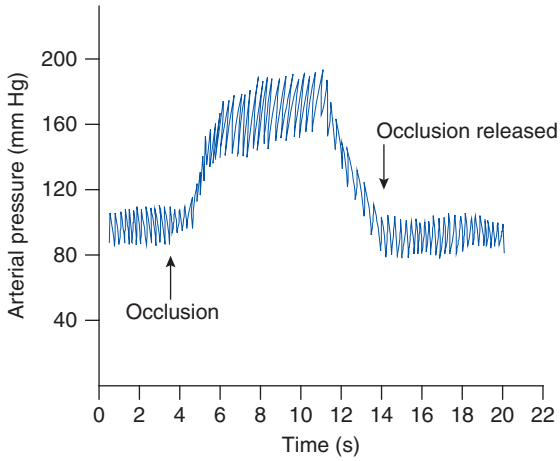


Fig. 4.5-10 Effect of occlusion of both common carotid arteries on arterial pressure in dog.

increase in response. Thus, the carotid baroreceptors exhibit a great sensitivity as they respond to pressure that varies from approximately 50–160 mm Hg.

- In the normal operating rate at 95–100 mm Hg, even a slight change in pressure causes a strong change in the baroreceptor reflex signals to readjust the arterial pressure back towards the normal.
- When pressure decreases below normal levels, the baroreceptor discharge decreases and reflexly brings the pressure to normal. Conversely, when pressure increases above normal, the baroreceptor discharge also increases and reflexly brings the pressure to normal. The effect of carotid receptors response to change in arterial pressure can be demonstrated experimentally as:

Bilateral occlusion of common carotid arteries at their origin reduces the carotid sinus pressure, as a result the carotid baroreceptors become inactive and lose their inhibitory effect on the VMC and the blood pressure is raised. Because aortic and cardiac baroreceptors respond to raised pressure so the occlusion of both carotid arteries cause only a moderate pressure response. When the common carotid occlusion is removed, the arterial pressure returns to normal (Fig. 4.5-10).

The carotid baroreceptors respond both to the mean pressure and the pulse pressure. Thus, the baroreceptor discharge would increase:

- When the mean pressure rises and the pulse pressure remains unchanged or
- When the pulse pressure rises and the mean pressure remains unchanged.

Carotid baroreceptors respond much more to a rapidly changing pressure than to a stationary pressure, i.e. if the mean arterial pressure is 150 mm Hg and at that moment is rising rapidly, the rate of impulse transmission may be as much as twice than that when the pressure is stationary at 150 mm Hg. Conversely, if the pressure is falling, the rate might be as little as one quarter that for the stationary pressure.

Pressure–buffer system of baroreceptors. From the above description it is clear that baroreceptor system opposes both increase as well as decrease in the arterial pressure. Therefore, it is called *a pressure–buffer system* and the nerves from the baroreceptors are called buffer nerves.

Baroreceptors resetting. Baroreceptors possess a property to reset themselves in 1–2 days to whatever pressure they are exposed. Therefore, in chronic hypertension, the baroreceptor reflex mechanism resets to maintain an elevated rather than a normal arterial pressure. Because of this property, the *baroreceptor system has no role to play for long-term regulation of the mean arterial pressure.* Thus, the baroreceptor reflex mechanism plays an important role only in preventing the extreme variations in blood pressure which occur for a short term.

APPLIED ASPECTS

Carotid sinus massage is used clinically to interrupt paroxysmal atrial tachycardia by inducing a vagally mediated slowing of the heart.

Stokes–Adams syndrome refers to an increased sensitivity of the carotid sinus seen in some elderly individuals who experience syncope as a result of vagally mediated sinus arrest, which causes a prolonged period of ventricular systole.

Effect of common carotid clamping and bilateral vagotomy

- Bilateral clamping of the common carotid arteries proximal to the carotid sinus lowers the pressure in the sinuses which is followed by a decline in discharge rate from the carotid baroreceptors leading to rise in the blood pressure and heart rate.

- When along with bilateral occlusion of the common carotid arteries, bilateral vagotomy is also performed, the blood pressure rises to 300/200 mm Hg or higher and is unstable.
- Bilateral destruction of NTS, the site of termination of the baroreceptor afferents, also causes a marked pressure response producing severe hypertension which can be even fatal.
- These forms of experimentally induced hypertension due to neurogenic lesions are called *neurogenic hypertension.*

CARDIAC BARORECEPTORS

Cardiac baroreceptors are located in the walls of heart subendocardially. All cardiac receptors are innervated by the vagus nerve. These include:

Atrial stretch receptors

Atrial stretch receptors present in the walls of atria are also called *low-pressure receptors.*

Types of atrial stretch receptors. Atrial stretch receptors have been studied in detail by Prof. A. S. Paintal (an Indian scientist) in 1953. These can be divided into following types:

1. Atrial stretch receptors with large myelinated afferent fibres. These receptors are:
- *Atriocaval receptors,* which are located in the right atrium just at the entrance of superior and inferior vena cavae.
- *Pulmonary venoatrial receptors,* which are located in the left atrium just at the entrance of pulmonary vein.

Depending upon the discharge pattern, the atrial stretch receptors are of three types:

(i) *Type A receptors* discharge during the atrial systole only and their impulse activity occurs in the PR interval of the electrocardiogram (Fig. 4.5-11).

(ii) *Type B receptors* discharge in the later part of the atrial diastole when the atria are distended with blood. That is, these receptors discharge just before the onset of atrial contraction and reach peak after T-wave of the electrocardiogram (Fig. 4.5-11). Type B receptor discharge increases when the venous return is increased.

(iii) *Intermediate type of receptors* discharge both during the atrial systole as well as the diastole. Therefore, their discharge pattern is characterized by type A receptors discharge followed by type B receptors discharge.

2. Atrial stretch receptors with non-myelinated afferent fibres. These receptors are scattered throughout the atria and the interatrial septum.

Role of atrial stretch receptors

The atrial stretch receptors have been associated with following roles in the cardiovascular control:

1. As low-pressure receptors, the atrial stretch receptors (especially type B receptors) along with pulmonary receptors play an important role to minimize arterial pressure changes in response to change in blood volume. Low-pressure receptors cannot detect the systemic arterial pressure, they do detect simultaneous increase in pressure in the low-pressure area of circulation caused by increase in volume, and they elicit reflexes parallel to the baroreceptor reflexes to make the total reflex system much more potent for control of arterial pressure. In other words, the atrial stretch receptors provide information about the circulating blood volume, i.e. greater the venous return, greater will be the discharge from the receptor fibres.

2. Atrial reflex control of heart rate (Bainbridge reflex). Bainbridge noted that sudden rise in the atrial pressure after rapid infusion of saline or blood in anaesthetised animals produced tachycardia, if the initial heart rate was low. This effect is known as Bainbridge reflex. Atrial stretch receptors may be responsible for this reflex. The afferent signals from these receptors pass through the vagus nerves to the medulla of brain. The efferent signals are transmitted back through both the vagal and the sympathetic nerves to increase the heart rate and force of contraction. Thus, this reflex helps to prevent damming of blood in the veins, atria and pulmonary circulation.

3. Atrial reflex control of blood volume (volume reflex). When there is volume overload the atrial stretch, receptors help to return the blood volume back towards normal by following mechanisms, which collectively are called volume reflex:

(i) Stretch of the atria causes very significant reflex dilatation of the afferent arterioles in the kidney leading to rise in glomerular capillary pressure with resultant increase in filtration of fluid into the kidney tubules.

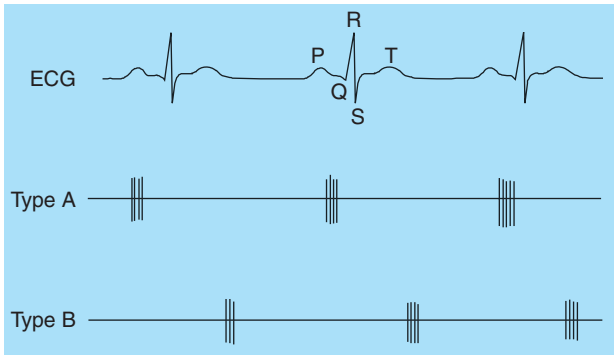


Fig. 4.5-11 Discharge pattern of type-A, and type-B atrial stretch receptors as recorded from vagus nerve and their correlation with electrocardiogram recording.

(ii) *Stretch of the atria also transmits signals to the hypothalamus* to decrease the secretion of antidiuretic hormone (ADH), which diminishes the reabsorption of water from tubules.

(iii) Stretch of the atria also causes release of a chemical *called atrial natriuretic peptide* (ANP) which causes powerful diuresis and thus blood volume back to normal.

The above described mechanisms (i–iii), which combined constitute volume reflex, act as a volume controller and thus indirectly act *as a pressure controller* as well. Because excess volume increases the cardiac output and thus the arterial pressure as well.

Ventricular receptors

The ventricular baroreceptors are scattered throughout the left ventricle and interventricular septum. They discharge irregularly and no physiological significance can be attached to these receptors.

Bezold–Jarisch reflex or coronary chemoreflex refers to the reflex apnoea followed by rapid breathing, hypotension and bradycardia which occur following injection of certain drugs like serotonin, veratridine or nicotine into the coronary arteries supplying the left ventricle (injection into the right coronary artery is ineffective) in experimental animals. This reflex is probably produced by the chemical stimulation of the left ventricular stretch receptors.

- *Physiological significance of* this reflex is uncertain, but it has been speculated that the persistent hypotension in some patients of acute myocardial infarction may be due to stimulation of the ventricular receptors by substances released from the necrotic cardiac tissue.

PULMONARY BARORECEPTORS

Pulmonary baroreceptors are located in the walls of pulmonary trunk and its divisions, the right and left pulmonary artery. The pulmonary receptors along with the atrial receptors constitute the so-called low-pressure *receptors or cardiopulmonary receptors* and play an important role to minimize arterial pressure changes in response to change in blood volume as discussed above (see page 258).

ROLE OF CHEMORECEPTOR REFLEXES IN CARDIOVASCULAR CONTROL

Chemoreceptors are chemosensitive cells that respond to following changes in blood:

- Oxygen lack (decreased PO_2),
- Carbon dioxide excess (increased PCO_2) and
- Hydrogen ion excess (decreased pH).

Location of chemoreceptor. The chemoreceptors are present in (Fig. 4.5-7):

1. *Carotid bodies.* These are 1–2 mm in size and are located in the bifurcation of each common carotid artery. These are innervated by carotid sinus nerve which is a branch of glossopharyngeal nerve.
2. *Aortic bodies* are one to three in number located adjacent to arch of aorta. These are innervated by aortic nerve (branch of vagus nerve).

Functions of chemoreceptors

1. *Respiratory control.* Chemoreceptors are primarily concerned with the regulation of pulmonary ventilation and are discussed in much more detail in Chapter 5.6, page 342.
2. *Cardiovascular control.* The chemoreceptors exert their role in cardiovascular regulation in following conditions:
   - *In hypoxia,* there occurs increased chemoreceptor discharge, which not only produces hyperventilation but also excites the VMC leading to peripheral vasoconstriction and increase in the arterial blood pressure. Thus, unlike the inhibitory action of arterial baroreceptors, the chemoreceptors have an excitatory effect on the VMC.
   - *In hypotension due to severe haemorrhage,* the increased chemoreceptor discharge may help to raise the arterial blood pressure.

Note. It is important to note that the chemoreceptors are not stimulated strongly until the arterial pressure falls below 60 mm Hg. Therefore, it is at lower pressures that this reflex becomes important and helps to prevent still further fall in pressure.

DIRECT EFFECTS ON VASOMOTOR AREA

The vasomotor centre is directly affected by locally produced hypoxia and hypercapnia. Examples of direct effects are central nervous system ischaemic response and Cushing reflex.

1. Central nervous system ischaemic response

- When blood pressure falls below 60 mm Hg, the blood flow to the vasomotor area in the brainstem is decreased enough to cause central nervous system (CNS) ischaemia.
- As a result of CNS ischaemia, the CO_2/lactic acid are accumulated locally near the VMC and excite the neurons of VMC strongly.
- Excitation of VMC causes strong sympathetic stimulation leading to vasoconstriction. There occurs immediate increase in the blood pressure. This most powerful response that activates sympathetic vasoconstrictor system strongly is called *CNS ischaemic response.* This acts as an emergency arterial pressure control system.

2. Cushing reflex

When intracranial pressure is increased and becomes equal to the arterial pressure, it compresses the arteries in the brain and blood supply to the vasomotor area is compromised. The hypoxia and hypercapnia produced locally increase the discharge from VMC. The resultant rise in the systemic pressure tends to restore the blood supply to medulla. This effect is called *Cushing reflex.* The resultant increase in blood pressure also causes reflex bradycardia via baroreceptor response. Thus, bradycardia is an important feature of raised intracranial pressure.

Afferents from nociceptive stimuli

Afferents carrying pain sensations also affect VMC and evoke either pressor or depressor reflex effect as:

Pressor effect in the form of an increase in the blood pressure and tachycardia is caused due to the sympathetic activity by *somatic pain afferents,* i.e. unmyelinated C-fibres which stimulate the pressor area of VMC.

Depressor effect in the form of hypotension and bradycardia, is produced by *visceral pain afferents,* i.e. thin myelinated fibres which synapse with depressor area of VMC and cause inhibition of sympathetic activity.

ROLE OF SKELETAL NERVES AND MUSCLES IN CONTROLLING BLOOD PRESSURE

1. Abdominal compression reflex

Whenever VMC is stimulated, e.g. by baroreceptor reflex or chemoreceptor reflex, other areas of reticular formation of brainstem are also stimulated along with. They send simultaneous impulses through the skeletal nerves to skeletal muscles of the body especially abdominal muscles. The contraction of abdominal muscles compresses the abdominal venous reservoirs increasing the venous return to heart and thereby the cardiac output. This response is called *abdominal compression reflex.*

2. Role of skeletal muscles during exercise

During exercise, the skeletal muscles especially that of limbs contract and compress the venous reservoirs. This causes translocation of large quantities of blood from the peripheral vessels into heart and lungs. This increases the cardiac output.

HUMORAL CONTROL MECHANISMS

Humoral regulation of circulation refers to the regulation by substances secreted into or absorbed into body fluids, e.g. hormones, ions, etc. Most important humoral factors affecting circulation are:

- Circulating vasodilators,
- Circulating vasoconstrictors and
- Ions and other chemical factors.

CIRCULATING VASODILATORS

The circulating vasodilators include:

- Kinins
- Vasoactive intestinal peptide
- Atrial natriuretic peptide (ANP)

KININS

Kinins are peptides which cause vasodilation. Two forms of kinins with similar action found are:

- ***Bradykinin*** is nonapeptide found in the plasma and
- ***Lysyl-bradykinin*** or kallidin is a decapeptide found in body tissues.

Synthesis and secretion

- The kinins are formed from high molecular weight kininogen (HMWK) and low molecular weight kininogen (LMWK) by the action of plasma and tissue kallikreins:

  HMWK $\xrightarrow{\text{Plasma kallikrein}}$ Bradykinin

  LMWK $\xrightarrow{\text{Plasma kallikrein}}$ Lysyl-bradykinin
- Lysyl-bradykinin can be converted to bradykinin by aminopeptidase.

  Bradykinin $\xrightarrow{\text{Aminopeptidase}}$ Lysyl-bradykinin

Functions of kinins

- They cause *vasodilation* by relaxing vascular smooth muscle (VSM) via nitric oxide (NO) and increase *capillary permeability.*
- Kinins play role in regulating blood flow especially to skin, salivary glands and GIT glands. Therefore, they are formed during active secretion in sweat glands, salivary glands and in exocrine portion of pancreas.
- By regulating blood flow to skin, the kinins probably play a role in thermoregulatory vascular adjustments.
- Kinins appear to be responsible for some episodes of vasodilation in patients with carcinoid tumours.
- Kinins are responsible for the inflammation because of their following actions:
  - Increase in vascular permeability especially of venules and capillaries leading to escape of plasma proteins into tissues,
  - Excitation of sensory nerve endings and production of pain (which is enhanced by 5HT) and
  - Attraction and migration of leucocytes from blood to tissues.

- Kinins cause contraction of visceral smooth muscles of ileum, uterus and bronchioles (leading to bronchoconstriction in patients with asthma).

VASOACTIVE INTESTINAL PEPTIDE (SEE PAGE 455)

ATRIAL NATRIURETIC PEPTIDE (SEE PAGE 616)

CIRCULATING VASOCONSTRICTORS

The circulating vasoconstrictors include catecholamines, angiotensin II and vasopressin.

CATECHOLAMINES

Catecholamines are released on the sympathetic stimulation and include:

Epinephrine. It stimulates both α and β adrenergic receptors:

- *Stimulation of* α*-receptors* results in vasoconstriction in skin and splanchnic areas.
- *Stimulation of* β*-receptors* results in dilation of the vessels in the skeletal muscles, liver, and coronary arteries.
- The β-receptor-induced vasodilation is more dominant than α-receptors-induced vasoconstriction. So, the net effect is slight lowering of peripheral resistance producing slight fall in diastolic blood pressure.
- β-receptor-induced increase in the stroke volume and heart rate results in higher cardiac output, a rise in the systolic blood pressure and widening of pulse pressure.

Norepinephrine. It has a generalized vasoconstrictor action as it has much greater effect on α than on β receptors.

- Therefore, it increases peripheral resistance and raises the diastolic blood pressure.
- Since it has negligible effect on β receptors, so direct cardiac stimulation is insignificant.

RENIN–ANGIOTENSIN SYSTEM

The renin–angiotensin system has important roles in the regulation of blood pressure and in the regulation of extracellular fluid volume.

Renin secretion and angiotensin formation

- *Renin,* a protease enzyme is secreted *by juxtaglomerular cells* of the kidney into the blood. Its secretion is stimulated by a decrease in the blood pressure.
- Renin catalyzes the conversion of *angiotensinogen* (α_2-globulin substrate present in the plasma) to angiotensin I.
- *Angiotensin I* is converted into angiotensin II by the action of *angiotensin converting enzyme* (ACE) present in the endothelium of blood vessels throughout the body, especially in the lungs and kidneys.

$$\text{Angiotensinogen} \xrightarrow{\text{Renin}} \text{Angiotensin I}$$

$$\text{Angiotensin I} \xrightarrow{\text{ACE}} \text{Angiotensin II}$$

Effects of angiotensin II

Angiotensin II has three principal effects by which it can elevate the arterial pressure:

1. Vasoconstriction. Angiotensin II is the most potent pressor substance being four to eight times more potent than norepinephrine. This effect of angiotensin II is important in the intermediate blood pressure control during circumstances, such as acute haemorrhage.

2. Decrease in salt and water excretion by kidney. Angiotensin II causes salt and water retention by the kidney. This action slowly increases extracellular fluid volume, which increases arterial pressure over a period of hours and days. Thus, this effect of angiotensin II plays an important role in the long-term control of arterial pressure.

Angiotensin II causes salt and water retention by the kidney in two ways:

(i) *By following direct actions on the kidneys:*
- Angiotensin II constricts the efferent arterioles which diminishes blood flow through the peritubular capillaries, allowing rapid osmotic reabsorption from the tubules.
- Angiotension II directly stimulates the epithelial cells of renal tubules to increase reabsorption of sodium and water.

(ii) *By stimulating secretion of aldosterone.* Angiotensin II stimulates the adrenal glands to secrete aldosterone which in turn increases salt and water reabsorption by the epithelial cells of the renal tubules.

3. Stimulation of thirst. Angiotensin II is a powerful stimulator of thirst. It leads to consumption of large volumes of water, leading to a rise in blood volume. This mechanism also plays some role in long-term control of blood pressure.

VASOPRESSIN

Vasopressin or ADH is secreted in minute quantities and therefore mainly affects water reabsorption in renal tubules. However, after a severe haemorrhage its concentration rises to a high level and then it has vasoconstrictor effect. For details see page 547.

IONS AND OTHER CHEMICAL FACTORS

The increased concentration of many different ions and chemical factors can also alter local blood flow by causing vasodilation or vasoconstriction.

- Calcium ions cause vasoconstriction,
- Potassium ions cause vasodilation,
- Hydrogen ions (decreased pH) cause vasodilation,
- Carbon dioxide causes vasodilation in most tissues and marked vasodilation in the brain,
- Glucose or other vasoactive substances, when increased in quantities, raise the osmolarity of blood and cause vasodilation.

LOCAL CONTROL MECHANISMS

Local cardiovascular control mechanisms are primarily concerned with the control of blood flow to the tissues locally. The ability of the tissues to regulate their own blood flow locally serves many *functions,* a few examples are:

- Local control of blood flow permits the tissues to maintain adequate nutrition and perform necessary functions in maintaining homeostasis. In general, the blood flow to the tissues is related to the rate of metabolism of the organ, greater the rate of metabolism, greater is the blood flow.
- These mechanisms help in temporarily curtailing blood flow to some organs so as to divert more blood to a metabolically active organ.
- Control of blood flow to skin helps in control of body temperature.

CLASSIFICATION OF LOCAL CONTROL MECHANISMS

A. Mechanisms involved in acute control of blood flow.
B. Mechanisms involved in long-term blood flow regulation.

MECHANISMS INVOLVED IN ACUTE CONTROL OF BLOOD FLOW

Acute control occurs within seconds to minutes through constriction or dilation of arterioles, meta-arterioles, and precapillary sphincters. The mechanisms involved in acute control of blood flow include:

- General mechanisms and
- Special mechanisms.

I. GENERAL MECHANISMS

These are the mechanisms that are present in most tissues of the body are:

1. Autoregulation, i.e. control of flow during changes in the arterial pressure,
2. Role of local vasodilator metabolites and factors,
3. Role of local vasoconstrictors and
4. Role of substances secreted by the endothelial cells.

1. Autoregulation (control of flow during changes in arterial pressure)

Autoregulation is the ability of an organ or tissue to adjust its vascular resistance and maintain a relatively constant blood flow over a wide range of arterial pressure. For details see page 236.

2. Role of local vasodilator metabolites and other factors

The accumulation of local vasodilator metabolites increases local blood flow. The greater the rate of metabolism in the tissue, the greater the rate of production of tissue metabolites. These include:

- *Decrease in O_2 tension and pH* causes vasodilation in most tissues. These changes cause relaxation of the arterioles and precapillary sphincters.
- *Increase in pCO_2 and osmolality* also dilates the vessels. The direct dilation action of CO_2 is most pronounced in the skin and brain.
- *Rise in temperature* exerts a direct vasodilator effect and the temperature rise in the active tissues (due to the heat of metabolism) may contribute to vasodilation.
- *Potassium (K^+) and lactate ions* are other substances that accumulate locally and play role in vasodilation especially in skeletal muscles.
- *Histamine* released from the damaged cells in injured tissues increases capillary permeability.
- *Adenosine* may play a vasodilator role in the cardiac muscles but not in the skeletal muscles. It also inhibits the release of norepinephrine.

Local vasodilator metabolites increase blood flow during following conditions:

Active hyperaemia refers to the vasodilation which occurs when the tissue metabolic rate increases. The dilation of local blood vessels helps the tissues to receive the additional nutrients required to sustain its new level.

Metabolic theory of autoregulation states that any vasodilator metabolites which accumulate in the tissues during active metabolism will produce autoregulation. When blood flow decreases, they accumulate and the vessels dilate; when blood flow increases, they are washed away.

Reactive hyperaemia is a phenomenon by which the local blood flow to the organ is controlled after a period of ischaemia. This phenomenon also appears to be a manifestation of local metabolic blood flow regulation mechanisms. After vascular occlusion, there occurs accumulation of

tissue vasodilator metabolites and the development of oxygen deficiency in the tissues.

3. Role of localized vasoconstrictors

Serotonin released from platelets in the injured tissue is responsible in part for the vasoconstriction which occurs in haemostasis.

Decrease in tissue temperature causes vasoconstriction and this local response to cold plays a part in temperature regulation.

4. Role of substances released by endothelium

Vascular endothelial cells make up a large and important organ. These cells secrete many growth factors and vasoactive substances which play an important role in the local control of blood flow. The vasoactive substances include:

- Prostaglandins and thromboxane A_2,
- Endothelium-derived relaxing factor (EDRF) and
- Endothelins.

(i) Prostaglandins and thromboxane A_2

- *Prostacyclin* is prostaglandin produced by the endothelial cells from arachidonic acid via cyclooxygenase pathway. It inhibits platelet aggregation and promotes vasodilation.
- *Thromboxane A_2* is produced by platelets also from arachidonic acid. It promotes platelet aggregation and vasoconstriction.
- Balance between prostacyclin and thromboxane A_2 fosters localized platelet aggregation and consequent clot formation while preventing excessive extension of clot and maintaining blood flow around it.

APPLIED ASPECTS

The prostacyclin–thromboxane A_2 balance can be shifted towards prostacyclin by administering low doses of aspirin. Aspirin produces irreversible inhibition of cyclooxygenase. Obviously, this reduces production of both prostacyclin and thromboxane A_2. However, endothelial cells produce new cyclooxygenase in a matter of hours whereas platelets cannot manufacture the enzyme, and the level rises only as new platelets enter the circulation. This is slow process because platelets have a half-life of about 4 days. Therefore, administration of small amounts of aspirin for prolonged periods reduces clot formation and has been shown to be of value in preventing myocardial infarction, unstable angina, transient ischaemic attacks and stroke.

(ii) Endothelium-derived relaxing factor

EDRF is the name given to a substance which is released by vascular endothelial cells and produces vasodilation. Later on, it was identified to be *nitric oxide (NO)* in chemical structure.

Mechanism of vasodilation by NO. The NO that is synthesized in the endothelium diffuses to smooth muscle cells, where it activates soluble guanylyl cyclase, producing cyclic GMP, which in turn mediates the relaxation of VSM by decreasing intracellular Ca^{2+} concentration (Fig. 4.5-12).

Relaxation of vascular smooth muscle produced by NO serves various functions in different circumstances:

- *Flow-induced vasodilation* is thought to occur due to local release of NO. When flow to a tissue is suddenly increased by arteriolar dilation, the large arteries to the tissue also dilate as during physical exercise.
- *Tonic release of NO* under basal physiological conditions is necessary to maintain normal blood pressure.
- *Penile erection* which is consequent to vasodilation and engorgement of corpora cavernosa is also thought to be produced by release of NO.

(iii) Endothelins (ET)

Endothelins are family of three similar polypeptides: ET-1, ET-2 and ET-3. Endothelin-1 produced by the endothelial cells is the most potent vasoconstrictor agents.

Endothelin-1 are secreted into the blood, but most of the endothelin-1 is secreted into the tunica media of the blood vessels and act in a paracrine fashion.

Factors affecting endothelin-1 secretion. *Stimulators* of endothelin-1 secretion are angiotensin II, catecholamines, growth factors, hypoxia, insulin, oxidized LDL, HDL, shear stress and thrombin. *Inhibitors* of endothelin-1 secretion include NO, ANP, PGE and prostacyclin.

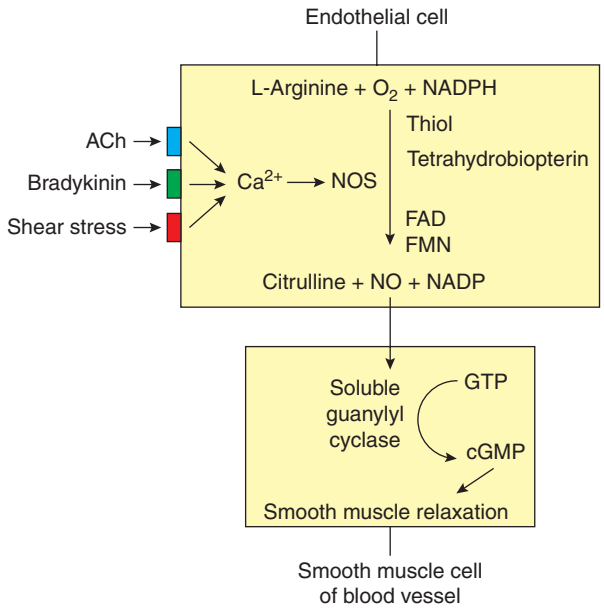


Fig. 4.5-12 Synthesis and mechanism of action of endothelium-derived relaxing factor (EDRF).

Mechanism of action. Endothelin-1 exerts its vasoconstrictor effect through the ET_A receptor which is found predominantly on VSM. ET_B receptor is found on the endothelial cells and responds to all the three endothelins (ET-1, ET-2 and ET-3). When activated ET_B receptor stimulates release of NO and thus favours vasodilation.

II. SPECIAL MECHANISMS

In addition to the above described general mechanisms, there are special mechanisms that control blood flow in special areas. These are discussed in relation to specific organs, but the following two are notable mechanisms:

1. Tubuloglomerular feedback mechanism in kidneys

In the kidneys, blood flow control is by a special mechanism called *tubuloglomerular feedback,* in which the composition of fluid in the early distal tubule is detected by the macula densa. When too much fluid filters from the blood through the glomerulus into the tubular system, feedback signals from the macula densa cause constriction of the afferent arterioles, thereby reducing renal blood flow and glomerular filtration rate (GFR) back toward normal.

2. Role of concentration of CO_2 and hydrogen controlling blood flow to brain

In the brain, the concentrations of CO_2 and H^+ play prominent roles in local blood flow control. An increase in CO_2 and H^+ dilates the cerebral blood vessels, which allows rapid washout of the excess CO_2 and H^+ ions.

MECHANISM INVOLVED IN LONG-TERM BLOOD FLOW REGULATION

The long-term blood flow regulation develops over a period of days to months to match the metabolic needs of the tissues. Long-term blood flow regulation is required by:

- Ischaemic tissues,
- Tissues that are growing rapidly and
- Tissues that become chronically hyperactive.

The long-term blood flow regulation is brought by an increase in the physiological size of the vessels in a tissue and in certain circumstances even by an increase in the number of blood vessels. One of the major factors that stimulate the increased vascularity of the tissues is a low oxygen concentration.

ANGIOGENESIS

The growth of the new vessels is called angiogenesis.

Angiogenic factors. These are the substances which are responsible for angiogenesis. Three of the best characterized angiogenic factors which have been isolated from tumours or from other tissues that are rapidly growing or have inadequate blood supply are:

- Vascular endothelial growth factor,
- Fibroblast growth factor and
- Angiogenin.

Development of collateral blood vessels. Collateral blood vessels refer to those new vessels which develop around a blocked artery or vein and allow the affected tissue to be at least partially resupplied with blood. An important example is the development of collateral blood vessels after thrombosis of one of the coronary arteries in old people.

Chapter 4.6

Regional Circulation

INTRODUCTION

After discussing the 'dynamics of circulation' and 'cardiovascular regulation mechanisms', it will be worthwhile to know how these basic principles apply to circulation in various regions of the body. This chapter includes:

- Coronary circulation,
- Cerebral circulation,
- Cutaneous circulation,
- Skeletal muscle circulation and
- Splanchnic circulation

Circulation to other regions, such as pulmonary circulation and renal circulation have been described in the concerned sections.

CORONARY CIRCULATION

CORONARY BLOOD VESSELS

Coronary arteries

Two coronary arteries (right and left) arise from the root of ascending aorta and supply blood to the myocardium (Fig. 4.6-1).

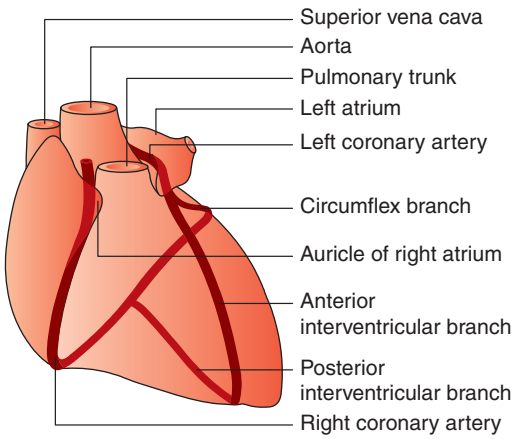


Fig. 4.6-1 Major coronary arteries and their branches.

Right coronary artery supplies blood to the right ventricle, the right atrium, the posterior part of left ventricle, the posterior part of interventricular septum and major portion of the conducting system of heart including SA node.

Left coronary artery supplies blood mainly to the anterior part of left ventricle, left atrium, anterior part of the interventricular septum and a part of the left branch of bundle of His.

Predominant supply by the right coronary artery described above is seen in about 50% individuals. In 20% individuals the predominant supply to myocardium is by left coronary

artery. In 30% individuals it is the balanced supply, i.e. equal supply by the two arteries.

Major coronary arteries (i.e. right coronary artery and its main branch posterior interventricular branch and left coronary artery and its main branches the circumflex artery and anterior interventricular artery) travel in the epicardium of heart *(superficial vessels)* (Fig. 4.6.1) and subdivide sending *penetrating branches* through the myocardium. The penetrating branches subdivide into arcades that distribute blood to the myocardium.

End arteries. Normally, the coronary arteries appear to function as *end arteries.* However, the presence of an arterial plaque or occlusion allows the *anastomoses* present between vessels to become functional. These anastomoses are of two types:

Cardiac anastomoses are those which are present between branches of two coronary arteries and between the branches of coronary artery and deep venous system.

Extracardiac anastomoses include those present between the branches of coronary arteries and vessels lying near the heart, such as vasa vasora of aorta, vasa vasora of pulmonary arteries, intrathoracic arteries, bronchial arteries and phrenic arteries.

Coronary veins

Coronary sinus is a wide vein about 2 cm long, which drains most of the venous blood from the myocardium (mainly left ventricle) into the right atrium. Its tributaries are the *great cardiac vein,* the *small cardiac vein, the posterior vein of left ventricle* and the *oblique vein of left ventricle* (Fig. 4.6-2).

Anterior cardiac vein draining venous blood mainly from the right ventricle opens directly into the right atrium.

Thebesian veins and *coronary-luminal vessels* (connections between the coronary vessels and the lumen of heart) constitute the deep venous system. These vessels drain only less than 10% of the venous blood from the myocardium directly into the various cardiac chambers, contributing to an *anatomic shunt effect.* The coronary luminal connections carry a larger proportion of the flow in the right ventricle than in the left ventricle.

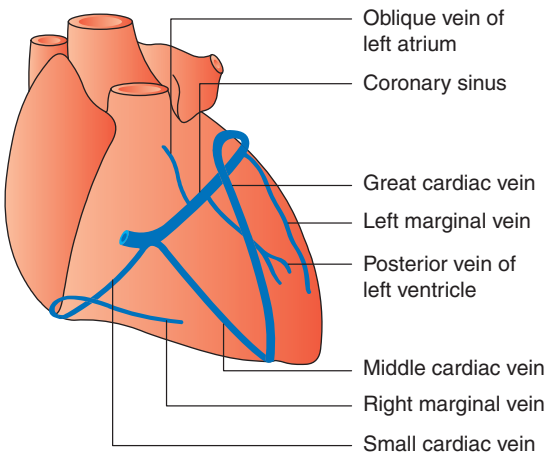


Fig. 4.6-2 Coronary sinus and its tributaries.

CORONARY BLOOD FLOW: CHARACTERISTIC FEATURES

Normal coronary blood flow and oxygen demand

A continuous flow of blood to the heart is essential to maintain an adequate supply of O_2 and nutrients.

Normal coronary blood flow at rest is about 250 mL (70 mL/100 g tissue/min), i.e. about 5% of the resting cardiac output (5 L). Three to six fold increase in the coronary blood flow may occur during exercise.

Oxygen consumption by the myocardium is very high (8 mL/min/100 g at rest). Because of this, even at rest 70–80% of the oxygen is extracted from each unit of the coronary blood as compared to the whole body (average of 25%) oxygen extraction at rest. The increased oxygen demand of the myocardium during exercise is met with by almost total (nearly 100%) extraction of oxygen and by manifold increase in the coronary blood flow. Oxygen supply and utilization by myocardium vis-a-vis rest of the body (average) is shown in Table 4.6-1.

Phasic changes in coronary blood flow

The coronary blood flow shows changes during phases of the cardiac cycle. The blood flow is determined by the balance between *pressure head* (i.e. aortic pressure) and the resistance (i.e. extravascular pressure exerted by the myocardium on the coronary vessels) offered to the blood flow during various phases of cardiac cycle is shown in Table 4.6-2 and Fig. 4.6-3 and described here:

Table 4.6-1 Oxygen supply and consumption by myocardium vis-a-vis rest of the body (average)

Parameter	Rest of the body (average)	Myocardium
Oxygen content – arterial blood – venous blood	 19 mL% 14 mL%	 19 mL% 06 mL%
A–V O_2 difference	5 mL%	13 mL%
Coefficient of O_2 Utilization	5/19 × 100 = 26%	13/19 × 100 = 70%
Oxygen saturation of venous blood	14/19 × 100 = 70% with PO_2 40 mm Hg	6/19 × 100 = 35% with PO_2 < 20 mm Hg

Table 4.6-2 Pressure gradients between ventricles and aorta during systole and diastole

Phase of cardiac cycle	Pressure (mm Hg)			Pressure gradient (mm Hg) between aorta and ventricle	
	Right ventricle	Left ventricle	Aorta	Right ventricle	Left ventricle
Systole	25	121	120	95	–1
Diastole	0	0	80	80	80

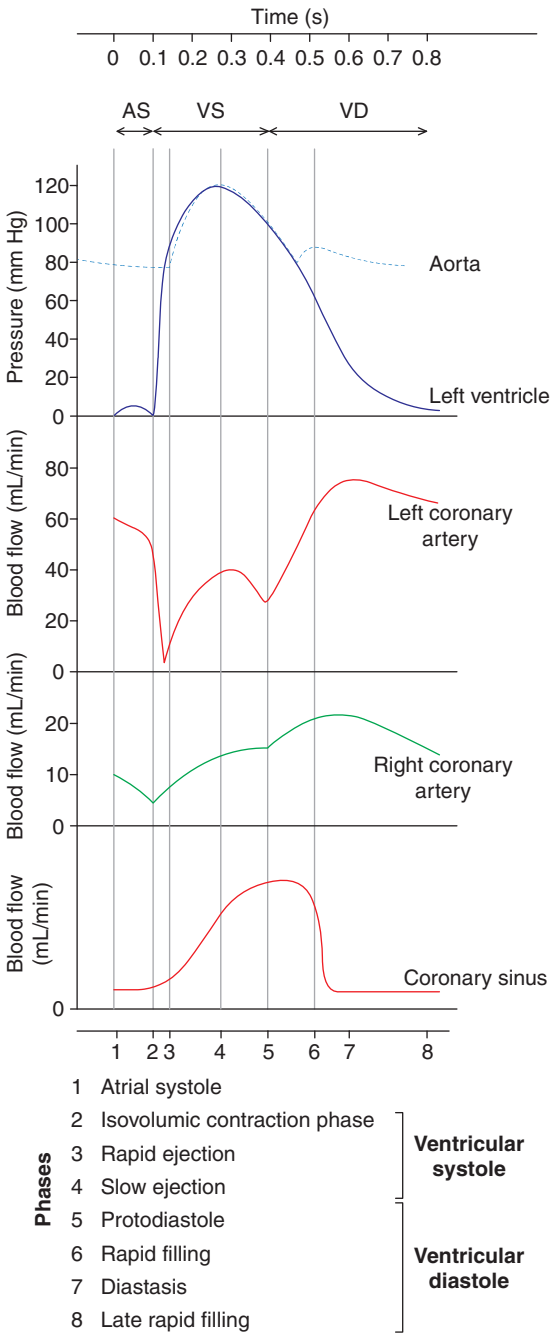


Fig. 4.6-3 Blood flow in right and left coronary arteries and coronary sinus during different phases of cardiac cycle.

Blood flow to left ventricle

During systole, the tension developed in the left ventricle is so high that it has throttling effect on the branches of the coronary arteries penetrating through them. As a result, the average blood flow through the capillaries of left ventricles falls to the extent that during isometric contraction phase, the blood flow to the left ventricle practically ceases, i.e. becomes zero.

During diastole, the cardiac muscles relax and blood flow increases. Thus, most of the coronary blood flow (over 70%) occurs during diastole (Fig. 4.6-3). In severe tachycardia, the duration of diastole is drastically reduced. This tends to reduce the coronary blood flow during diastole as well, but due to local metabolic regulation the blood flow to the myocardium is not seriously affected.

Blood flow to right ventricle and atria

Blood passing through coronary capillaries of right ventricle also shows phasic changes similar to the left ventricle. However, the changes in the right ventricular flow are far less because force of contraction of the right ventricle is much less (Table 4.6-2). Thus, the blood flow to the right ventricle and atria occurs both during systole and diastole.

Blood flow through coronary sinus

As shown in Fig. 4.6-3, in the coronary sinus the inflow of blood gradually rises from the *isovolumic ventricular contraction phase* and reaches its peak during *protodiastole phase* and then gradually falls.

Clinical importance of phasic coronary blood supply

1. Subendocardial region of the left ventricle receives no blood supply during systole so this region is particularly vulnerable to ischaemia and is the *most common site of myocardial infarction.* This is true in spite of the fact that this region has been provided with following *compensatory (protective) mechanisms:*

Capillary density in subendocardial region of left ventricle is much higher (1100 capillaries/mm^2) than the epicardial region (750 capillaries/mm^2). Therefore, during diastole, flow to the subendocardial region of the left ventricle is considerably higher.

Minimum diffusion distance between the capillaries and myocardial cells is 20% shorter in the subendocardial region of left ventricle (16.5 µm) as compared to the epicardial region (20.5 µm).

Myoglobin content (O_2 storage pigment) is higher in the subendocardial region than the epicardial region of the left ventricle.

2. In aortic stenosis, pressure in the left ventricle is much higher than that in aorta because the ventricle has to force the blood against a narrow aortic orifice. This leads to severe compression of coronary vessels during systole and thus chances of myocardial infarction are increased in such cases.

3. In congestive heart failure (CHF), increase in venous pressure decreases aortic diastolic pressure. As a result, the effective coronary perfusion pressure falls and coronary blood flow decreases.

MEASUREMENT OF CORONARY BLOOD FLOW

1. Nitrous oxide method (Kety method)

Principle. Nitrous oxide method is the most common method used for measuring coronary blood flow. It gives almost accurate value and is based on the Fick's principle (see page 216).

Procedure. The individual is made to inhale a mixture of 15% nitrous oxide and air for 10 min.

- During inhalation of gases, serial samples of arterial and coronary sinus venous blood (through a catheter introduced) are taken at fixed intervals for 10 min.
- The coronary blood flow (CBF) is then determined from the amount of nitrous oxide taken up per minute (N_2O/min) and the difference of nitrous oxide content of arterial (A) and venous (V) blood, i.e.

$$CBF = \frac{N_2O \text{ taken up/min}}{(A - V)}$$

2. Radionuclides utilization technique

Principle. The radioactive tracers are pumped into cardiac muscle cells by the enzymes Na^+–K^+ ATPase and equilibrate with the intracellular K^+ pool. Distribution of radioactive tracers is directly proportional to myocardial blood flow and this forms the basis of this technique.

Procedure. Radionuclide such as thallium-201 ($^{201}T_1$) is injected intravenously. After 10 min, the amount of $^{201}T_1$ taken up by the myocardial cells is then measured with the help of gamma-scintillation camera over the chest. The amount of coronary blood flow is calculated from these values. Areas of ischaemia are detected by their low uptake. Some radiotracers such as technetium-99m stannous pyrophosphate (^{99m}Tc-PYP) are selectively taken up by the infarcted tissues only by an unknown mechanism. These substances are used to detect areas of myocardial infarcts which stand out as *hot spots* on the scintiscans of the chest.

3. Coronary angiographic technique

Coronary angiography when combined with measurement of ^{133}Xe washout using a multiple crystal scintillation camera provides detailed analysis of coronary blood flow.

4. Electromagnetic flowmeter technique

- This technique is employed in animals to measure the coronary blood flow. The main advantages of this technique are that it tells the phasic flow and the flow per minute.
- In this technique, blood flow through the left ventricle is determined with the help of electromagnetic flowmeter implanted around the main left coronary artery or around its circumflex branch. The arterial and venous (with the help of a catheter passed into coronary sinus) blood samples are analyzed for the O_2 content. From the flow measured and the difference of O_2 content of arterial and venous blood, the myocardial consumption of O_2 is then determined directly.

REGULATION OF CORONARY BLOOD FLOW

I. Local control mechanism

1. Autoregulation. Like other vital organs of the body, coronary circulation shows well developed phenomenon of autoregulation, as described in detail on page 236. However, this phenomenon of autoregulation of coronary blood flow fails when blood pressure falls below 70 mm Hg and coronary perfusion is seriously compromised.

2. Role of local metabolites. Metabolic local factors are the most important factors which regulate the coronary blood flow.

- It is important to note that under resting state about 50–70% of O_2 is released to the myocardium from the haemoglobin of arterial blood (in contrast, haemoglobin releases about 25% of its O_2 content for the body as a whole). Therefore, not much additional oxygen can be provided to myocardium unless the blood flow increases. The O_2 consumption regulates the coronary blood flow probably by the following mechanisms:
- *Role of adenosine* (Berne's hypothesis). Adenosine is considered the major factor in production of coronary vasodilation during hypoxic states. In myocardial ischaemia, either due to generalized hypoxia or due to increased myocardial metabolism the intracellular myocardial *adenine nucleotides* are degraded to *adenosine*. The adenosine is capable of crossing myocardial cell membrane and thus comes out in the extracellular fluid and gains access to the resistance vessels including the precapillary sphincters of the coronary system producing an extremely strong vasodilator response (Fig. 4.6-4). The adenosine is then reabsorbed back into the cardiac cells to be reused.
- *Direct effect of O_2*. It has been proposed that a decrease in the tissue PO_2 could also act directly on the arterioles and cause vasodilation.
- *Role of other local metabolites.* Hydrogen ions, bradykinin, CO_2 and prostaglandins are the other suggested vasodilator substances.

3. Role of endothelial cells

- Endothelial cells release several *vasodilator* autacoids that contribute to the physiologic regulation of coronary

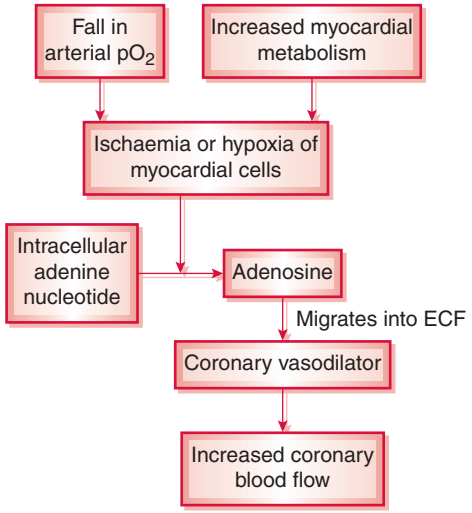


Fig. 4.6-4 Berne hypothesis (adenosine mechanism) of increase in coronary blood flow.

vasomotor tone. These include endothelium-derived relaxing factor (EDRF), prostacyclin and endothelium-derived hyperpolarizing factor.

- Endothelial cells also release *vasoconstrictor autacoids* that may have a pathologic role, such as endothelin-1 (ET-1), angiotensin II and endothelium-derived contracting factors.

II. Nervous control mechanism

Autonomic nerves control the coronary blood flow directly as well as indirectly.

1. Direct nervous control

Direct nervous control on coronary circulation is exerted through sympathetic and parasympathetic nerve supply to the coronary vessels.

Parasympathetic nerve fibres to coronary vessels through vagus are so less that the parasympathetic stimulation has very little direct effect, causing vasodilation.

Sympathetic nerve fibres extensively innervate the coronary vessels. The transmitters released at their nerve endings are epinephrine and norepinephrine. The coronary vessels contain both α and β receptors. The net result of direct effect of sympathetic stimulation is vasoconstriction.

2. Indirect nervous control

Indirect control of nervous stimulation on the coronary blood flow is through their action on the heart.

Sympathetic stimulation causes increase in the heart rate and increase of force of contraction of the heart. Thus, an increased activity of heart helps conversion of ATP to ADP which by producing coronary vasodilation increases the coronary blood flow.

Parasympathetic stimulation causes decreased heart rate and decreased force of contraction of heart. Thus, indirectly the coronary blood flow is reduced.

3. Neurohumoral control factors

Several nonadrenergic noncholinergic neurotransmitters also play a modulatory role, which include:

- ATP (purine) which is co-released with norepinephrine from the nerve terminals produces vasoconstriction through P_1 receptors and vasodilation through P_2 receptors present on the vascular smooth muscles.
- *Neuropeptide-Y* (NPY) is released with norepinephrine during sympathetic stimulation and causes severe vasoconstriction.
- *Calcitonin gene-related peptide* and substance P which are also found in cardiac nerves cause release of EDRF and produce maximal dilation of epicardial coronary arteries.

FACTORS AFFECTING CORONARY BLOOD FLOW

1. Mean aortic pressure. This is the force for driving blood into the coronary arteries. Rise in mean aortic pressure increases the blood flow and vice versa.

2. Muscular exercise. Normal CBF at rest is about 70 mL/100 g tissue/min. During exercise, CBF increases about four times because of sympathetic stimulation by the following mechanisms:

- Increased activity of heart
- Increased cardiac output (> 5 folds)
- Increase in mean arterial pressure

3. Emotional excitement. During emotional excitement states, such as fright, auditory and olfactory stimuli, the CBF is increased due to increased sympathetic discharge.

4. Hypotension. There occurs reflex increase in noradrenergic discharge during hypotension which produces coronary vasodilation to increase CBF. This effect is observed secondary to the metabolic changes in the myocardium at a time when there occurs vasoconstriction of splanchnic, renal and cutaneous vessels.

5. Hormones affecting CBF are:

- *Thyroid hormones* increase CBF because of increase in metabolism.
- *Adrenaline and noradrenaline* cause increase in CBF as already explained indirectly.
- *Acetylcholine* may increase CBF by its action on heart similar to parasympathetic stimulation.
- *Pitressin* is known to decrease CBF by increasing coronary resistance.
- *Nicotine* is reported to increase CBF through the liberation of norepinephrine.

6. Heart rate. When heart rate is increased, stroke volume decreases, therefore, phasic CBF and O_2 consumption per beat also decreases.

7. Effect of ions. Potassium ions (K^+) in low concentration cause dilatation of coronary vessels increasing CBF, whereas high K^+ ion concentration causes constriction of coronary vessels decreasing CBF.

8. Metabolic factors. Increased metabolism of the heart increases O_2 consumption leading to relative hypoxia. Hypoxia causes vasodilation due to direct effect and also due to release of adenosine leading to increased CBF.

9. Temperature. Hyperthermia increases metabolism and so causes increase in the CBF, while hypothermia decreases metabolic rate and thus decreases CBF as well.

CORONARY ARTERY DISEASE

Coronary artery disease (CAD) also known as ischaemic heart disease results due to the insufficient coronary blood flow.

It is a condition associated with development of atherosclerosis in the coronary arteries, which supply the heart muscles (myocardium). With atherosclerosis, the arterial wall is hardened and its lumen becomes narrow due to plaque formation which may consist of calcium deposits, fatty deposits, smooth muscle proliferation and abnormal inflammatory cells.

Risk factors for CAD include:

- *Age and sex*. Men over 60 and women over 65 are more prone.
- *Family history* is a predisposing factor.
- *Diseases* like diabetes, hypercholesterolaemia and hypertension are proven risk factors.
- *Smoking* is a big risk factor.
- *Obesity*
- *Diet* rich in saturated fats and low in antioxidants.
- *Life style*. Sedentary worker with lack of exercise.

Pathophysiology and manifestations of CAD

Coronary artery atherosclerosis per se or with superadded arterial spasm or thrombus leads to limitation of blood flow to the heart muscle causing ischaemia (cell starvation due to a lack of oxygen) of the myocardial cells. Myocardial ischaemia clinically may manifest as:

1. Stable angina pectoris or
2. Acute coronary syndrome: Unstable angina, Myocardial infarction

Angina pectoris

Definition. Angina pectoris refers to a transient form of myocardial ischaemia, especially occurring during increased oxygen demand (e.g. during exercise) in patients with coronary artery disease having about 60–70% narrowing of coronary arteries. Superadded thrombus formation causing incomplete coronary occlusion results in an unstable angina.

Characteristic features. Typically, the angina is described as a feeling of uncomfortable pressure, fullness, squeezing or pain in the substernal region, which may be localized or may be referred to the inner border of left arm, neck or jaw. Pain occurs due to accumulation of anoxic myocardial metabolites and *factor P* which stimulates pain nerve endings.

Types. Angina is of two types; stable and unstable.

Stable angina, also known as effort angina refers to the occurrence of above described features of angina precipitated by some activity (walking, running, exercise etc.) with minimal or non-existent symptoms at rest.

Unstable angina occurs at rest and usually lasts for more than 10 min. Attacks are frequent.

Myocardial infarction

Myocardial infarction (MI) or acute myocardial infarction (AMI), commonly known as a 'heart attack' refers to a degree of myocardial ischaemia (due to interruption of blood supply) that causes irreversible changes (necrosis i.e. cell death or infarction) in the myocardium. Commonly MI occurs when partially occluded coronary artery is constricted further by vasospasm or plaque (most common cause), which triggers formation of thrombus and occludes coronary artery.

Signs and symptoms

- Sudden severe chest pain is a classical symptom of MI. Pain lasts for more than 30 min and typically may radiate to left arm and left side of neck. Pain occurs due to the anoxic metabolites and necrotic tissue products.
- Associated symptoms with pain, often complained by patients are shortness of breath, nausea, vomiting, palpitation, sweating and anxiety (often described as a sense of impending doom).

Note. Approximately 25% of all myocardial infarction are 'silent' i.e. without chest pain or other symptoms. Silent MI usually occurs in diabetics with associated autonomic neuropathy in elderly and also in patients with heart transplants.

Diagnosis of MI is made by triad of:

- Typical signs and symptoms associated with
- ECG changes seen on serial tracings and
- Changes in serum levels of certain enzymes and proteins (cardiac biomarkers).

ECG changes in myocardial infarction are very important to diagnose, localize the area of infarction and to know the duration of infarction. Typical ECG changes (hallmark) seen in MI include:

- Elevation of ST segment in the leads overlying the infarct area and
- Depression of ST segments in the reciprocal leads.

For details see page 202.

Measurement of serum enzymes and protein related to MI (cardiac biomarkers). Certain enzymes and proteins (called as Cardiobiomarkers) leak into the circulation from the damaged myocardial cells. These include:

- *Troponin-T and Troponin-I* are cardiac specific proteins and so most sensitive and specific for MI, released 2–4 h after MI. Peak levels are seen after 12 h and persist up to 7 days.
- *Creatine Kinase (CK-MB)* levels increase within 4–6 h and lasts for 2–3 days. It is relatively specific when skeletal muscle damage is not present.
- *Lipoprotein-(a) [Lp(a)].* There is a relation between atherosclerosis and circulating levels of Lp(a). Lp(a) interferes with the fibrinolysis by decreasing plasmin generation.
- *Lactate dehydrogenase (LDH),* levels peak within 72 h. LDH is not specific as troponin.
- *Homocysteine.* This substance damages endothelial cells of coronary vessels. There is a strong positive correlation of circulating levels of Homocysteine and MI.

  Note. Homocysteine is converted into non-toxic substance (methionine) by vitamin B_{12} and folic acid. Therefore supplements of folate and vitamin B_{12} lower the incidence of MI.

- Highly sensitive *C-reactive protein* and other inflammatory markers are correlated with the presence of inflammatory cells in the atherosclerotic lesion. Therefore, estimation of plasma C-reactive protein is also helpful in diagnosis.
- *Glycogen phosphorylase isoenzyme BB (GPBB)* is one of the new cardiobiomarkers. A rapid rise in blood levels can be seen in MI and unstable angina within 3 h after process of ischaemia and peak levels are seen within 7 h. It has a high sensitivity and specificity, if estimated early after chest pain.
- *Serum transaminases (AST/ALT).* These are non-specific cardiac biomarkers as they exist in other tissues, namely liver, skin, RBC, etc. However, rise and fall with characteristic symptoms and ECG changes, these can be used as a marker.

CEREBRAL CIRCULATION

CEREBRAL BLOOD VESSELS

Arteries of the brain

- The arteries which supply blood to the brain are derived from two internal carotid arteries and the basilar artery (formed by union of the right and left vertebral arteries). Branches of the internal carotid arteries and of basilar artery anastomose on the inferior surface of the brain to form the circulus arteriosus (circle of Willis).
- The *circle of Willis* (Fig. 4.6-5) is thus basically a free anastomoses between the two internal carotid arteries and the two vertebral arteries which equalize pressure on the arteries of the two sides. In this way, the circulus arteriosus allows blood that enters by either internal carotid or vertebral artery to be distributed to any part of both cerebral hemispheres.
- *Six large arteries* taking part in the formation of circle of Willis supply by their central and cortical branches to the brain substance.
- *Anatomical peculiarities of cerebral capillaries* and their role in *blood–brain barrier* are discussed at page 775.

Venous drainage of the brain

The cerebral hemisphere has two sets of veins: the superficial and deep. The veins draining the brain open into the various dural venous sinuses. Ultimately, the blood from all these sinuses reaches the sigmoid sinuses, which become continuous with the two internal jugular veins. The venous drainage of individual part of the brain is as follows:

Veins of the cerebral hemisphere

- The veins of the cerebral hemisphere consist of two sets: superficial and deep veins,
- *Superficial veins* drain into neighbouring venous sinuses and
- *Deep veins* of the cerebral hemisphere are two internal cerebral veins and two basal veins. The internal cerebral veins join to *great cerebral veins.* Basal veins end in the great cerebral vein which in turn ends in straight sinus.

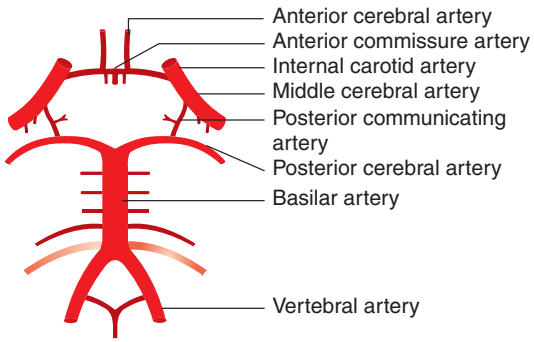


Fig. 4.6-5 The circle of Willis.

Note. The veins have no valves but are opened by structures (dura) around their orifices.

CEREBRAL BLOOD FLOW: CHARACTERISTIC FEATURES

Normal blood flow

- Brain relies on a continuous blood flow for adequate function. It is most susceptible to ischaemia. Interruption of blood flow only for 5–10 s causes a loss of consciousness and circulatory arrest for only 3–4 min results in irreversible brain damage. The vegetative structures in brainstem are more resistant to hypoxia than cerebral cortex.
- Brain has a very rich blood supply. Normal blood flow to brain (which forms less than 2% of the body weight) represents approximately 15% of the resting cardiac output (75 mL/min), or about 50–55 mL blood/100 g tissue/min.
- When cerebral blood flow falls below 18 mL/100 g tissue/min *(critical flow level),* there occurs unconsciousness.

Normal O_2 consumption

- Total O_2 consumption of the brain is approximately 50 mL/min (3.3 mL/100 g tissue/min), i.e. 20% of the whole body at rest.
- O_2 consumption of grey matter (GM) is much more than the white matter. The greater O_2 demand of GM is essential for its functioning and is made possible because of high density of capillary network (4000 capillaries/mm^2 of GM).

MEASUREMENT OF CEREBRAL BLOOD FLOW

1. Kety method

Principle. It is based on the Fick's principle (see page 216).

Procedure. In this method, subject is made to breathe a mixture of 15% nitrous oxide and air for 10 min. During this period, serial samples are taken every minute from the internal jugular vein (representing venous blood from brain) and a peripheral artery. The CBF then is calculated as:

$$\text{CBF} = \frac{\text{N}_2\text{O taken up by brain tissue per min}}{\text{A} - \text{V diff of N}_2\text{O concentration}}$$

Disadvantage of this technique is that it gives an average value but no information about regional differences in the blood flow.

2. By using radioactive substances, i.e. single photon emission computed tomography (SPECT)

Principle. In this technique, blood flow to a region of brain is measured from the clearance curve of an inert radioactive tracer.

Procedure. Commonly used radioactive substance in this method is radioactive xenon (^{133}Xe or ^{123}Xe). A known amount of ^{133}Xe gas dissolved in saline is injected within 1–2 s in the internal carotid artery with the help of a thin catheter. The arrival and clearance of the tracers is detected for 10 min by multiple collimated scintillation detector built into a helmet that fits over the cranium. Each detector collimated to scan about 1 cm^2 of brain surface. The output from the detectors is processed in computer and displayed on coloured television screen. The colour is proportional to amount of blood flow. Resolution can be improved by computerized tomographic reconstruction. The mean CBF in mL/g/min is calculated from following equation:

$$\text{CBF} = \frac{\lambda b\ (H_{max} - H_{10})}{A_{10}}$$

λb = brain blood partition coefficient,
H_{max} = maximal height of clearance,
H_{10} = Clearance height at 10 min,
A_{10} = Area under clearance curve.

Advantage of this technique is that blood flow to different regions of the cerebral cortex can be measured in conscious humans.

3. Positron emission tomography (PET)

It can also be employed to measure regional cerebral blood flow.

4. Magnetic resonance imaging (MRI)

In this technique, regional concentration of individual metabolites is measured and changes in local O_2 utilization are mapped from which regional cerebral blood flow is studied.

REGULATION OF CEREBRAL BLOOD FLOW

The perfusion pressure which determines cerebral blood flow is the difference between the mean arterial pressure at the head level and the internal jugular pressure (cerebral venous pressure). Therefore, the factors which affect the cerebral blood flow are:

- Arterial blood pressure,
- Intracranial pressure,
- Resistance, i.e. viscosity of the blood and
- Diameter of the cerebral blood vessels.

The cerebral blood flow is regulated by following mechanisms:

1. Metabolic regulation

The important metabolic factors which play important role are:

(i) Carbon dioxide. Physiologically, pCO_2 is the most potent vasodilator of cerebral blood vessels (Fig. 4.6-6). A rise

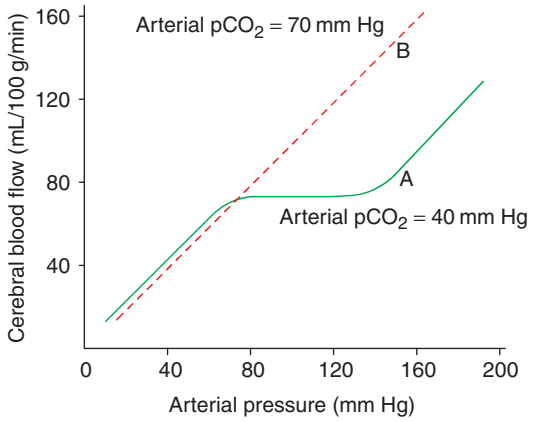


Fig. 4.6-6 Effect of pCO_2 on cerebral blood flow and autoregulation of cerebral blood flow.

in pCO_2 is associated with a rise in H^+ concentration. Carbon dioxide easily diffuses through the blood–brain barrier and reaches the CSF. In the CSF, CO_2 combines with water to form carbonic acid, which partially dissociates to form H^+ ions. The H^+ ions induce cerebral vasodilation in proportion to their concentration. The H^+ ions, however, do not cross the blood–brain barrier. Therefore, any substance that increase the acidity of brain and therefore, the H^+ ion concentration will increase cerebral blood flow.

(ii) pO_2. Slight fall in pO_2 causes vasodilation and produces increase in cerebral blood flow due to rapid formation of *adenosine,* which is potent dilator of pial arterioles.

(iii) K^+ ions. An increase in K^+ concentration in the CSF following hypoxia cause rapid increase in cerebral blood flow.

2. Autoregulation of cerebral blood flow

Like other vital organs of the body, cerebral circulation also shows phenomenon of autoregulation. Due to autoregulation, cerebral blood flow remains nearly constant between 60 and 140 mm Hg blood pressure (Fig. 4.6-6A) when blood pressure falls below 60 mm Hg, the cerebral blood flow becomes extremely compromised and syncope may result. When blood pressure rises above 140 mm Hg, there may occur disruption of blood–brain barrier due to stretching and cerebral oedema or cerebral haemorrhage may result.

3. Role of intracranial pressure in regulation of cerebral blood flow

The intracranial pressure level regulates cerebral blood flow by the following two mechanisms:

(i) Monro-Kellie doctrine. According to this doctrine the brain, CSF and blood in the cerebral vessels are three elements enclosed in a rigid cranial cavity and when any of them increases, it is at the expense of other two. This relationship helps to maintain the cerebral blood flow when changes in the arterial blood pressure occur at the level of head.

(ii) Cushing reflex. When intracranial pressure is increased and becomes equal to the arterial pressure, it compresses the arteries in the brain and blood supply to vasomotor area is compromised. The hypoxia and hypercapnia produced locally increases the discharge from VMC. The resultant rise in a systemic pressure tends to restore the cerebral blood flow (see page 260).

4. Nervous regulation of cerebral blood flow

The cerebral blood vessels are innervated by the noradrenergic vasoconstrictor fibres and cholinergic vasodilator fibres. However, under normal conditions vasomotor nerves do not regulate the cerebral blood flow.

- *In severe hypertension,* the noradrenergic sympathatic nerves cause vasoconstriction reducing cerebral blood flow. This prevents cerebral vascular haemorrhage and stroke and also protects the integrity of blood–brain barrier which gets disrupted at high blood pressure.
- *In severe hypotension,* the cholinergic sympathetic vasodilator nerves play role in maintaining the cerebral blood flow.

CUTANEOUS CIRCULATION

CUTANEOUS BLOOD VESSELS (FIG. 4.6-7)

Cutaneous arterioles form a dense network just under the dermis layer of the skin.

- *Meta-arterioles* which arise from the arterioles are relatively high-resistance conduits present between the arterioles and capillaries.
- *Cutaneous capillaries.* The meta-arterioles subdivide into capillary loops which provide a large surface area for heat exchange.
- *Venules* form an extensive subpapillary venous plexus which holds large quantity of blood and lie parallel to the surface of skin and play an important role in maintaining the body temperature.
- *Arteriovenous anastomoses* are located in the distal parts of the extremities (hands and feet), the nose, lips and ear lobules. These vessels serve as shunts and allow blood to bypass the superficial capillary loops and play a major role during control of body temperature.

CUTANEOUS BLOOD FLOW: CHARACTERISTIC FEATURES

Main function of cutaneous circulation is to aid in the regulation of body temperature.

Resting cutaneous blood flow, i.e. the flow when a person is at thermal equilibrium with the environment (at approximately

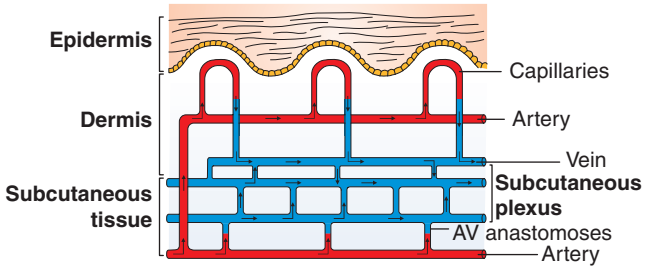


Fig. 4.6-7 Arrangement of blood vessels in subcutaneous region.

27°C atmospheric temperature), is about 10–15 mL/min/100 g of skin tissue.

During exposure to cold, when sweating is minimal, the cutaneous blood flow falls to about 1/10th of resting blood flow, i.e. about 1 mL/min/100 g tissue.

During exposure to heat, when sweating is maximum, the cutaneous blood flow may increase ten times of resting blood flow, i.e. about 150 mL/min/100 g tissue.

Regional variation in cutaneous blood flow exists due to presence of A–V anastomoses in abundance in certain area, such as hands, feet, nose and ear lobules. During heat stress, the blood flow to the area with rich A–V anastomoses increases much more (about 75 mL/100 g/min) as compared to the rest of the skin (about 25 mL/min/100 g tissue).

Cutaneous blood flow and skin colour. The colour of skin is basically determined by the pigment present; however, the amount of blood and degree of oxygenation also affect the skin colour tinge which may be reddish, bluish or some shade in between.

REGULATION OF CUTANEOUS BLOOD FLOW

The cutaneous blood flow is predominantly regulated by the nervous control instead of metabolic control.

Nerve supply of cutaneous vessels

Sympathetic vasoconstrictor nerves supplying the cutaneous vessels exhibit a sympathetic constrictor discharge under resting condition. The sympathetic tonic discharge is more marked on A–V anastomoses vessels than the other vessels.

Parasympathetic vasodilator nerves do not supply the cutaneous blood vessels. Vasodilation of cutaneous vessels results due to:

- Reduction of sympathetic vasoconstrictor effect,
- Local production of bradykinin (a potent vasodilator polypeptide) in sweat glands and
- Production of other local vasodilator substances.

Neural control mechanisms

The cutaneous blood flow is regulated by the following neural control mechanisms.

1. Hypothalamic control mechanism

The reflex increase or decrease in the sympathetic discharge to cutaneous vessels during thermoregulation is mediated through the temperature regulation centres of the hypothalamus as:

Under resting conditions, i.e. when the person is at thermal equilibrium with the environment (at about 27°C atmospheric temperature) the sympathetic vasoconstrictor fibres have a mild tonic discharge. The tonic sympathetic discharge normally keeps the A–V anastomoses closed.

During exposure to heat stress, the tonic sympathetic discharge is reflexly abolished by a hypothalamic mechanism. Thus the blood flow to skin is increased by following responses in a chronological sequence:

- First of all, A–V anastomoses of hands, feet, and ear lobules dilate due to reduction in the sympathetic tonic discharge.
- Secondly, rest of the cutaneous vessels dilate due to progressive withdrawal of sympathetic vasoconstrictor activity.
- Thirdly, sweat glands get activated due to the cholinergic sympathetic discharge. The *bradykinin* produced by the secretory activity of the sweat glands acts locally as a powerful vasodilator and increases blood flow to skin.

All the above mechanisms combinedly may increase the cutaneous blood flow to as high as 150 mL/min/100 g tissue. The increased blood flow carries heat to the surface of the body, where it is dissipated by *radiation, evaporation* and *conduction* to the environment.

During exposure to cold stress, via hypothalamic mechanism, there occurs widespread cutaneous vasoconstriction due to increased sympathetic discharge. Consequently, cutaneous blood flow is markedly decreased to as low as 1 mL/min/100 g. In this way, *heat conservation* is accomplished by markedly diminishing the rate of blood flow to skin.

2. Baroreceptor-mediated reflex

Cutaneous blood vessels participate in the baroreceptor-mediated reflexes during conditions of circulatory stress, such as exercise and haemorrhage.

3. Cortical control mechanism

The *emotions* affect the cutaneous circulation through the corticohypothalamic pathway. The effects of emotions on cutaneous circulation manifest in following forms:

Blanching of skin during situations of fear (pale with fear) occurs due to vasoconstriction mediated through cortical mechanism.

Phenomenon of blushing, i.e. emotional embarrassment occurs due to vasodilation of vessels.

CUTANEOUS VASCULAR RESPONSES

Certain peculiar cutaneous vascular responses are:

- White reaction,
- Triple response,
- Dermatographia,
- Axon reflex,
- Reactive hyperaemia,
- Cold vasodilatation and
- Cold vasoconstriction.

White reaction

White reaction refers to an appearance of a pale stroke line when a pointed object is drawn lightly over the skin. This occurs due to the fact that the mechanical stimulus initiates contraction of the precapillary sphincter and blood drains out of the capillaries and small veins. This response appears in about 15 s.

Triple response

Triple response is a three-part response, consisting of the red reaction, wheal and flare, when the skin is stroked more firmly with a pointed instrument, instead of white reaction, there occurs triple response.

The red reaction refers to the red line which appears at the site of injury in about 10 s. It occurs due to dilatation of the precapillary sphincters in the injured area. The dilatation of the precapillary sphincters is produced by histamine and/or some polypeptides such as bradykinin released from the damaged skin.

The flare refers to the diffusely spreading and irregularly outlined redness of the skin surrounding the red line.

- It occurs after a few minutes of the appearance of red line.
- It occurs due to the dilatation of the arteriole and precapillary sphincters.
- The dilatation of arteriole is mediated by nerves, since it is abolished by the local anaesthetic agents.
- The flare is mediated by the *axon reflex* in the cutaneous fibres, which does not involve the CNS like a typical reflex.

Wheal refers to the swelling or localized oedema that develops within the area of flare.

- It occurs due to increased capillary permeability with consequent extravasation of fluid.
- The increase in capillary permeability responsible for the wheal formation is produced by histamine or histamine-like substance (released from local mast cells) and by substance P (the transmitter released at the central termination of the sensory fibre neurons).

Dermatographia

Dermatographia refers to a striking triple response that occurs as an unusual reaction in some individuals. Thus, in the prone individuals anything drawn on the skin even with a blunt point becomes conspicuous within a few minutes. Possibly, it is due to excessive release of the histamine from the involved skin area.

Reactive hyperaemia

Reactive hyperaemia is a phenomenon by which the vessels control blood flow to the organ after a period of *ischaemia* following occlusion of the artery to an organ or tissue. This response of the blood vessels occurs in many organs but is visible in the skin as fiery red skin.

Cold vasodilatation

As discussed above, normally during exposure to cold stress, there occurs widespread cutaneous vasoconstriction via hypothalamic mechanism. However, prolonged and severe vasoconstriction may lead to tissue damage known as *frost-bite.* This usually occurs when skin temperature falls below 10°C. The tissue injury so produced is painful and associated with release of histamine and/or some other polypeptide which excites the sensory terminals and produce *vasodilatation* due to *axon reflex* operating particularly on A–V anastomoses. Cold vasodilatation is the cause of ruddy cheeks seen in fair complexioned individuals on a cold day.

Cold vasoconstriction

Exposure to cold causes vasoconstriction via hypothalamic mechanism. Prolonged cold-induced vasoconstriction especially in damp conditions results in cutaneous ischaemia producing lesions such as *trench foot.*

SKELETAL MUSCLE CIRCULATION

SKELETAL MUSCLE BLOOD FLOW: CHARACTERISTIC FEATURES

At rest, the blood flow to the skeletal muscle is about 2–4 mL/min/100 g of the muscle tissue.

- Since the whole body skeletal muscles weight is approximately 30 kg in adults, so the total blood flow to the body muscle mass is about 750–800 mL/min.
- At rest only 20–25% of muscle capillaries have flowing blood.

During exercise. During strenuous exercise, muscle blood flow can increase up to 20 times, i.e. about 50–80 mL/min/100 g muscle tissue or over 20 L/min to the whole body skeletal mass. This is called exercise hyperaemia.

- The tremendous increase in the muscle blood flow mainly occurs due to local metabolite-induced vasodilatation. During exercise there occurs dilatation of the arterioles and precapillary sphincters, and all the dormant capillaries open up, greatly increasing the surface area and the rate of blood flow to the skeletal muscles.
- During exercise the blood flow to the muscle is intermittent because during each contraction, muscle fibres squeeze the blood vessels passing through them and thus the blood flow decreases or even stops (Fig. 4.6-8). During relaxation period, muscle blood flow increases and myoglobin acts as an O_2 acceptor and it yields its O_2 to the myofibrils during the subsequent muscle contraction.
- Sustained and severe contractions lasting more than 10 s lead to cessation of blood flow, myoglobin supply of O_2 is exhausted and anaerobic metabolites accumulate causing fatigue and ischaemic pain.
- Following heavy phasic exercise, the blood flow does not subside immediately, but falls exponentially from its high level during the exercise to resting values. This is due to the *oxygen debt of exercise.*

REGULATION OF MUSCLE BLOOD FLOW

The blood flow to the skeletal muscles is regulated by an autoregulation mechanism, metabolic control mechanism and nervous control mechanism.

1. Autoregulation mechanism

Mechanism of autoregulation is well developed in the skeletal muscles like that of kidney, heart and brain. The precapillary resistance vessels in the skeletal muscles have a *high basal myogenic tone.* A rise of the transmural pressure excites a stretch-induced contraction of the sphincter smooth muscles which by raising the precapillary vessel tone protects the capillaries from an undue rise of pressure *(myogenic theory of autoregulation)* also (see page 236).

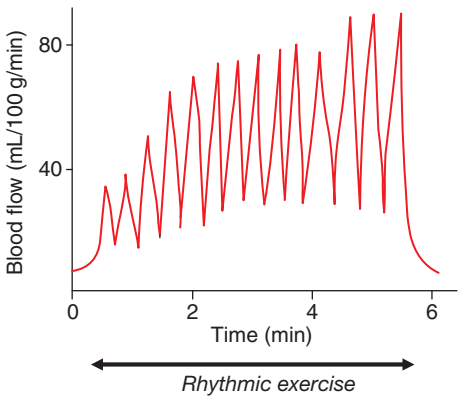


Fig. 4.6-8 Blood flow through skeletal muscles.

2. Metabolic control mechanism

Local metabolic control mechanism is chiefly responsible for the tremendous increase in the skeletal muscle blood flow during exercise (exercise hyperaemia). The decreased tissue pO_2 leads on to vasodilatation. In addition, the levels of adenosine, potassium ions, hydrogen ions, lactic acid and carbon dioxide rise in exercising skeletal muscle. Further, rise of *tissue temperature* due to muscular activity may also contribute to the dilatation of arterioles and precapillary sphincter. As a result there is 10 to 100 fold increase in the number of open capillaries in the skeletal muscles.

3. Nervous control mechanisms

Vessels of the skeletal muscles are supplied by both sympathetic vasoconstrictor and sympathetic vasodilator fibres.

Sympathetic vasoconstrictor control

Under resting conditions, the noradrenergic sympathetic nerve fibres discharge at a rate of 1 impulse/s in recumbent position and 2–3 impulses/s in the upright position. This sympathetic discharge contributes relatively small part to the high basal tone of the resistance vessels of the muscles. Most of the basal tone of these vessels is myogenic because of this relatively low sympathetic contribution to the total basal tone.

During muscular exercise. Because of the low sympathetic contribution to total basal tone, it is obvious that *exercise hyperaemia* is independent of the sympathetic discharge to muscle vessels and is due to metabolic factors as described above.

Further, skeletal blood vessels contain both α adrenergic and β adrenergic receptors. Alpha (α) receptors are located in close proximity of the terminals of vasoconstrictor sympathetic nerve fibres and β receptors are independent of any innervation. During strenuous exercise, norepinephrine and epinephrine are released into the systemic circulation from adrenal medulla, which tend to produce vasoconstriction and vasodilatation by acting on α and β receptors, respectively. These two factors tend to counter each other and hence may not contribute significantly to the increased skeletal blood flow during exercise.

During circulatory shock and other type of circulatory stress, the sympathetic vasoconstrictor mechanism assumes a great physiological importance. Sympathetic vasoconstriction reduces muscle blood flow profoundly. Thus help in diverting substantial amount of blood from the muscles towards the heart and other vital organs.

Sympathetic vasodilator fibres

- Skeletal blood vessels are also supplied by the sympathetic vasodilator fibres which have acetylcholine as neurotransmitter.
- These fibres are activated by *corticohypothalamic reticulospinal* pathways (see page 253).
- Sympathetic vasodilator fibres play no significant role in increasing the muscle blood flow either during or before exercise but.
- These fibres play an important role in preventing the sudden rise in the systemic blood pressure at the beginning of exercise, i.e. these provide a *safety valve mechanism as* just before the start of exercise, there occurs a considerable increase in the sympathetic activity.

SPLANCHNIC CIRCULATION

SPLANCHNIC VESSELS

Splanchnic circulation includes the combined vascular beds of the intestines, pancreas, spleen and liver. The main vessels which constitute the splanchnic circulation are:

Arteries supplying the blood to the intestines, pancreas, spleen and liver (Fig. 4.6-9) include:

- *Coeliac trunk* is about 1 cm long and after arising from the abdominal aorta it divides into three main branches the *left gastric artery hepatic artery and splenic artery,*
- *Superior mesenteric artery* and
- *Inferior mesenteric artery.*

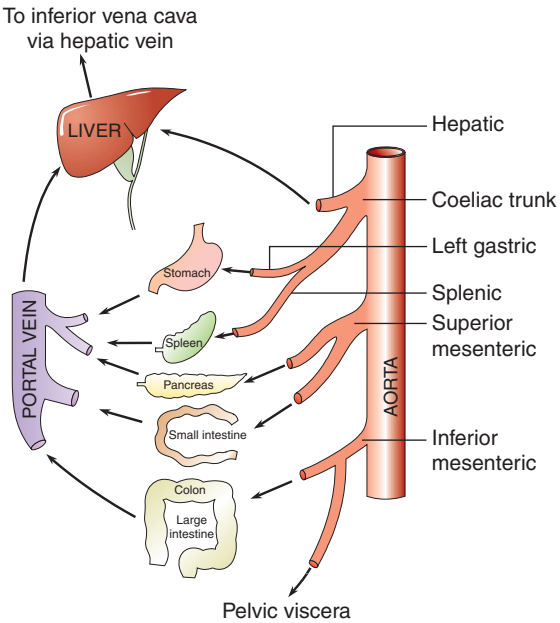


Fig. 4.6-9 Splanchnic circulation.

Hepatic portal system is formed by the veins draining blood from the abdominal part of the gastrointestinal tract (GIT). The veins comprising the hepatic portal system are shown in Fig. 4.6-10. All these veins end in the *portal vein.* The portal vein supplies the blood collected from GIT to the liver by its right and left branches.

Hepatic veins are terminal parts of an elaborate venous tree that permeates the liver. The hepatic veins emerging from the liver tissue end in the inferior vena cava (Fig. 4.6-11)

SPLANCHNIC CIRCULATION: CHARACTERISTIC FEATURES

- During rest the abdominal GIT, viscera and liver receive about 1500 mL blood per minute (about 30% of cardiac output) via coeliac, superior mesenteric and inferior mesenteric arteries (Fig. 4.6-10).
- If the entire GIT become simultaneously active, the splanchnic blood flow would have increased to about 4.0 L/min. However, since during digestion and absorption, the GIT is sequentially activated, the maximum circulation is about 3.0 L/min.
- The unique feature of the splanchnic circulation is that the venous blood from GIT viscera is not directly carried to the heart through systemic veins, but is carried to the liver forming hepatic portal system.

For the purpose of discussion, the splanchnic circulation is considered to consist of three parts:

- Intestinal (mesenteric) circulation,
- Splenic circulation and
- Hepatic circulation.

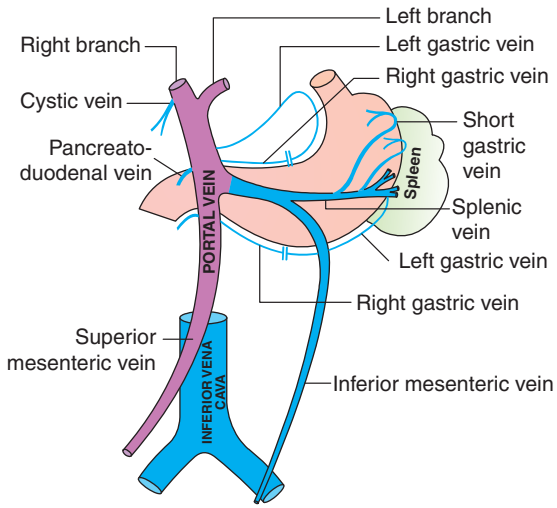


Fig. 4.6-10 Tributaries of portal vein.

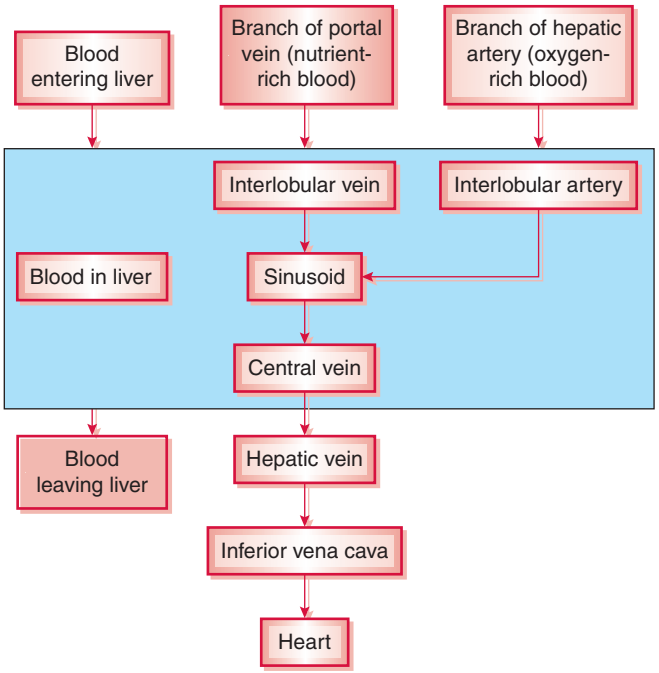


Fig. 4.6-11 Blood flow through the liver.

INTESTINAL CIRCULATION

Intestinal or mesenteric circulation is constituted by the blood supplied to the intestines and pancreas (about 100 mL/min) by a series of parallel circulations via the branches of superior and inferior mesenteric arteries.

Extensive anastomoses between the vessels constituting mesenteric circulation, but blockage of a large intestinal artery still leads to infarction of the below.

The blood flow to the intestinal mucosa is much more (about five times) than that of rest of the intestinal wall (Table 4.6-3).

During metabolic activity, the blood flow to GIT increases (Table 4.6-3) due to vagal activity (in the stomach), humoral activity, the local release of bradykinin from the mucosal glands and metabolites in the intestinal tract itself.

Counter-current system exists in the capillaries and venules in a villus, i.e. the direction of blood flow in the capillaries and venules in a villus is opposite to that in the main arteriole (Fig. 4.6-12).

This system permits diffusion of O_2 from the ascending arterial limb of villi into the descending venous limb (Fig. 4.6-12B). In this way, at low flow rates substantial amount of O_2 from the arterioles is shifted to the venules near the base of villi resulting in decrease in O_2 supply to the mucosal cells at the tips of villi. When intestinal blood flow is very low, the transfer of O_2 from arterioles to venules is exaggerated and may cause extensive necrosis of intestinal villi.

Table 4.6-3 Blood flow to intestines (mL/100 g/min)

Structure	At rest	During maximum metabolic activity
Intestinal mucosa	50–60	300–400
Rest of the intestinal wall	10	40

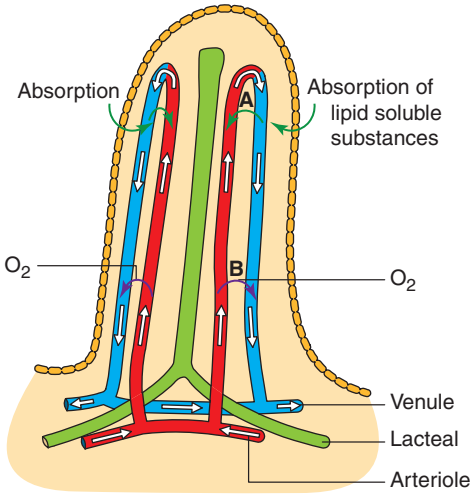


Fig. 4.6-12 Counter-current system of villus blood vessels depicting transfer of lipid soluble substances from (A) venous limb to arterial limb and that of O_2 from (B) arterial to venous limb.

SPLENIC CIRCULATION

Splenic artery which is a branch of coeliac trunk supplies about 200 mL of blood/min to the spleen during rest via its splenic branches which enter the hilum of the spleen.

Spleen serves as a reservoir of blood. In spleen, two structures are involved in the storage of blood, namely splenic venous sinuses and splenic pulp. The small arteries and arterioles open directly into the splenic venous sinuses (Fig. 4.6-13). Due to dilatation of venous sinuses, a large amount of blood is stored in spleen and the spleen distends. The capillaries of the splenic pulp are highly permeable. So, lot of blood cells pass through the capillary membrane and are stored in the pulp.

The constriction of splenic venous sinuses by the sympathetic stimulation causes release of blood into the circulation.

HEPATIC CIRCULATION

Characteristic features of hepatic circulation

Source of blood. Liver receives about 1500 mL blood/min from two sources:

Hepatic artery which is a branch of coeliac trunk supplies about 20–25% (300–400 mL) of the total blood which caters to the metabolic requirements of the liver tissue.

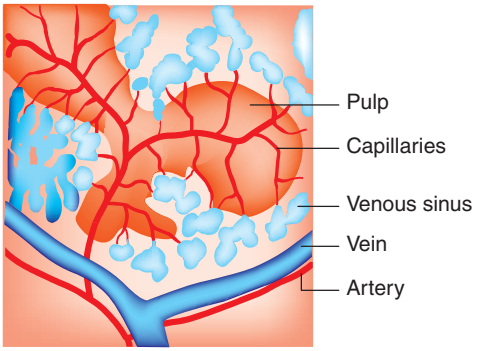


Fig. 4.6-13 Storage of blood in splenic venous sinuses and splenic pulp.

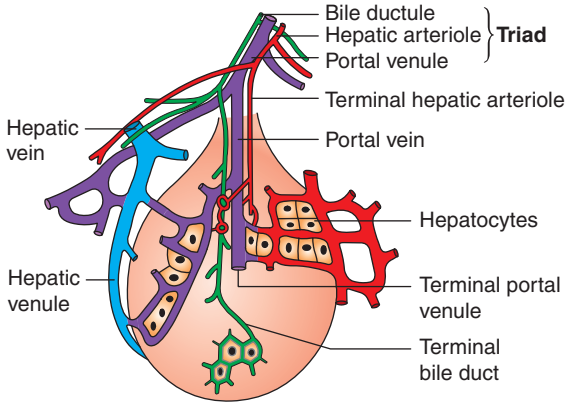


Fig. 4.6-14 Concept of acinus as a functional unit of liver.

Portal vein which collects blood from the mesenteric and splenic vascular bed supplies about 75–80% (1100–1200 mL/min) of the total blood.

The hepatic and portal blood streams meet in the sinusoids.

Functional unit of liver

The functional unit of liver is *acinus.* There are about 10,000 acini in human liver. Thus, each acinus is at the end of vascular stalk containing terminal branches of portal vein, hepatic arteries and bile ducts (Fig. 4.6-14). Blood flows from these terminal vessels into the sinusoids, which represent the capillary network of the liver. The sinusoids radiate towards the periphery of acinus, where they drain into the terminal branches of hepatic veins. Blood from these terminal hepatic venules drains into progressively larger branches of the hepatic veins, which are tributaries of the inferior vena cava.

Zones of acinus. Each acinus can be considered to have three zones: 1, 2, and 3 based on the pattern of vessels in the acinus described above. The blood supply to different zones of acinus is:

- *Zone 1* refers to the central portion of acinus immediately surrounding the terminal hepatic arteriole and terminal portal venule. This zone is *well oxygenated.* Enzymes involved in oxidative metabolism and glucogenesis predominate here.
- *Zone 2,* i.e. the intermediate zone which is present in between zone 1 and 3 is *moderately well oxygenated.* It contains a mixed complement of enzymes.
- *Zone 3* refers to the most peripheral part of the acinus. It is least well oxygenated and most susceptible to an anoxic injury. It is rich in enzymes involved in glycolysis, lipid and drug metabolism.

Regulation of hepatic circulation

1. Autoregulation. The hepatic arterial blood flow is autoregulated and the portal blood flow is not autoregulated. As described above the hepatic arterial blood flow changes reciprocally with the portal blood flow and that the adenosine is involved in this adjustment.

2. Functional hyperaemia of the intestinal tract after meals is associated with an increased portal blood flow to liver.

3. Neural regulation. The hepatic vessels are innervated by the noradrenergic sympathetic nerve fibres. The liver serves as a blood reservoir, storing about 400 mL of blood in its sinusoids. The sympathetic nerves constrict the presinusoidal resistance vessels in the portal venous system and hepatic arterial system. As described in the neural control of intestinal blood flow, the neural effects on capacitance vessels are more important. Sympathetic stimulation causes a marked reduction in the capacitance of the portal system and other splanchnic capacitance vessels and mobilizes about 1 L of blood towards the heart in less than a minute. In severe shock, hepatic blood flow gets reduced markedly and may produce patchy necrosis of the liver.

APPLIED ASPECTS

- Blood supplied by hepatic arteries to the liver does not take part in portal circulation. This blood supplies oxygen and nutrients to the liver cells.
- Obstruction of the portal vein or its tributaries causes increased blood pressure in portal venous system, a condition known as portal hypertension. This results in enlargement of spleen, oesophageal vein (varices) and formation of haemorrhoids (piles) in the rectum. Haemorrhage from oesophageal varices can be fatal.

Chapter 4.7

Cardiovascular Homeostasis in Health and Disease

CARDIOVASCULAR HOMEOSTASIS IN HEALTH

Cardiovascular homeostasis in health refers to the compensatory adjustments of the cardiovascular system to challenges faced by the circulation in everyday life. The common situations during which cardiovascular adjustments are required in day to day life include:

- Gravitational changes,
- Intrathoracic pressure changes and
- Exercise.

CARDIOVASCULAR ADJUSTMENTS DURING GRAVITATIONAL CHANGES

Gravitational changes occur under following conditions in life:

- Posture change from lying to standing and
- Prolonged quiet standing.

ADJUSTMENTS DURING POSTURE CHANGE FROM LYING TO STANDING

When posture is changed from lying (recumbent) to standing (erect), the haemodynamic changes occur as a result of the effect of gravity on the blood column which tend to reduce the cardiac output and blood pressure. However, since in humans the compensatory mechanisms are so well developed that in normal persons no effect is felt on the posture changes in day-to-day life. The sequence of events which occur during change in posture from lying to standing are:

In standing position, due to *hydrostatic (gravitational) effect of blood column,* for every centimetre below or above the heart level the pressure increases or decreases by 0.77 mm Hg, respectively. Therefore, in normal adults, the blood pressure at the level of feet (about 100 cm below heart level) in both the arteries and veins is increased approximately by 80 mm Hg.

Increased intraluminal pressure has no effect on thick-walled arteries, but the thin-walled veins distend and accommodate more blood (venous pooling).

Venous pooling (300–500 mL) results in decreased venous return and so the cardiac output and hence the blood pressure is also reduced.

A drop in the blood pressure in carotid sinus and aortic arch within seconds triggers the *baroreceptor-mediated compensatory mechanism* which causes following changes:

- *Heart rate* is increased by 5–10 beats/min.
- *Force of cardiac contraction* is increased leading to an increase in the stroke volume and cardiac output.
- *Peripheral resistance* is increased due to arteriolar constriction in the cutaneous, renal and splanchnic circulation. An increase in peripheral resistance increases the diastolic pressure.
- *Venoconstriction* in the body transfers blood from the capacitance vessels towards the heart increasing the venous return.
- *Increased secretion of renin and aldosterone,* which also help in normalization of blood pressure.

In spite of the above mentioned compensatory changes, the stroke volume and cardiac output in standing posture are about 25% less than in supine position.

However, due to a 25% increase in the total peripheral resistance, the blood pressure becomes almost normal.

The above events are summarized in Fig. 4.7-1.

Maintenance of cerebral blood flow

Cerebral blood flow in standing position is maintained by certain additional compensatory changes:

- *Decrease in jugular venous pressure to* 5–8 mm Hg due to gravity compensates for the drop in arterial pressure at head level to 20–40 mm Hg by reducing the drop in perfusion pressure (arterial pressure minus venous pressure).
- *Fall in intracranial pressure* due to fall in venous pressure reduces cerebral vascular resistance and thus facilitates cerebral blood flow.
- *Increased pCO_2 and decreased pO_2 and decreased pH* in brain tissue occurring due to decreased cerebral blood flow cause vasodilation improving the blood flow.

Because of the operation of the above mentioned *autoregulatory mechanisms,* the cerebral blood flow decreases only 20% on standing. In addition, the amount of O_2 *extracted per unit of blood increases,* and the net effect is that cerebral O_2 consumption is about the same in supine and erect positions.

The above events are summarized in Fig. 4.7-2.

CHANGES DURING PROLONGED QUIET STANDING

The prolonged quiet standing (a situation particularly met with military or police personnel, i.e. standing in attention for long periods) along with the venous pooling the fluid begins to accumulate in the interstitial spaces because of the increased hydrostatic pressure in the capillaries. The cardiac output is decreased due to decreased venous return. A stage may come when cerebral blood flow decreases to less than about 60% and symptoms of cerebral ischaemia develop. The individual may faint and fell down.

The fainting, in a sense, is also a homeostatic mechanism, because falling to horizontal position promptly restores venous return, cardiac output and cerebral blood flow to adequate levels.

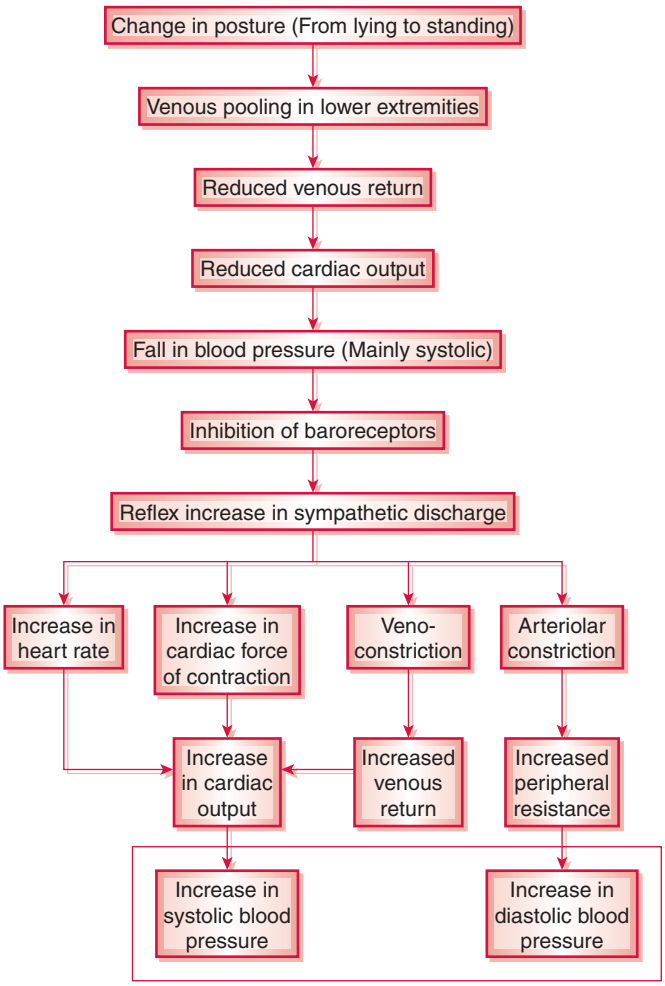


Fig. 4.7-1 Summary of events maintaining normal blood pressure during change of posture from lying to standing.

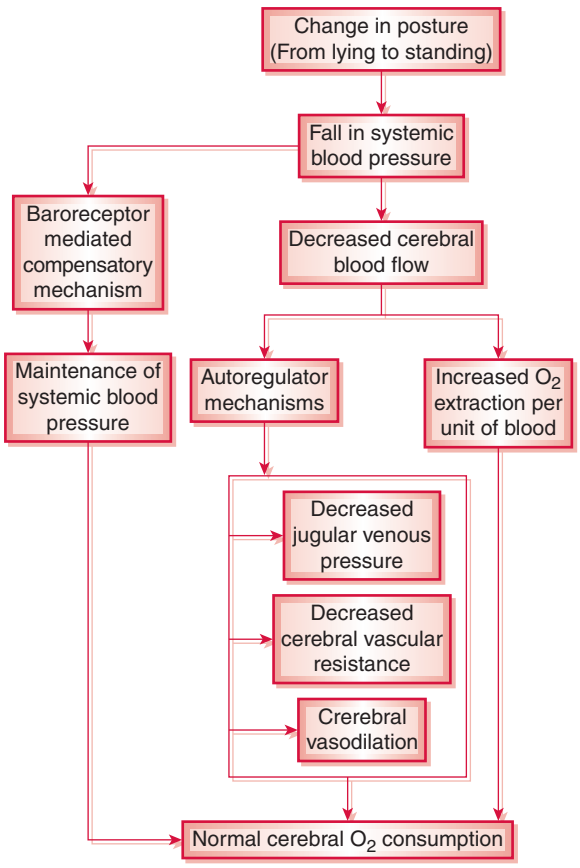


Fig. 4.7-2 Summary of events maintaining normal cerebral O_2 consumption in standing posture.

POSTURAL HYPOTENSION

Postural hypotension or orthostatic hypotension in which there occurs a sudden fall in blood pressure on changing posture from lying to erect, which causes symptoms of cerebral ischaemia. The individual experiences transient blurring of vision, dizziness or even fainting. It is diagnosed by recording blood pressure in lying and standing postures. A decrease in the systolic blood pressure by 30 mm Hg or more on standing from supine position is diagnostic.

Pathophysiology

Postural hypotension develops in individuals in whom the cardiovascular compensatory mechanism (described above) which maintain normal blood pressure and adequate cerebral blood flow are very slow to develop.

Causes of postural hypotension

1. Decreased blood volume. The effects of gravity on the circulation in humans depend in part upon the blood volume. When the blood volume is low, the compensatory mechanisms are slow to develop and the individual may suffer from postural hypotension.

2. Sympatholytic drugs. Postural hypotension is common in patients receiving sympatholytic drugs.

3. Dysfunctions of sympathetic nervous system are obviously associated with postural hypotension. Dysfunction of sympathetic nervous system may be grouped as:

(i) Surgical sympathectomy.
(ii) Autonomic neuropathy occurring in diseases, such as diabetes mellitus, syphilis and Parkinson's disease.
(iii) Primary autonomic failure.

The prolonged standing presents an additional problem due to increasing interstitial fluid volume in the lower extremities. If the person keeps on moving the operation of 'muscle pump' (see page 219) keeps the venous pressure below 30 mm Hg at the feet level and maintains adequate venous return.

Treatment

Mineralocorticoids are used to treat patients with postural hypotension. Of course, wherever possible, the causative disease should be ameliorated.

CARDIOVASCULAR EFFECTS OF GRAVITY ACCELERATION AND DECELERATION

When the body moves upwards, the force due to acceleration acting in the long axis of the body from head to foot is called *positive g*. 'g' represents the unit of gravity force, and 1 g refers to the force of gravity on earth's surface.

When the body moves downwards from height towards the earth, the force due to deceleration acting in the long axis of the body from foot to head is called *negative g*.

Positive g and negative g effects are experienced during the takeoff and landing of space rockets, during landing of airplanes, during parachute jumping and in elevators (lifts) while going up and down.

- Since the 'positive g and negative g' effects are experienced when there is acceleration or deceleration along the long axis of the body, the astronauts avoid these effects by positioning themselves perpendicular to the direction of g i.e. in a *chest-to-back direction.*

Effect of positive g

The effects of 'positive g' on the cardiovascular system are due to throwing down of blood in lower part of the body and similar to those occurring from change of posture from lying to standing, but they are multiplied depending upon the speed of acceleration.

At acceleration less than 5 g, the compensatory mechanisms (described in effect of posture change) are able to maintain vital cardiovascular status. *Cardiac output* is maintained for a time because blood is drawn from the pulmonary venous reservoir and because the force of cardiac contraction is increased *cerebral circulation* is protected due to associated fall in the jugular venous pressure and intracranial pressure.

At acceleration more than 5 g with body in long axis, the pressure in the veins in the lower limbs rises to over 450 mm Hg. The consequent passive dilatation of veins of the lower limbs retains so much blood that the venous return and therefore, cardiac output is markedly reduced. Under such a situation vision fails *(blackout)* in about 5 s and unconsciousness almost immediately thereafter.

Antigravity suit or the g-suit is used by the astronauts to effectively cushion the effects of gravitational force. The 'g-suit' is a double-walled pressure suit containing water or compressed air. When there is 'positive g', there is tendency of venous pooling simultaneously the water in the g-suit also rushes to the lower parts. The g-suit is regulated in such a way that it compresses the abdomen and legs with a force proportionate to the positive g. This decreases venous pooling and helps to maintain venous return.

Effect of negative g

The effects of 'negative g' on the cardiovascular system are due to rushing up of the blood towards head when the body suddenly moves down. As a result of accumulation of blood in the head and neck following changes occur:

- *Cardiac output* is increased due to general increase in the venous return. Most of the cardiac output however moves towards upper parts of the body.

- *Cerebral arterial pressure* is increased markedly. In spite of the great increase in cerebral arterial pressure, the vessels in the brain do not rupture because there occurs a corresponding *increase in intracranial pressure and their walls are supported.* In other words, the cerebrospinal fluid acts like a g-suit.
- Blood vessels of head and neck show intense congestion,
- *Ecchymosis* appears around the eyes,
- *Severe throbbing headache,* pain and eventually,
- *Mental* confusion (red out).

Effect of zero gravity

The 'zero gravity' situation occurs when the astronauts in a spacecraft go out of the earth's gravitational effect, e.g. during orbital flights to other planets. The absence of gravity leads to:

- Weightlessness,
- Movements of the body become effortless,
- Absence of hydrostatic pressure on the blood column.
- A data of 14 months stay in the zero gravity zone are available, which shows following documental effects.

Effects on the cardiovascular system

- *Transient postural hypotension* has been present after return to earth from space flights and full readaptation to normal gravity has been reported to occur in 4 weeks.
- *Some atrophy of myocardium* is reported to occur because of the fact that heart did not have to function for increases in cardiac output required in everyday life as on earth. *More severe disuse atrophy of myocardium* is speculated in prolonged period of weightlessness during future trips to planets.

Other effects of zero gravity

- *Flaccidity and atrophy of skeletal muscles* to some extent occurs since due to zero gravity; the muscular effect is much reduced when objects to be moved are weightless and the normal proprioceptive input is decreased. A programme of regular exercises against resistance, e.g. pushing against a wall of spacecraft or stretching a heavy rubber band may decrease the muscle atrophy.
- *Space motion sickness,* the nausea, vomiting and vertigo that occur in astronauts, develops when they are first exposed to 'zero gravity' and often wears off after a few days of space flight. It can occur with re-entering in the gravity. It occurs due to vestibular apparatus dysfunctioning.
- *Changes in the blood* noted are:
  - Loss of plasma volume, probably because of head ward shift of body fluids, with subsequent diuresis,
  - Loss of red cell mass and
  - Alterations in the plasma lymphocytes.
- *Bone mineral* is lost steadily with increased Ca^{2+} excretion. A loss of body Ca^{2+} is equivalent to 0.4% of the total body Ca^{2+} per month initially but later tapers off during prolonged space flight. Further, a high calcium diet helps to overcome this problem.
- *Psychological problems* associated with isolation and monotony of prolonged space flight are also a matter of concern.

CARDIOVASCULAR ADJUSTMENTS DURING INTRATHORACIC PRESSURE CHANGES

Intrathoracic pressure changes are not uncommon in everyday life. Depending upon the mechanism of intrathoracic pressure changes, the activities responsible for these may be grouped into Valsalva manoeuvre and Muller's manoeuvre.

VALSALVA MANOEUVRE

Valsalva manoeuvre refers to a forced expiration against a closed glottis. The common every day activities in which Valsalva manoeuvre effect is seen on the intrathoracic pressure are: straining during defaecation, initial phase of coughing and straining during parturition.

Intrathoracic pressure changes during Valsalva manoeuvre and their effects on cardiovascular system

The changes exerted on the cardiovascular system due to a sudden and sharp rise in intrathoracic pressure occur due to the Valsalva manoeuvre effect can be described in four phases (Fig. 4.7-3).

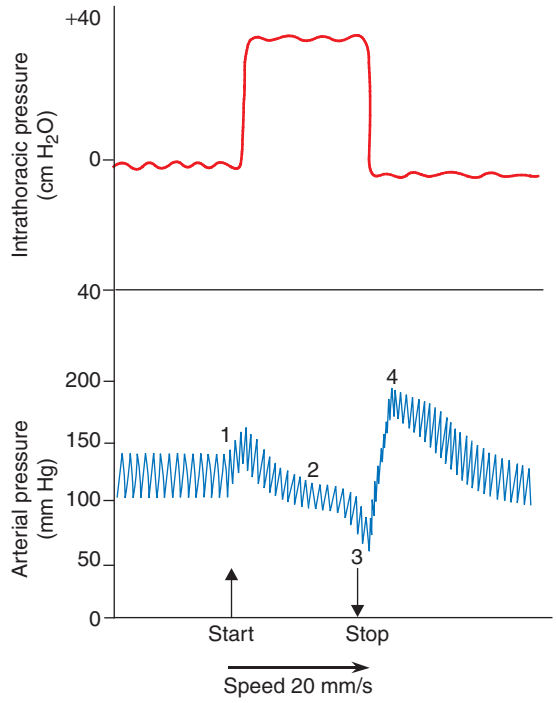


Fig. 4.7-3 Response to Valsalva manoeuvre in a normal man: A, intrathoracic pressure changes and B, arterial pressure changes (in phase 1–4).

Phase 1 is characterized by a transient rise in the arterial pressure. It coincides with the compression of the aorta due to sudden increase in the intrathoracic pressure.

Phase 2, which follows phase 1, is characterized by:

- A fall in the arterial pressure which plateau after few seconds owing to reflex vasoconstriction and
- Heart rate usually increases slightly.

Mechanism. The phase 2 changes in cardiovascular system are initiated by a decrease in venous return which occurs due to increase in the intrathoracic pressure changes.

Phase 3 is characterized by a transient fall in blood pressure which follows 1–2 s after release of the strain. It coincides with release of pressure compressing the aorta due to decreased intrathoracic pressure. In other words, events in phase 3 are just reverse of the events in phase 1.

Phase 4 is characterized by:

- *Increase in arterial blood pressure* above the resting level within 10 s. This overshoot of blood pressure is due to the lingering effect of vasoconstriction induced during phase 2. The blood pressure returns to the resting level after about 1.5 min of strain release.
- *Slowing of heart rate* occurs due to the baroreceptor-mediated vagal stimulation in response to the overshoot of blood pressure.

Clinical application of Valsalva manoeuvre

In clinical practice, Valsalva manoeuvre is employed for testing of baroreceptor reflexes and also as a test for autonomic insufficiency.

Procedure. The subject is asked to blow for about 15 s into a mouthpiece which is attached to a sphygmomanometer at a pressure of 40 mm Hg. A continuous recording of ECG is made during and after this manoeuvre.

Interpretation of results. Results of this test are interpreted in the form of *Valsalva ratio.* Valsalva ratio is the ratio of longest R-R interval (noted within 20 beats of the end of Valsalva manoeuvre) to the shortest R-R interval (noted during the Valsalva manoeuvre).

MULLER'S MANOEUVRE

Muller's manoeuvre refers to a forced inspiration against closed glottis. It is just reverse of the Valsalva manoeuvre, i.e. it reduces the intrathoracic pressure (up to –80 mm Hg). Therefore, the cardiovascular changes occurring during Muller's manoeuvre are exactly opposite to those which occur during the Valsalva manoeuvre described above.

CARDIOVASCULAR ADJUSTMENTS DURING MUSCULAR EXERCISE

Severe muscular exercise is the most stressful physiological condition that the cardiovascular homeostasis mechanisms face in everyday life. Since in addition to cardiovascular system adjustments, respiratory and other adjustments also occur in the body, so they are comprehensively discussed in the Chapter 5.8 on 'Physiology of Exercise' (see page 367).

CARDIOVASCULAR HOMEOSTASIS IN DISEASES

Cardiovascular homeostasis mechanism operates in almost all cardiorespiratory and many other multiorgan diseases. The most important condition which needs special description is circulatory shock.

CIRCULATORY SHOCK

Circulatory shock or simply called as shock is a syndrome (collection of different entities that share certain common features) characterized by *serious reduction of tissue perfusion* with a relatively or absolutely inadequate cardiac output. In other words, the shock is a condition characterized by an inadequate delivery of oxygen and nutrients to critical organs, such as heart, brain, liver, kidneys and gastrointestinal tract.

TYPES AND CAUSES OF SHOCK

Depending upon the cause of inadequacy of cardiac output (relative or absolute), the circulatory shock may be of following types:

I. Hypovolaemic shock,
II. Low-resistance or distributive or vasogenic shock,
III. Cardiogenic shock and
IV. Obstructive shock.

I. Hypovolaemic shock

Hypovolaemic shock, also known as cold shock, is caused by a low blood volume resulting in decreased cardiac output.

Causes. Depending on the causes, the hypovolaemic shock may be of following types:

Haemorrhagic shock occurs as a result of external or internal blood loss caused by the ruptured vessels.

Dehydration shock. Fluid loss when insufficient amount can dehydrate the body and reduce the circulating blood volume. Fluid loss can occur from:

- GIT in diarrhoea or vomiting,
- *Kidney* in diabetes mellitus, diabetes insipidus, or excessive use of diuretics and
- *Skin* in burns.

Traumatic shock *is* a special type of hypovolaemic shock in which there is associated neurogenic shock caused by severe pain which inhibits the vasomotor centre.

II. Low-resistance or distributive or vasogenic shock

Low-resistance or distributive or vasogenic shock occurs when neural reflexes or toxic substances cause excessive vasodilation within the vascular system. Due to vasodilation the size of capacitance vessels is increased and thus the cardiac output is decreased in spite of normal blood volume. Low-resistance shock is also called *warm shock* because skin is warm and not cold and moist as it is in hypovolaemic shock.

Causes. Depending upon the causes, the low-resistance shock is of following types:

1. Neurogenic shock occurs due to two types of nervous effects:

(i) *Marked reduction in sympathetic vasomotor tone* and
(ii) *Pronounced increased in the vagal tone* as in vasovagal syncope or emotional fainting.

2. Anaphylactic shock. Anaphylaxis refers to an acute allergic reaction. Large quantities of histamine and histamine-like substances released in the allergic reaction cause widespread and marked *vasodilation* reducing peripheral resistance. Also, there is marked *increase in capillary permeability* leading to a fluid loss and adding hypovolaemic element to the low-resistance shock.

3. Septicaemic shock. Septicaemia is a condition in which bacteria circulate and multiply in the blood and form toxic products and cause high fever and marked vasodilation due to peripheral arteriolar paralysis.

4. Endotoxic shock refers to the shock produced by the endotoxins released by gram-negative bacteria. Endotoxins produce shock due to their following effects:

- Marked vasodilation reducing peripheral resistance,
- Depressing myocardial contractility reducing cardiac output and
- Increasing capillary permeability and causing hypovolaemia.

III. Cardiogenic shock

Cardiogenic shock occurs due to the decreased pumping ability of the heart because of some cardiac abnormality.

Causes of cardiogenic shock are:
- Myocardial infarction,
- Cardiac arrhythmias,
- Congestive heart failure and
- Severe valvular dysfunctions.

IV. Obstructive shock

Obstructive shock or more precise to be the extra-cardiac obstructive shock occurs due to the impairment of ventricular filling during diastole due to some external pressure on the heart. Due to decreased ventricular filling, the stroke volume and hence the cardiac output is decreased causing circulatory shock.

Causes of obstructive effusion shock are:
- Pericardial cardiac tamponade, i.e. bleeding into the pericardium with external pressure on the heart,
- Tension pneumothorax,
- Constrictive pericarditis and
- Pulmonary embolism.

STAGES AND CLINICAL FEATURES OF SHOCK

Stages and clinical features of all types of shock are similar with minor differences. However, since the haemorrhagic shock is of more common occurrence the discussion in this section will be centred on it. Depending upon the severity, the circulatory shock can be divided into three stages:

- First stage or non-progressive shock
- Second stage or progressive shock
- Third stage or refractory shock

I. First stage of shock or non-progressive shock

Non-progressive shock also known as *compensated shock* or initial stage of shock occurs when there is a moderate reduction in cardiac output.

Hypovolaemic shock due to acute blood loss, in other words, the *haemorrhagic shock,* occurs when at least 10–15% of total blood volume is lost. Thus, about 10% of the total blood volume may be lost without any significant effect.

Compensatory mechanisms of rapid onset (short-term mechanisms) immediately set into motion following the acute loss of blood and try to maintain the blood flow to vital organs in spite of reduced cardiac output. The compensatory mechanisms are described below in detail:

A. Rapid compensatory mechanisms (Neural mechanism)

Rapid or the so-called short-term control mechanisms primarily include the following three nervous reflexes:

1. Baroreceptor reflex. When the blood pressure is decreased, there occurs decrease in the impulse discharge from the arterial baroreceptors (for detail see page 255). As a result there is a generalized increase in the sympathetic vasomotor discharge to heart, arterioles and veins. There occurs generalized vasoconstriction (sparing vessels of brain and heart). Vasoconstriction is more marked in the cutaneous, splanchnic, renal and skeletal muscle vessels.

This causes shifting of greater amount of blood in circulation. All these mechanisms maintain blood pressure at such a level that the blood flow to the vital organs like heart and brain is not affected. However, blood flow to the vital organs is provided at the cost of other body tissues such as skin, abdominal viscera, kidney and skeletal muscles.

2. Chemoreceptor reflex. An acute haemorrhage causes loss of red blood cells leading to reduced O_2 carrying capacity. The resultant anaemia and stagnant hypoxia as well as acidosis stimulate chemoreceptors which also excite vasomotor centre to cause the same effects as these caused by the baroreceptor reflex. Fall in blood pressure below 60 mm Hg usually initiates the chemoreceptor reflex.

3. Central nervous system ischaemic response. When blood pressure falls below 50 mm Hg, this response is initiated (for details see page 259). It causes more powerful sympathetic stimulation.

The rapid compensatory mechanisms discussed above account for following:

Symptoms and signs observed in patients with shock:

- *Pale, cold and moist skin* occurs due to decreased blood flow to skin and increased sweating (due to increased sympathetic discharge).
- *Cyanotic tinge of skin* may sometimes occur because of increased O_2 extraction from the blood.
- *Tachycardia* and *fall in pulse pressure* produce *thin and thready pulse,* the characteristic feature of hypovolaemic shock.
- *Increased rate and force of respiration* is due to greater sino-aortic chemoreceptors discharge.
- *Oliguria,* another important feature of the hypovolaemic shock is due to the renal arteriolar constriction.
- *Restlessness and apprehension* may occur due to the stimulation of brainstem reticular formation by circulating catecholamines, since haemorrhage is a potent stimulator of secretion of these hormones from the adrenal medulla.

B. Intermediate compensatory mechanisms

- Renin–angiotensin vasoconstrictor mechanism (see page 261),
- Reverse stress relaxation (see page 247) and
- Capillary fluid shift mechanism (see page 247).

C. Long-term compensatory mechanisms

1. Restoration of plasma volume and proteins. After a moderate haemorrhage, the plasma volume is restored to normal in 12–72 h. There is rapid entery of preformed albumin from the extravascular stores.

2. Restoration of red cell mass. In the mean time, there is excess release of erythropoietin which increases the rate of cell production in the bone marrow within 10 days. Normal cell mass is restored in 4–8 weeks.

Note. Under normal circumstances the circulatory compensatory mechanisms described above eventually cause full recovery without help of outside therapy during the stage of non-progressive shock. However, a timely outside therapy may hasten the recovery.

II. Second stage of shock or progressive shock

Second stage of the circulatory shock is progressive stage which occurs after 15–25% loss of total blood volume. In this stage, the compensatory mechanisms are not able to stop the progression of the shock. In progressive shock, the structures of circulatory system begin to deteriorate and various types of *positive feedback mechanisms* develop. Therefore, timely therapeutic interventions are essential in this stage, otherwise the vicious cycle of positive feedback mechanisms will cause progressive decrease in the cardiac output and ultimately patient will go into the stage of refractory shock.

Positive feedback cycles

Positive feedback cycles responsible for the continuous progression of shock, if not interrupted by therapeutic intervention, lead into refractory stage are (Figs 4.7-4 and 4.7-5):

1. Cardiac failure. Due to severe decrease in arterial pressure, particularly diastolic pressure, the coronary blood flow also decreases and coronary ischaemia occurs. This positive feedback causes progressive cardiac deterioration and may ultimately cause complete heart failure.

2. Vasomotor failure. Very severe fall in blood pressure, there may occur cerebral ischaemia and failure of medullary vasomotor centre (VMC). Failure of VMC results in marked vascular dilatation causing venous pooling and decreased venous return. The cardiac output and blood pressure are further decreased.

3. Peripheral circulatory failure. Due to prolonged and intense vasoconstriction, there occurs hypoxia and accumulation of metabolites in the body tissues results in increase in capillary permeability, and large quantity of fluid begin to transudate into the tissues. This decreases the blood volume and increases the shock.

4. Septicaemia and toxicaemia. Due to prolonged vasoconstriction of splanchnic vessels, there occurs hypoxia of gastrointestinal tract (GIT). The hypoxic damage causes a breakdown of normal protective mucosal barrier in the gut leading to entry of the intestinal bacteria into the portal circulation. Simultaneous deterioration of hepatic functions permits bacteria and bacterial endotoxins to reach into the systemic circulation leading to septicaemia and toxicaemia. The endotoxins cause widespread failure of arteriolar and precapillary sphincter functions and cardiac depression.

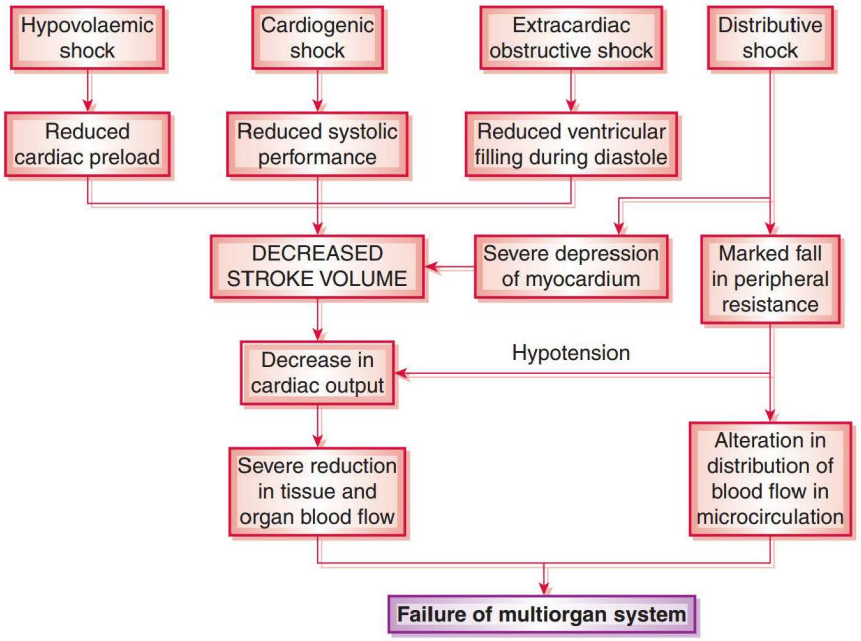


Fig. 4.7-4 Chain of events that ultimately cause failure of multiple organ system and refractive shock in different types of shocks.

There occurs extensive vasodilatation. At this stage, no amount of treatment can restore the circulatory functions to normal.

Effect on body tissues in progressive shock

In progressive shock, there occurs *widespread cellular degeneration* in the body tissues. Generalized cellular damage usually occurs first in highly metabolic tissues, such as liver, lung and heart. Liver cells are usually the first to be affected because hepatic cells have very high rate of metabolism and also is the first organ exposed to toxins from the intestine through the portal vein.

Among the different damaging cellular effects that are known to occur include:

- Great decrease in active transport of sodium and potassium through cell membrane resulting in an accumulation of sodium in the cells and loss of potassium from the cells. So, the cell begins to swell.
- Mitochondrial activity in the liver cells as well as in the cells of many other tissues of the body is decreased.
- Lysosomes begin to split in tissues throughout the body with release of hydrolases that cause widespread intracellular damage.
- Cellular metabolism of nutrients, such as glucose is eventually greatly depressed in last stages of shock. The activity of some hormones is depressed as well, including a marked suppression of insulin action.
- Poor delivery of oxygen to tissues greatly diminishes oxidative metabolism and the cells switch to anaerobic glycolysis. This leads to accumulation of lactic acid in the blood. Moreover, the sluggish blood flow through the tissue also results in the accumulation of CO_2 in tissue. The CO_2 dissolves in water to produce H^+ ions causing acidosis. Acidosis causes vasodilation which further aggravates the shock and a vicious cycle starts.

III. Refractory shock

As mentioned above, when the shock is in the progressive stage and is not treated adequately, a vicious cycle of various positive feedback mechanisms (Fig. 4.7-4) set in and patient passes into the third stage of shock, the *refractory shock* (previously known as *irreversible shock). In this stage, all therapeutic* interventions are usually ineffective and eventually the patient dies.

Causes of refractiveness (point of no return) of shock. The main factor responsible for the irreversibility of shock is the *depletion of high-energy phosphate compounds.* The high-energy phosphate reserves in the cells of the body, especially liver and heart are greatly diminished in severe degrees of shock.

Tissue damage in refractory shock. Slowly necrosis of cells of body sets in.

- In the kidney, there may occur *acute tubular necrosis* leading to acute renal failure and uraemic death.
- Deterioration of the lungs often leads to respiratory distress, the *shock lung syndrome.*

TREATMENT OF SHOCK WITH PHYSIOLOGICAL BASIS

The treatment of shock is aimed at correcting the cause and helping physiological compensatory mechanisms.

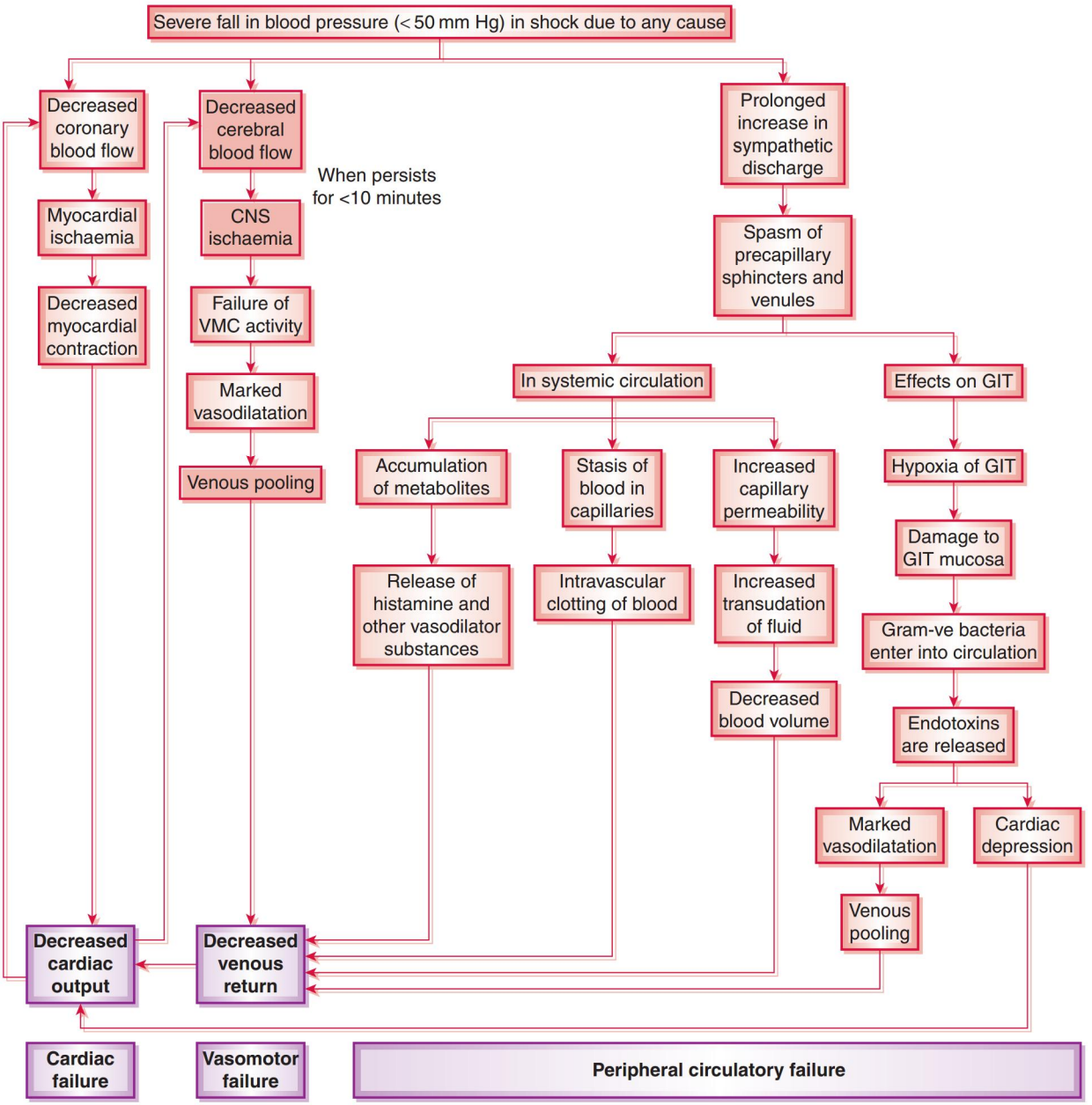


Fig. 4.7-5 Chain of events of various positive feedback cycles which cause progression of shock to refractory stage. ***Note.*** In severe haemorrhage (blood loss > 40% of blood volume) also the refractory shock develops due to these chain of events.

1. General measures for shock treatment are:

- *Room temperature.* Where the patients of shock are kept should be cold. If exposed to warmth there will be sweating which will cause further hypovolaemia and aggravate the shock.
- *Raising the foot end of the patient's bed* by 6–12″ *(Trendelenburg position)* helps in promoting the venous return and thereby increasing the cardiac output. It is especially useful in haemorrhagic and neurogenic shock when the blood pressure is too low.

2. Replacement therapy is very useful in the hypovolaemic shock.

- *In haemorrhagic shock* the best therapy is transfusion of whole *blood.* When whole blood is not available the *plasma* may be used. Plasma maintains the colloid osmotic pressure of the blood but the haematocrit decreases with this therapy. If neither whole blood nor plasma is available, a plasma substitute such as *dextran* may be used.
- *In patients with burns or interstitial obstruction,* plasma infusion is the appropriate therapy. If plasma is not available, *dextran* can be used.
- *In hypovolaemic shock* due to dehydration intravenous infusion of balanced electrolytic solution (e.g. lactated Ringer's solution is the appropriate therapy.

3. Sympathomimetic drugs are useful as:

- Sympathomimetic drugs are usually not useful in the haemorrhagic shock where the sympathetic system is already very active.
- These are especially useful in the *neurogenic shock* and *anaphylactic shock.*
- *Dopamine* should be the sympathomimetic drug of choice.
- Epinephrine or norepinephrine drug may also be used when dopamine is not available.

4. Oxygen therapy. Oxygen therapy may have some beneficial effects.

5. Glucocorticoids. Glucocorticoids are particularly useful in the anaphylactic shock.

HEART FAILURE

Heart failure is a pathophysiological state of the heart when cardiac performance is too low to maintain the cardiac output to meet the demands of the metabolizing tissues.

PATHOPHYSIOLOGY

Cardiac failure can occur in following three situations:

1. Increase in preload. According to the *Starling law of heart,* an increase in preload (end-diastolic volume) augments cardiac function, but when there is too much increase in preload then (it operates through descending limb of Starling curve, see page 218) leads to ventricular dilatation and heart failure.

2. Increase in afterload (resistance), e.g. in hypertension there is resistance to outflow of blood from the heart which causes overstretching leading to ventricular dilatation and heart failure.

3. Reduction in myocardial contractility, i.e. decrease in pumping ability of the heart decreases the cardiac output.

The interaction of these three variables leads to the development of cardiac failure.

TYPES AND CAUSES OF HEART FAILURE

Heart failures can be classified by various ways:

I. Depending on the involvement of side of heart

1. Left heart failure. Anatomically, left heart comprises left atrium, left ventricle, aortic valve and mitral valve. The left heart failure, therefore, refers to the reduction in left ventricular output leading to elevation of left ventricular volume and pressure and its transmission to left atrium and pulmonary veins. The conditions causing left heart failure are:

(i) Left ventricular outflow obstruction due to:

- Systemic hypertension
- Aortic valve stenosis
- Coarctation of aorta

(ii) Left ventricular inflow obstruction due to mitral stenosis

(iii) Reduced ventricular contractility due to:

- Cardiomyopathy particularly involving left ventricle and
- Anterior wall myocardial infarction.

2. Right heart failure. Like left heart anatomically right heart includes right atrium, right ventricle and tricuspid and pulmonary valves. Right heart failure is a condition in which there is reduction in the right ventricular output leading to rise in the right ventricular and right atrial pressure, which further causes rise in jugular venous pressure, oedema, congestive hepatomegaly and congestion of viscera except lungs.

3. Biventricular (congestive) heart failure. In this condition, there is simultaneous involvement of right and left heart due to disease of myocardium, or left ventricular failure after sometime involves the right heart also.

II. Depending on inadequate cardiac output

1. Forward heart failure results due to inadequate cardiac output and

2. Backward heart failure is the one in which decreased cardiac output results in elevation of the end-diastolic volume and thus increases the ventricular pressure. The elevation of left and right ventricular pressure results in pulmonary and systemic congestion, respectively.

III. Systolic and diastolic heart failure

1. Systolic heart failure occurs due to poor myocardial contractility (systolic dysfunction) and

2. Diastolic heart failure results due to poor ventricular filling because of defective relaxation.

Both systolic and diastolic heart failures coexist particularly in myocardial infarction.

IV. High output and low output failure

1. High output failure is a state in which cardiac output remains high, i.e. at the upper limit of normal cardiac output (3.5 L/m^2/min) even though cardiac functions are depressed. Various conditions which result in high output failure are:

- Fever,
- Thyrotoxicosis,
- Anaemia, and
- Beriberi.

2. Low output failure. In this state, cardiac output remains at its lowest limit (2.5 L/m^2/min) at rest and in stressful conditions becomes further depressed, as in the case of

heart failure secondary to the ischaemic heart disease, hypertension, valvular and pericardial diseases.

CLINICAL FEATURES

Clinical features depend on the underlying disease and type of heart failure and these are as under:

1. *Features due to low cardiac output* are:
   - Fatigue,
   - Hypotension,
   - Poor tolerance to stress and
   - Oliguria.
2. *Features due to the left heart failure* are:
   - Cardiomegaly,
   - Cardiac arrhythmia,
   - Dyspnoea and orthopnoea.
3. *Features due to the right heart failure* are:
   - Rise in jugular venous pressure,
   - Hepatomegaly,
   - Peripheral oedema, ascites and hydrothorax.
4. *Features of chronic heart failure:*
   - Raised jugular venous pressure,
   - Oedema and
   - Congestive hepatomegaly.

DIAGNOSIS

Diagnosis of heart failure is mainly based on its clinical features and following investigations which are performed to establish the nature, severity and complications which have occurred.

1. ***Electrocardiography*** (ECG) findings may reveal arrhythmia, ventricular hypertrophy and myocardial infarction.
2. ***Radiography of chest*** may show enlargement of heart, congestion of lungs and certain valvular defects.
3. ***Biochemical tests*** include estimation of blood urea and electrolytes for renal failure, hypokalaemia and hyponatraemia.

TREATMENT

The basic principles of treatment of heart failure are aimed at to:

- Remove the precipitating factors,
- Correct the underlying cause,
- Control the congestive heart failure state and
- Prevent complications.

In general following measures are employed as a treatment of cardiac failure:

A. To reduce cardiac work load

- For this, complete bed rest is advised or patient is hospitalized for 1–2 weeks,
- Small and light meals are recommended and
- Drugs (like sedatives and antianxiety) are prescribed.

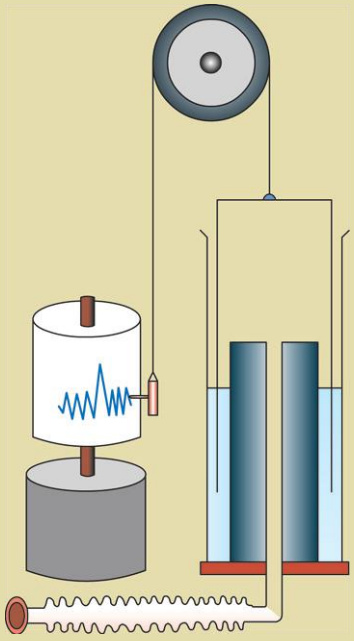
B. To improve myocardial contractility. Drugs like cardiac glycosides (digitalis) and sympathomimetic amines are prescribed.

C. To control fluid retention. Dietary salt intake is re-stricted and diuretics are given.

D. To reduce afterload, use of vasodilator drugs, especially angiotensin converting enzyme inhibitors (captopril and enalapril) is recommended.

SECTION 5

Respiratory System



The word respiration has been derived from the Latin word *respirae* which means to breath. The primary role of the respiratory system is to provide O_2 to the tissues for metabolic needs and remove the CO_2 formed by them. An adult body consumes about 250 mL of O_2 and produces about 200 mL of CO_2 per minute. Respiration entails two processes: the external respiration and internal respiration.

The *internal respiration* or tissue respiration refers to the utilization of O_2 and production of CO_2 by the tissues.

The *external respiration* includes supply of O_2 to the tissues from the environment and excretion of CO_2 released by the tissues into the atmosphere. The process of external respiration involves three major events:

- *Pulmonary ventilation,* i.e. exchange of gases between the environment and lungs. It includes mechanics of respiration.
- *Pulmonary diffusion* refers to transfer of gases from alveoli to the blood by diffusion across the respiratory membrane.
- *Transport of gases* from the blood to the body cells and back.

The *respiratory adjustments in health and diseases* are essential for life; and to understand these, knowledge about *regulation of respiration* is must.

Chapter 5.1

Respiratory Tract: Structure and Functions

FUNCTIONAL ANATOMY

- Respiratory passages
- Pleura and pleural cavity
- Respiratory parenchyma
- Blood supply
- Innervation

FUNCTIONS OF RESPIRATORY SYSTEM

- **Respiratory functions**
- **Non-respiratory functions**
  - Functions subserved by lung defence mechanisms
  - Functions subserved by pulmonary circulation
  - Metabolic functions of lungs
  - Functions subserved by respiratory muscles

FUNCTIONAL ANATOMY

The chief organs of the respiratory system are right and left *lungs.* The oxygen in the atmospheric air reaches the lungs by passing through a series of *respiratory passages* which also serve for the removal of CO_2 from the alveoli to the atmosphere. The respiratory system also includes a pump that ventilates the lung. This pump consists of the chest wall and respiratory muscles. Salient points about the functional anatomy of respiratory passages and lungs are discussed here.

RESPIRATORY PASSAGES

Respiratory passages include the following structures (Fig. 5.1-1):

1. Nasal cavities. The air enters the body through right and left anterior (external) nares which open into the right and left nasal cavities. Through the posterior (internal) nares, the nasal cavities open into the pharynx.

Nasal cavity subserves following functions:

- Nasal cavities warm up the air to body temperature and humidify the air to 100% saturation.
- Clean and filter the air of its particulate contents by channeling the air through tortuous path between the turbinates and prevent the foreign bodies to reach the alveoli. The larger particles (>10 µm diameter) are filtered by hair in the nostrils, whereas the smaller particles (2–10 µm diameter) enter into the bronchi, they initiate reflex coughing and also by ciliary escalator activity move them away (see page 295).

2. Pharynx. From above downwards, the pharynx is divided into nasopharynx, oropharynx and laryngopharynx. Air from the nasal cavities enters the nasopharynx and passes down through the oropharynx and laryngopharynx to larynx. From the mouth the air can directly pass to oropharynx.

3. Larynx. It is continuous above with laryngopharynx and below with trachea. The air passes through the glottis

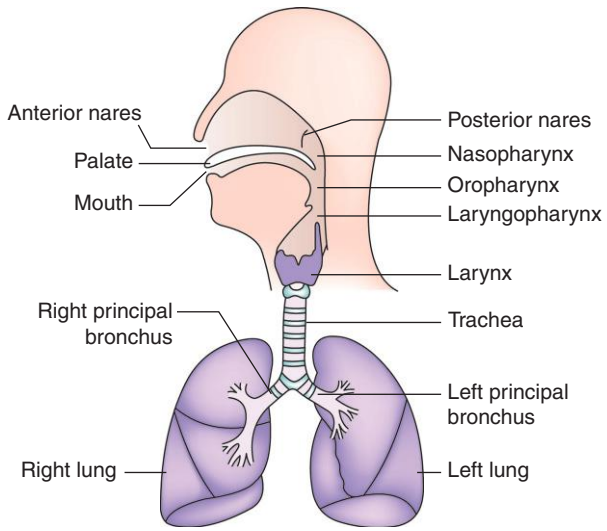


Fig. 5.1-1 The air passages.

(the triangular space between vocal cords) into the trachea. Apart from being a respiratory passage, the larynx also acts as a voice box.

4. Tracheobronchial tree. The air passages between trachea and alveoli divide 23 times to form the extensive tracheobronchial tree (Fig. 5.1-2). These multiple divisions greatly increase the total cross-sectional area of the airway from 2.5 cm^2 in the trachea to 11,800 cm^2 in the alveoli. Consequently, the velocity of air flow in the small airway declines to very low values. The 23 generation divisions of tracheobronchial tree have been numbered as:

The trachea is designated as generation zero.

The principal bronchi, right and left, which are two major divisions of trachea constitute the first generation.

- *Lobar bronchi,* which are division of the principal bronchus form the second generation.
- *Segmental bronchi* which are further division of each lobar bronchus forms the 3rd generation. Each segmental bronchus divides into several generations of branches that ultimately end in very small tubes called *bronchioles.*
- *Terminal bronchioles* is the name given to 16th generation of the divisions. Up to this generation of division, no exchange of gases is possible.

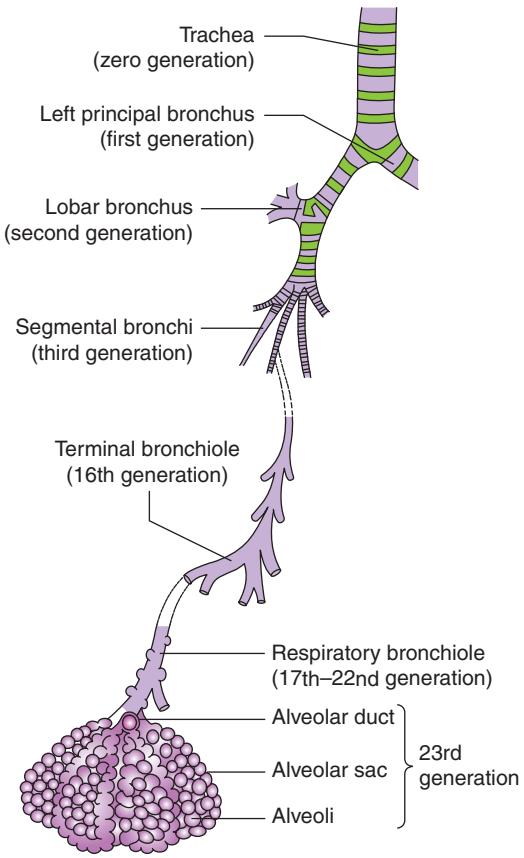


Fig. 5.1-2 The tracheobronchial tree.

- *Respiratory bronchiole* is the name given to 17th–22nd generation of divisions. These are labelled as respiratory bronchioles because some exchange of gases is possible in these tubes.

Alveolar ducts end in the alveoli or the alveolar sacs form the 23rd generation division. It is here that most of the O_2 and CO_2 exchange occurs.

Thus, from the functional point of view the tracheobronchial tree can be divided into two major zones:

(i) *Conducting zone* of the air passages is formed by the first 16 generations of passages and it only transports gases from and to the exterior. Thus, conducting zone starts from trachea and extends up to the terminal bronchioles. Since no exchange of gases is possible here, so starting from nose to the terminal bronchioles forms the so-called *dead space* which has a total capacity of approximately 150 mL.

(ii) *Respiratory zone.* The remaining seven generations of tracheobronchial tree which includes the respiratory bronchioles, alveolar ducts and alveoli form the respiratory zone where exchange of gases occurs. Its volume is approximately 4 L.

Histological features of tracheobronchial tree

1. Cartilaginous rings are present in the trachea and initial bronchi of few generations. These are absent in the terminal bronchioles and respiratory bronchioles. The cartilaginous rings of trachea are incomplete on their posterior aspect. This allows some contraction of trachea but tracheal lumen cannot be completely obliterated easily.

2. Smooth muscles. Ends of cartilaginous rings of trachea are approximated by transverse smooth muscle fibres. In terminal bronchioles they are present in large amount and form a sphincter. Smooth muscles are not present in alveoli.

3. Epithelial lining in trachea and large bronchi is columnar and becomes cuboidal in bronchioles and simple squamous in alveoli. The epithelial cells of tracheobronchial tree are ciliated. Cilia are absent in alveoli. Efficiency of ciliated cells of trachea and bronchi in propelling mucus and waste products is of higher order. Cilia are not influenced by nerve impulses. The mucous secreting goblet cells and deep serous glands are present in trachea and bronchi but are absent in bronchioles and alveoli.

PLEURA AND PLEURAL CAVITY

The lungs are covered by pleura which consists of two layers:

Parietal pleura is the outer layer which lies on the inner side of the chest wall and

Visceral pleura is the inner layer which covers the lung surface.

There is a potential space between layers of the pleura called as *pleural cavity*. The pleural cavity contains a small amount (about 2 mL) of serous fluid (*pleural fluid*). The adhesive and inexpansive nature of the fluid keeps the pleural layers together and help the lungs to slide easily on the chest wall.

RESPIRATORY PARENCHYMA

Each respiratory unit consists of one respiratory bronchiole which opens into a number of alveolar ducts and each alveolar duct in turn opens into number of alveoli. The two lungs contain about 300 million alveoli. Each alveolus has a diameter of about 0.2 µm. The alveoli are surrounded by the pulmonary capillaries. The total area of the alveolar walls in contact with capillaries in both the lungs is about 70 m^2.

Microscopic structure of alveolus (Fig. 5.1-3)

Each alveolus is lined by two types of epithelial cells:

- *Type I cells* are flat cells with large cytoplasmic extensions and are the primary lining cells.
- *Type II cells* (granular pneumocytes) are thicker and contain numerous lamellar inclusion bodies. These cells secrete *surfactant.*

The alveolar wall also contains other special type of epithelial cells:

- *Pulmonary alveolar macrophages* (PAM) which are active phagocytic cells,
- *Plasma cells* which form and secrete immunoglobulins,
- *Amine precursor uptake and decarboxylation (APUD) cells* which store and secrete many biologically active peptides, e.g. vasoactive intestinal peptide (VIP) and substance P, etc.
- *Mast cells* which contain heparin, various lipids, histamine and various proteases that participate in the allergic reactions.

Communication between the two alveoli occurs through small pores called *pores of Kohn.*

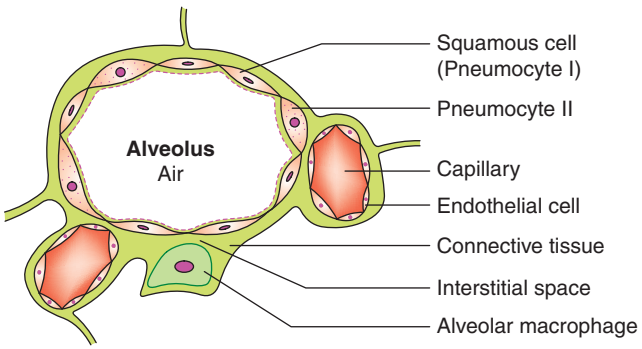


Fig. 5.1-3 Microscopic structure of alveolus.

BLOOD SUPPLY

Conducting airway is supplied by the systemic blood whereas the respiratory zone of the lung is supplied by the deoxygenated (venous) blood coming through pulmonary arteries to lungs. Blood is oxygenated in lungs and is returned to left atrium via pulmonary veins. For details of pulmonary circulation see Chapter 5.3.

INNERVATION

- *Parasympathetic* fibres pass through the vagus nerve. Their stimulation cause bronchoconstriction and increased bronchial secretion via muscarinic receptors. The nerve endings are also activated by leukotrienes, irritants, chemicals (e.g. CO, Pb, NO_2), hydrocarbons and even by cool air.

  Note. Bronchial smooth muscle tone has circadian rhythm, i.e. tone is maximum (maximum bronchoconstriction) in the morning (6 am) and minimum (maximum bronchodilation) in the evening (6 pm).

- *Sympathetic nerves* supplying the lungs when stimulated cause bronchodilation and decreased bronchial secretion via adrenergic receptors (predominantly β_2).
- *Non-cholinergic and non-adrenergic innervations.* Stimulation of these endings results in bron-chodilation due to release of VIP.

  Note. It has been observed that in large number of bronchial asthma patients VIP is absent.

- *Afferents* from the lungs pass through vagii.

FUNCTIONS OF RESPIRATORY SYSTEM

RESPIRATORY FUNCTIONS

The main function of the respiratory system in general and lung in particular is exchange of gases between atmosphere and blood.

NON-RESPIRATORY FUNCTIONS

Besides the respiratory functions, the respiratory system performs many important non-respiratory functions which include:

A. Functions subserved by lung defence mechanisms

1. Immunoglobulin-A (IgA) is secreted in the bronchial secretion and protects against respiratory infections. *Ciliary escalator action* is an important defence system against the air-borne infection. The dust particles in the inhaled air are often laden with bacteria. While passing through the repeatedly branched bronchial tree, the dust

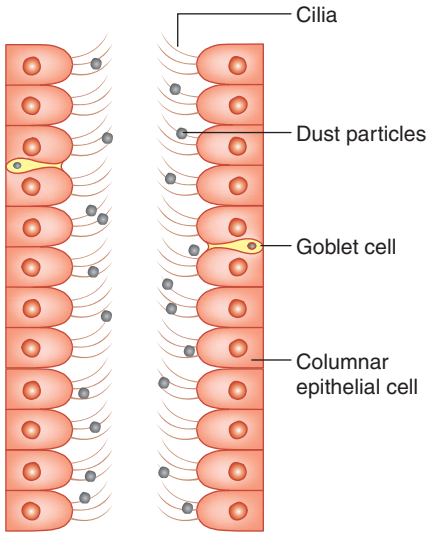


Fig. 5.1-4 Ciliary escalator action of the respiratory mucosa.

particles and the bacteria are caught in the mucous layer present at the mucosal surface of respiratory passages and are moved up towards pharynx by the rhythmic upward beating action of cilia (Fig. 5.1-4) and swallowed. Cigarette smoke disturbs the ciliary function.

2. Pulmonary alveolar macrophages play an important role in defence system.

- Being actively phagocytic cells they ingest the inhaled bacteria and small particles.
- Help in processing inhaled antigens for immunologic attack.
- PAMs secrete substances that attract polymorphonuclear cells to the lungs.
- By some secretions they stimulate granulocyte and monocyte formation in the bone marrow.

3. Cough reflex. The laryngeal, tracheal and bronchial mucous membranes contain vagal afferent terminals which act as *irritant receptors.* Stimulation of these receptors by chemical or mechanical stimuli (excessive mucus, inadvertently inhaled foodstuff, etc.) produces a bout of coughing which helps in expulsion of foreign material.

APPLIED ASPECTS

In Kartagener's syndrome patients, ciliary mobility may be congenitally absent due to the absence of axonemal dynein and ATPase. This condition is also associated with infertility due to lack of sperm motility.

B. Functions subserved by pulmonary circulation

1. Reservoir for left ventricle. When left ventricle output becomes transiently greater than systemic venous return, the blood stored in the pulmonary circulation helps in maintaining the left ventricular output for few strokes.

2. Pulmonary circulation acts as a filter and filters out particles from the blood, which may include: small fibrin or blood clots, detached cancer cells, fat cells, gas bubbles, agglutinated RBCs, masses of platelets, debris from the stored blood.

3. Removal of fluid from alveoli. Because of low pulmonary hydrostatic pressure, the fluid entering the alveoli is absorbed by the capillaries. This protects the gas exchange function of lungs and opposes transudation of fluid from capillaries to the alveoli.

4. Role in absorption of drugs. Certain drugs that rapidly pass through the alveolar capillary barrier by diffusion are administered by inhalation, e.g. anaesthetic gases, aerosol and other bronchodilators.

C. Metabolic functions of lungs

1. Surfactant produced in the lungs plays an important role in respiration. For details see page 306.

2. Conversion of angiotensin I to II is performed by the enzyme angiotensin converting enzyme (ACE) present in the pulmonary capillary endothelium.

3. Inactivation partly or completely of many vasoactive substances present in the blood is done by capillary endothelial cells as they pass through pulmonary circulation. These substances include bradykinin, serotonin, some prostaglandins, norepinephrine, acetylcholine, etc. Amount of serotonin and norepinephrine reaching the systemic circulation is decreased by lungs. Vasoactive substances that pass through the lungs without being metabolized include epinephrine, dopamine, oxytocin, vasopressin and angiotensin I.

4. Fibrinolytic mechanism present in the lung lyses clot in the pulmonary vessels.

5. Storage of hormones and certain biologically active peptides is done in the APUD cells and nerve fibres present in the alveoli. These substances include VIP, substance P, opioid peptides, cholecystokinin-pancreozymin (CCK-PZ) and somatostatin. These substances are later released into the systemic circulation.

D. Functions subserved by respiratory muscles

Respiratory muscles are also used during laughing and singing.

Chapter 5.2

Pulmonary Ventilation

INTRODUCTION

As we know, respiration, to be more precise, the *external respiration* (supply of O_2 from atmosphere to body tissues and removal of CO_2 from the body to atmosphere) involves three major processes:

- *Pulmonary ventilation,* i.e. exchange of gases between the environment and lungs.
- *Pulmonary diffusion,* i.e. transfer of gases from the alveoli to capillary blood across the respiratory membrane.
- *Transport of gases* from the blood to tissue cells and back.

In this chapter, following aspects and concepts related to pulmonary ventilation are discussed:

- Mechanics of pulmonary ventilation,
- Lung volumes and capacities,
- Pulmonary elastance and compliance and
- Work of breathing.

MECHANICS OF PULMONARY VENTILATION

MECHANISM OF BREATHING

Pulmonary ventilation is accomplished by two processes:

- *Inspiration* refers to the inflow of atmospheric air into the lungs. This obviously occurs when the intrapulmonary pressure falls below the atmospheric air pressure.
- *Expiration* refers to the outflow of air from the lungs into the atmosphere. This obviously occurs when the intrapulmonary pressure rises above the atmospheric air pressure.
- *Changes in intrapulmonary pressure* which govern the respiratory cycle of inspiration and expiration are related to changes in the intrapleural pressure.
- *Changes in intrapleural pressure* are brought about by the changes in the size of thoracic cavity. Expansion of thoracic cage leads to fall in the intrapleural pressure

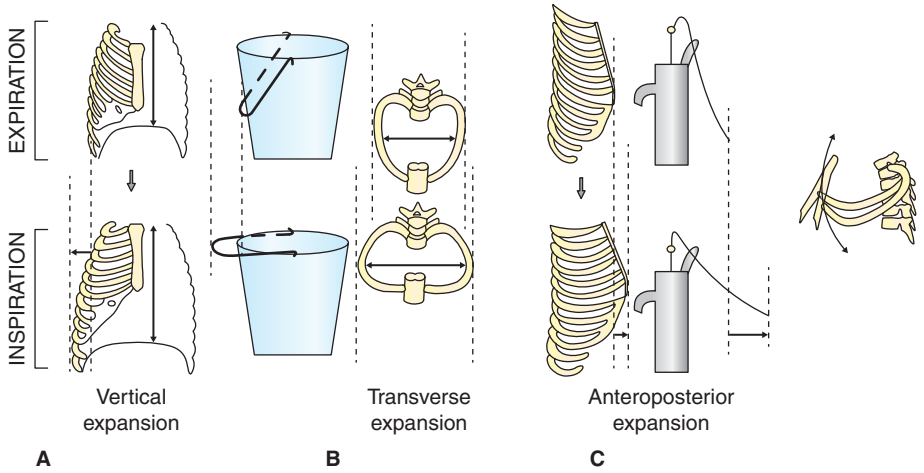


Fig. 5.2-1 Mechanism of increase in diameter of thoracic cavity: A, increase in vertical diameter (descent of diaphragm); B, increase in transverse diameter (bucket handle effect) and C, increase in anteroposterior diameter (pump handle effect).

and decrease in the size of thoracic cavity leads to rise in the intrapleural pressure.

- *Changes in size of the thoracic cavity* are brought about by the actions of respiratory muscles:
  - Muscles of normal tidal inspiration are diaphragm and external intercostal muscles.
  - Accessory muscles of inspiration are scaleni, sternomastoid and serratus anterior and alae nasi.
  - Muscles of expiration are internal intercostal muscles and abdominal muscles (abdominal recti muscles, transverse abdominis muscles and internal oblique muscles).

MECHANISM OF TIDAL RESPIRATION

Inspiration

Inspiration is an active process, normally produced by contraction of the *inspiratory muscles (negative-pressure breathing).* Use of respirator to inflate the respiratory system produces positive pressure *(positive-pressure breathing).* During tidal inspiration (quiet breathing), the diaphragm and external intercostal muscles contract and cause increase in all the three dimensions of thoracic cavity.

Role of diaphragm. The diaphragm is a dome-shaped, musculotendinous partition between thorax and abdomen. The muscle fibres of diaphragm arise from xiphisternum, inner surface of lower six ribs and lumbar vertebrae. The convexity of this dome is directed towards the thorax. *Innervation*: the motor supply to diaphragm is by phrenic nerve (C_3, C_4, C_5). In tidal inspiration (quiet breathing), 70–75% of expansion of chest is caused due to contraction of diaphragm.

When the diaphragm contracts following changes occur:

- The dome becomes flattened and the level of diaphragm is lowered *increasing the vertical diameter* of the thoracic cavity (Fig. 5.2-1A). During quiet breathing, the descent of diaphragm is about 1.5 cm and during forced inspiration it increases to 7 cm.
- The descent of diaphragm causes rise in the intra-abdominal pressure which is accommodated by the reciprocal relaxation of the abdominal wall musculature.
- Contraction of diaphragm also lifts the lower ribs causing thoracic expansion laterally and anteriorly (the bucket handle and pump handle effect, respectively) (Fig. 5.2-1B and C). The abdominal organs support the diaphragmatic dome and act as fulcrum while the diaphragmatic contraction raises the lower ribs.

Role of external intercostal muscles. The fibres of external intercostal muscles slope downward and forward. They are attached close to the vertebral ends of the upper ribs than the lower ribs (Fig. 5.2-2). From pivot-like joint with the vertebrae the ribs slope obliquely downwards and forwards. So, when the external intercostal muscles contract (because of the lever effect) the ribs are elevated causing lateral and anteroposterior enlargement of thoracic cavity due to the so-called bucket handle and pump handle effects, respectively (Fig. 5.2-1B and C).

Role of laryngeal muscles. The abductor muscles of the larynx contract during inspiration pulling the vocal cords apart.

APPLIED ASPECTS

- In bilateral phrenic nerve palsy, adequate ventilation at rest can be maintained by the external intercostal muscles alone but respiration is somewhat laboured to maintain life.
- Spinal cord transaction above the level of third cervical segment is fatal without artificial respiration, whereas in transaction below the fifth cervical segment respiration remains normal as the phrenic nerve (C_3, C_4, C_5) that innervates the diaphragm remains intact.

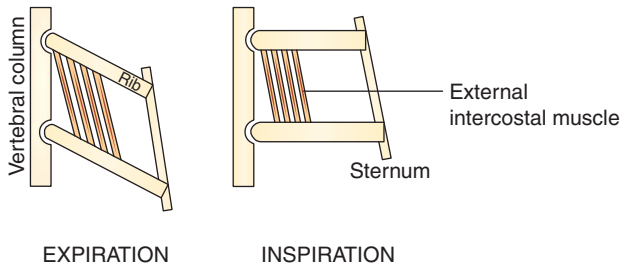


Fig. 5.2-2 Contraction of external intercostal muscle favours elevation of ribs due to mechanical leverage effect of the attachments of its fibres close to the pivot on the upper ribs as compared to the lower ribs.

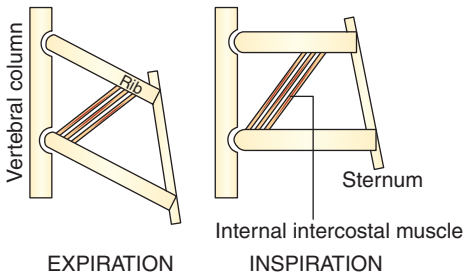


Fig. 5.2-3 Contraction of internal intercostal muscle pulls the ribs down due to mechanical leverage effect of its fibres close to the pivot on the lower ribs as compared to the upper ribs.

Expiration

Expiration in quiet breathing is largely a passive phenomenon and is brought about by the:

- Elastic recoil of the lungs,
- Decrease in size of the thoracic cavity due to relaxation of diaphragm and external intercostal muscles,
- An increase in the tone of muscle of the anterior abdominal wall which forces the relaxing diaphragm upward and
- The serratus posterior inferior muscles play a minor role in pulling down the ribs.

MECHANISM OF FORCED RESPIRATION

Forced inspiration

Forced inspiration is characterized by:

1. Forceful contraction of diaphragm leading to descent of diaphragm by 7–10 cm as compared to 1–1.5 cm during quiet inspiration.

2. Forceful contraction of external intercostal muscles causing more elevation of ribs leading to more increase in transverse and anteroposterior diameter of thoracic cavity.

3. Contraction of accessory muscles of inspiration which cause the following effects:

- Sternomastoid muscles contract and lift the sternum upwards,
- Anterior serrati muscles contract and lift many of the ribs upwards and
- Scaleni muscles contract and lift first two ribs.

Forced expiration

Forced expiration is required when respiration is increased during exercise or in the presence of severe respiratory disease. It is an active process caused as follows:

1. Contraction of abdominal muscles (abdominal recti, transversus abdominis, internal and external oblique).

- Compression of the abdominal contents which increases the intra-abdominal pressure and forces the diaphragm upward thereby reducing vertical diameter of the thoracic cavity.
- Downward pull on the lower ribs and thus decrease the anteroposterior diameter of the thoracic cavity.
- Fixation of the lower ribs so that internal intercostal muscles act more effectively.

2. Contraction of the internal intercostal muscles causes the effect which is just opposite to that of the external intercostal muscles. This is because of the leverage mechanism of the direction of the muscle fibres which slope downward and backward creating a longer force arm for the upper ribs (Fig. 5.2-3). Hence, their contraction tends to pull all the ribs downwards reducing: anteroposterior diameter (because of falling of pump handle effect) as well as the transverse diameter (because of action of ribs like falling of bucket handle) of the thoracic cavity.

Note. Besides their role in deep breathing, the expiratory muscles are also involved in other forced expiratory efforts, e.g. in coughing, vomiting, defaecation and Valsalva manoeuvre etc.

PRESSURE AND VOLUME CHANGES DURING RESPIRATORY CYCLE

Intrapulmonary pressure changes during respiratory cycle (Fig. 5.2-4)

The movement of air in and out of the lungs depends primarily on the pressure gradient between the alveoli and the atmosphere (i.e. *transairway pressure*). Intrapulmonary or alveolar pressure is the air pressure inside the lung alveoli.

At end-expiration and end-inspiration, i.e. when the glottis is open and there is no movement of air, the pressures in all parts of the respiratory tree are equal to atmospheric pressure, the intrapulmonary pressure is considered to be 0 mm Hg.

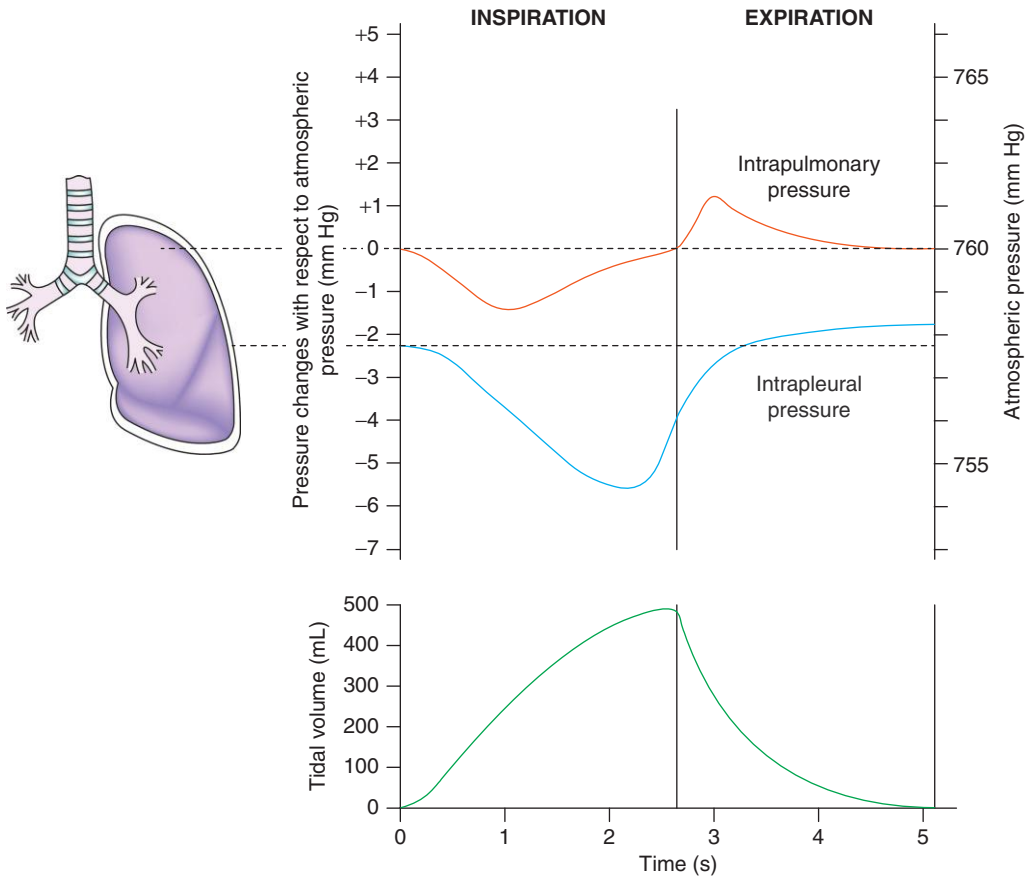


Fig. 5.2-4 Pressure and volume changes during respiratory cycle.

During inspiration in quiet breathing, the pressure in the alveoli decreases to about –1 mm Hg, which is sufficient to suck in about 500 mL of air into the lungs within 2 s period of inspiration. At the end-inspiration, the intrapulmonary pressure again becomes zero.

During forced inspiration against a closed glottis (Muller's manoeuvre), the intrapulmonary pressure may be as low as –80 mm Hg below the atmospheric pressure.

During expiration in quiet breathing, the elastic recoil of the lungs causes the intrapulmonary pressure to swing slightly to the positive side (+1 mm Hg) which forces the 500 mL of inspired air out of the lungs during 2–3 s of expiration. At the end-expiration, once again the alveolar pressure regains the atmospheric pressure (0 mm Hg).

Forceful expiration against the closed glottis (Valsalva's manoeuvre) may produce intrapulmonary pressure of as much as 100 mm Hg.

Intrapleural (pleural) pressure changes during respiratory cycle (Fig. 5.2-4)

Pleural pressure is the pressure of fluid in the space between the visceral pleura and parietal pleura.

Normal pleural pressure when the respiratory muscles are completely relaxed and the airways are open is about –2.5 mm Hg.

- The negative pleural pressure (–2.5 mm Hg) is the amount of suction required to hold the lungs at their equilibrium volume or the functional residual capacity (FRC). FRC is the lung volume at the end of normal (eupnoeic), relaxed expiration and is about 2–2.5 L of gas.
- The negative pleural pressure (–2.5 mm Hg) is the result of balance of two opposite forces, the recoil tendency of the lungs and the recoil tendency of thoracic cage.
- Recoil tendency of the lungs (continuous tendency to collapse) is caused by:
  - The presence of many elastic fibres in the alveolar walls which are under constant stretch in the inflated lungs.
  - Surface tension of the fluid lining the alveoli due to which the alveoli tend to become progressively smaller and collapse.
- Recoil tendency of the thoracic cage, i.e. a constant tendency to expand (to pop outward) is because of the fact that the chest wall is an elastic structure which is normally partially pulled inward. The elastic property of the

thoracic cage is because of the elastic nature of ribs, muscles and tendons.

During inspiration due to expansion of the chest wall the pleural pressure becomes still more negative (–6 mm Hg) and pulls the surface of lungs with greater force creating negative intrapulmonary pressure.

During expiration. At the end of inspiration, the inspiratory muscles relax and the recoiling force of lungs begins to pull the chest wall back to expiratory position. At end-expiratory position where the recoil force of the lungs and recoil force of thoracic cage balance, the pleural pressure returns back to –2.5 mm Hg.

Factors affecting pleural pressure

1. Deep inspirations. The pleural pressure may become as low as –30 mm Hg during deep inspiration.

2. Valsalva manoeuvre or when expiratory muscles are working against closed glottis, there occur marked decrease in the thoracic volume causing deflation of lungs. Under such circumstances, the intrapleural pressure can become positive by 60–70 mm Hg (greater than atmospheric pressure), e.g. during defaecation and coughing.

3. Gravity. The pleural pressure in the standing position is more negative (–7 mm Hg) at the apices of the lungs as compared to the bases (–2 mm Hg). Because of this fact, lungs are less expanded at the base and even the airway at the base may become closed at the end of expiration. This is why during the first part of inspiration, more of the inspired gas goes to the apices than bases of lungs.

4. Injury to the chest wall causing exposure of pleural cavity to the exterior causes air to enter the pleural cavity (pneumothorax) and the pleural pressure rises from subatmospheric level to atmospheric level. Since there is no opposing force to elastic recoil of the lung, so in pneumothorax the lung is collapsed.

5. Emphysema is a disease of the lungs in which lung elasticity is decreased or lost. Due to decrease in the recoiling force of lungs the intrapleural pressure becomes less negative (i.e. increases). Because of this reason in this disease the lung alveoli expand and chest also expands and becomes barrel shaped.

Measurement of pleural pressure

1. Manometric measurement can be made by inserting a needle into the intrapleural space whose other end is attached to a water manometer with the help of rubber tubing.

2. Intraoesophageal pressure can be recorded with the help of an air containing rubber balloon sealed over a catheter placed in the lower part of thoracic part of oesophagus. Intraoesophageal pressure in the thoracic part is equivalent to the intrapleural pressure because this part of oesophagus becomes a closed cavity due to closure of its lower end by cardiac sphincter and upper end by closure of glottis.

Lung volume changes during respiratory cycle

During tidal inspiration, the volume of air in the lungs increases by 500 mL (tidal volume).

During tidal expiration, the elastic forces compress the gas in the lungs which starts flowing out and at the end of expiration the volume of air in the lungs decreases by 500 mL (Fig. 5.2-4).

LUNG VOLUMES AND CAPACITIES

STATIC LUNG VOLUMES AND CAPACITIES

LUNG VOLUMES

The maximum volume to which a lung can be expanded has been divided into four non-overlapping volumes (Fig. 5.2-5):

1. Tidal volume (TV). It is the volume of air inspired or expired with each breath during normal quiet breathing. It is approximately 500 mL in normal adult male.

2. Inspiratory reserve volume (IRV). It is the extra volume of air that can be inhaled by a maximum inspiratory effort over and beyond the normal tidal volume. It is about 3000 mL (range 2000–3200 mL) in a normal adult male.

3. Expiratory reserve volume (ERV). It is the extra volume of air that can be exhaled by the maximum forceful expiration

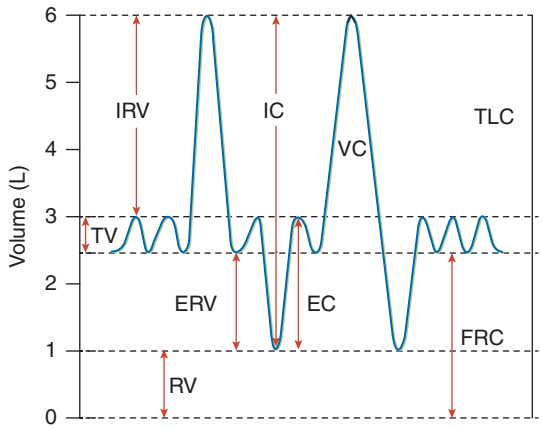


Fig. 5.2-5 A spirogram showing various lung volumes and capacities: RV = residual volume; TV = tidal volume; IRV = inspiratory reserve volume; ERV = expiratory reserve volume; FRC = functional residual capacity; TLC = total lung capacity; IC = inspiratory capacity; VC = vital capacity; EC = expiratory capacity.

over and beyond the normal tidal volume, (i.e. after the end of normal passive expiration). It is approximately 1100 mL in a normal adult male.

4. Residual volume (RV). It is the volume of the air that still remains in the lungs after the most forceful expiration. It is about 1200 mL in a normal adult male. RV can be calculated from function residual capacity.

LUNG CAPACITIES

Lung capacities are combination of two or more pulmonary volumes and include (Fig. 5.2-5):

1. Inspiratory capacity. This is the maximum volume of the air that can be inspired after normal tidal expiration. Therefore, it equals the tidal volume plus inspiratory reserve volume (TV + IRV) and is approximately 3500 mL in a normal adult male.

2. Expiratory capacity. It is the maximum volume of air that can be expired after normal tidal inspiration. It equals tidal volume plus expiratory reserve volume (TV + ERV) and is approximately about 1600 mL in a normal adult male.

3. Functional residual capacity. It is the volume of the air remaining in the lungs after normal tidal expiration. Therefore, it equals the expiratory reserve volume plus the residual volume (ERV + RV) and is about 2300 mL in a normal adult male.

Significance of FRC. The FRC has several advantages:

(i) *Continuous exchange of gases* is possible due to the presence of some air always in the lungs and thereby concentration of O_2 and CO_2 in blood is maintained constant. Without FRC, PO_2 would have risen to 150 mm Hg during inspiration and reduced nearly to zero during expiration, which is maintained at about 100 mm Hg due to the FRC.

(ii) *Breath holding* is made possible due to the FRC.

(iii) *Dilution of toxic inhaled gases* occurs due to the reserve of 2300 mL of air in the lungs (FRC) most of the times.

(iv) *Load on respiratory mechanism and left ventricle would* have been much more if there was no FRC since:

- Without FRC lungs would have collapsed at the end of each expiration and re-expansion would have required tremendous breathing effort and
- The collapsed lungs would have increased pulmonary vascular resistance and imposed a heavy load on the left ventricle.

Factors affecting FRC. Hyperinflation of the lungs seen in following conditions may be associated with increased FRC:

- Old age due to loss of elasticity of lungs,
- Emphysema and
- Bronchial asthma.

Measurement of FRC. Functional residual capacity can be measured by nitrogen wash-out method or helium dilution method.

Nitrogen wash-out method for measuring FRC. This method is based on the assumption that the alveolar air has 80% nitrogen. In this method, the subject is made to wash out nitrogen from the lungs completely by inhaling pure O_2 for 5 min and expiring into a large gas bag called Douglas bag (washed with pure O_2 and hence made nitrogen free). The volume of gas collected in the *Douglas bag* and its nitrogen content are measured and value of FRC is calculated as:

- Suppose the expired gas collected at the end of 5 min is 40 L and its nitrogen concentration is 5%.
- Then volume of nitrogen (Vn) in the expired gas =

$$\frac{40 \times 5}{100} = 2\,\text{L or } 2000\,\text{mL}$$

- Since the nitrogen content of the expired gas is solely from the subject's lungs where it forms 80% of the gas present in the lungs (FRC); therefore;

$$\text{FRC} = \frac{2000 \times 100}{80}\,\text{mL}$$
$$= 2500\,\text{mL or } 2.5\,\text{L}$$

Helium dilution method for FRC estimation. This method is based on the basic principle that the amount (A) of a substance present is equal to its volume (V) multiplied by its concentration (C), i.e.

$$A = V \times C$$

- In this method, a closed circuit system is prepared with a mixture of helium, oxygen and air. The amount of helium added in the mixture is such as to achieve a helium concentration of 10%.
- The subject is then made to breathe in the closed system till the concentration of helium in the lungs and the bag becomes equal. The helium concentration is measured at the end of a normal expiration (because the FRC is the volume of air remaining in the lungs at the end of normal expiration).
- Since no helium was present in the lungs to start with so the same amount of helium was added to the mixture in the close circuit system gets distributed between the bag and the lung (Fig. 5.2-6A and B).
- The FRC can be calculated from the following known values as:
  - Suppose the volume of helium added in the mixture to make its concentration to 10% is 560 mL,
  - Volume of the bag is 5000 mL and
  - Concentration of helium (He) in bag and lungs after breathing in close circuit system is 8%.

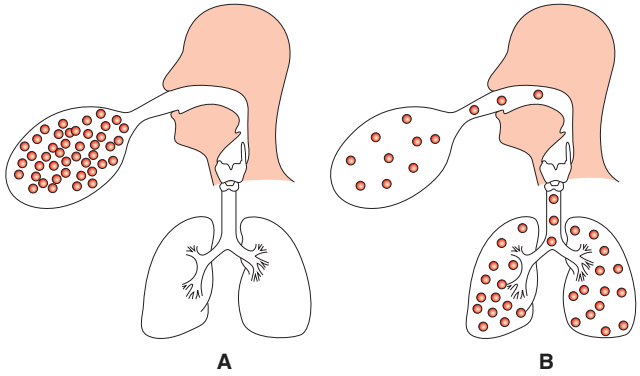


Fig. 5.2-6 Helium dilution method for estimation of functional residual capacity (FRC): A, in the beginning no helium present in the lungs and B, the helium present in close circuit distributed equally between bag and the lungs.

$$\text{Then volume of He} = \frac{(\text{Volume of lungs} + \text{bag})}{\text{Concentration of helium}} \times \frac{8}{100}$$

$$\text{i.e.} \qquad 560 = (\text{volume of lungs} + 5000) \times \frac{8}{100}$$

or

$$\text{volume of lungs} + 5000 = \frac{560 \times 100}{8}$$

or

$$\text{volume of lungs} = 7000 - 5000$$

Hence FRC = 2000 mL.

4. Vital capacity (VC). This is the maximum amount of air a person can expel from the lungs after the deepest possible inspiration. Therefore it equals tidal volume plus inspiratory reserve volume plus expiratory reserve volume (TV + IRV + ERV) and is about 4600 mL in a normal adult male.

Significance of vital capacity

- Estimation of VC allows assessment of maximum inspiratory and expiratory efforts and thus gives useful information about strength of the respiratory muscles.
- VC also provides useful information about other aspects of pulmonary functions through forced expiratory volume.

Factors affecting vital capacity

A. Physiological

(i) *Size of the thoracic cavity.* VC is more in males (2.6 L/m^2 body surface area) because of large chest size and more muscle power than females.

(ii) *Age.* In old age, VC is decreased due to decrease in elasticity of the lungs.

(iii) *Strength of respiratory muscles.* In swimmers and divers, VC is more because of the increased strength of the respiratory muscles.

(iv) *Gravity.* In standing position, VC is more than in sitting and lying position because of:

- Increased size of thoracic cavity in standing (diaphragm moves down) and
- Reduced pulmonary blood flow due to the decreased venous return.

(v) *In pregnancy,* VC is reduced due to pushing up of diaphragm and reduced capacity of the thoracic cavity.

B. Pathological

(i) *In ascites* (accumulation of fluid in the abdominal cavity), VC is reduced due to same reason as in pregnancy.

(ii) *Pulmonary diseases* like pulmonary fibrosis, emphysema, respiratory obstruction, pulmonary oedema, pleural effusion and pneumothorax are associated in decreased VC.

5. Total lung capacity (TLC). It is the volume of air present in the lungs after the maximal inspiration. It equals the vital capacity plus the residual volume (VC + RV) and is about 5800 mL in a normal adult male.

Calculation of residual volume from the FRC. Since the FRC is the volume of air remaining in the lungs after normal tidal expiration, so it equals the expiratory reserve volume plus the residual volume (ERV + RV). Therefore, RV = FRC – ERV.

Calculation of total lung capacity. Since total lung capacity is the volume of air present in the lungs after the maximal inspiration, so it equals the vital capacity plus residual volume (VC + RV).

Measurement of static lung volumes and capacities. All volumes and capacities except residual volume, functional residual capacity and total lung capacity are recorded by a *spirometer.*

Recording of lung volumes and capacities are the important lung function tests (for details see page 365).

DYNAMIC LUNG VOLUMES AND CAPACITIES

Timed vital capacity (TVC) or forced vital capacities (FVC)

Forced vital capacity is the volume of the air that can be expired rapidly with a maximum force following a maximum inspiration. The volume of air expired can be timed by recording the vital capacity on a spirograph moving at the known speed.

Components of TVC or FVC *(Fig. 5.2-7)*

(i) Forced expiratory volume in 1 s (FEV_1). It represents the volume expired in the first second of a FVC.

- Estimation of FEV_1 is the most commonly used screening test for airway diseases.
- The FEV_1 is actually a flow rate: its unit are L/s.
- FEV_1% is the percent of FVC expired in 1 s (i.e. $FEV_1\% = FEV_1/FVC \times 100$) normally FEV_1% is about 80% of the FVC (Fig. 5.2-8A).

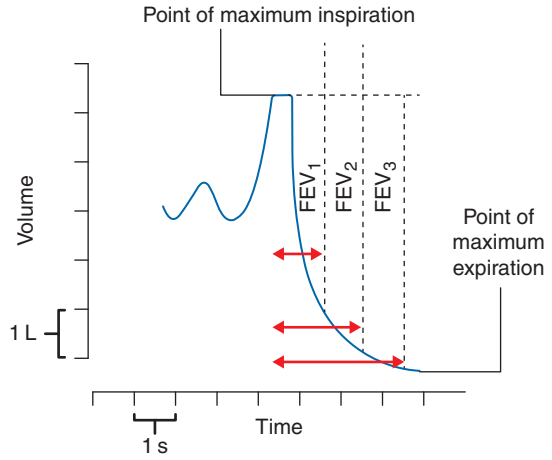


Fig. 5.2-7 Components of timed vital capacity.

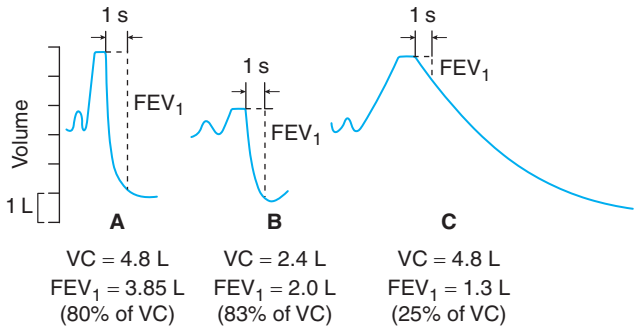


Fig. 5.2-8 Forced expiratory volume in first second (FEV_1) component of timed vital capacity: A, in normal subject; B, in a patient with restrictive lung disease (RLD) and C, in a patient with obstructive lung disease (OLD).

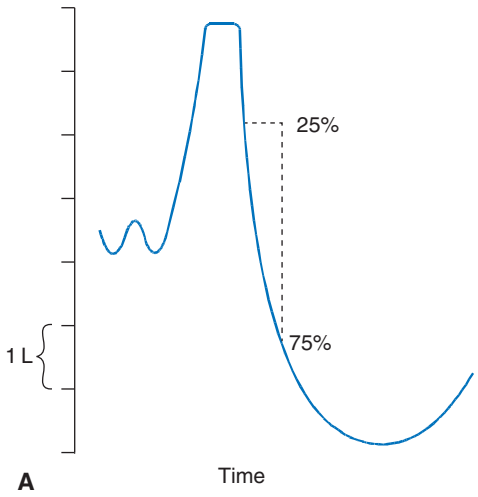
Clinical application. Useful in distinguishing between restrictive and obstructive lung diseases:

- Patients with *restrictive lung disease* (e.g. kyphoscoliosis and ankylosing spondylitis) have a reduced FVC but are able to achieve relatively high flow rates; therefore their FEV_1% exceeds 80% (Fig. 5.2-8B).
- Patients with *obstructive lung disease* (e.g. bronchial asthma) have low flow rates as a result of high airway resistance therefore their FEV_1% is abnormally low (Fig. 5.2-8C).

Note. FEV_1 is reproducible and is highly sensitive index of obstructive lung disorder. However, it does not differentiate the different causes of obstructive lung diseases.

(ii) Forced expiratory volume in 2 s (FEV_2). It represents the volume of air expired in first 2 s of FVC, FEV_2% is about 90% of FVC under normal condition.

(iii) Forced expiratory volume in 3 s (FEV_3). It represents the volume of air expired in first 3 s. Normally, FEV_3% is 98–100% of FVC.



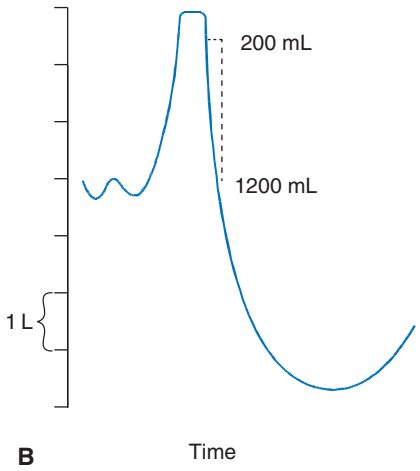


Fig. 5.2-9 Forced expiratory flow: A, during 25–75% of expiration (FEF 25–75%) and B, during 200–1200 mL of expiration (FEF 200–1200 mL).

Forced expiratory flow during 25–75% of expiration (FEF 25–75%)

- It is the mean expiratory flow rate during middle 50% of FVC (Fig. 5.2-9A).
- Normal FEF 25–75% is 300 L/min.
- Mid expiratory time (MET) refers to the time taken for FEF 25–75%. Its normal value is 0.5 s which is increased in obstructive lung disorders.

Forced expiratory flow during 200–1200 mL of expiration (FEF 200–1200)

- FEF 200–1200 refers to the mean expiratory flow rate between 200–1200 mL segments of FVC (Fig. 5.2-9B).
- Normal values of FEF 200–1200 is 350 L/min.

Minute ventilation (MV) or pulmonary ventilation (PV)

It is the volume of air inspired or expired per minute. It equals the tidal volume multiplied by respiratory rate (TV × RR) the

TV at rest averages 500 mL (0.5 L) and the normal respiratory rate is 12–15 breaths/min; therefore normal minute volume is 6–7.5 L/min.

Maximum breathing capacity (MBC) or maximum voluntary ventilation or maximum ventilation volume (MVV)

It is the maximum volume of air that can be ventilated on command during a given interval.

This index of ventilatory function depends upon the complete co-operation of the subject, who is asked to breathe as rapidly and deeply as he/she can for a 15 s interval. The volume of the air moved is either recorded by a spirometer fitted with a writing point or is collected in a Douglas bag; the result is expressed as L/min.

Normal adult male can attain a maximum ventilation volume (MVV) of 80–170 L/min (average 120 L/min).

MVV is profoundly reduced in patients with emphysema, airway obstruction and very poor respiratory muscle strength.

IMPORTANT NOTE

Hyperventilation causes CO_2 washout which leads to respiratory depression. This may result in fainting, therefore, voluntary hyperventilation should be done for brief periods (15 s only).

Pulmonary reserve (PR) or breathing reserve

- PR refers to the maximum amount of the air above the pulmonary ventilation that can be inspired or expired in 1 min.
- It equals maximum ventilation volume minus pulmonary ventilation (minute ventilation), i.e.

$$PR = MVV - PV/min$$

- Pulmonary reserve is usually expressed as percentage of MVV and is known as percentage pulmonary reserve or *dyspnoeic index (DI), i.e.*

$$DI = \frac{MVV - PV}{MVV} \times 100$$

- Normal values of DI or % PR range from 70% to 95% with an average of 75%.
- Dyspnoea is usually present when the value of DI becomes less than 60%.

PULMONARY ELASTANCE AND COMPLIANCE

PULMONARY ELASTANCE

Elastance refers to the *recoil* (retractive) tendency of a structure. Both the thoracic cage and lungs have elastance.

The lungs and chest wall are elastic structures. With the initiation of breathing at birth on first inspiration there is enlargement of chest and lungs (virtually lungs are dragged after chest wall because of adhesive and inexpansile properties of pleural fluid present in the pleural cavity that keeps the two layers of pleura together). At the end of expiration their tendency to recoil from chest wall is balanced by tendency of chest wall to recoil in the opposite direction.

ELASTANCE OF THORACIC CAGE

Elastance or recoil tendency of the thoracic cage refers to the constant tendency of the thoracic cage to expand (to pop outward). The elastance of the thoracic cage is because of the fact that the chest wall is an elastic structure which is normally kept partially pulled inward. The elastic property of the thoracic cage is because of the elastic nature of ribs, muscles and tendons.

ELASTANCE OF LUNGS

Elastance or recoil tendency of the lungs refers to the constant tendency of the lungs to collapse. The recoil forces in the lungs are generated by:

Tissue forces. These are due to the presence of many elastic tissues such as smooth muscle, elastic and collagen in the lung parenchyma which are kept under constant stretch in the inflated lungs.

Surface forces. These are generated at the alveolar surface lined by fluid *(alveolar surface tension)* due to which the alveoli tend to become progressively smaller and tend to collapse.

ALVEOLAR SURFACE TENSION

Alveolar surface tension is generated because of the unbalanced attraction of the liquid molecules at the surface of alveolar membrane. A phase change occurs between the alveolar gas and the surface of the alveolar membrane. Alveolar surface tension has a tendency to reduce the size of each alveolus, thus resulting into recoil tendency of the lung, i.e. surface tension increases the tendency of the lungs to deflate. An increased transmural pressure is necessary to counteract the effects of surface tension. According to the *law of Laplace,* in a spherical structure like alveoli, the transmural pressure generated equals two times the (surface) tension divided by radius, i.e. P = 2T/r. Therefore, small alveoli tend to become still smaller whereas large alveoli tend to become still larger (Fig. 5.2-10). Thus, surface tension tends to produce collapse of the alveoli.

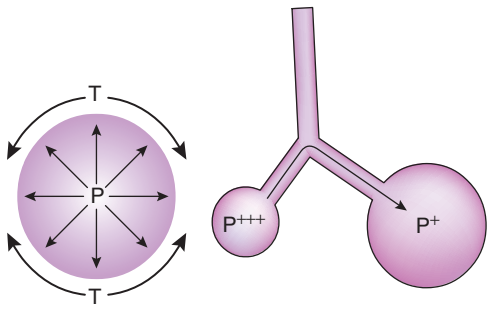


Fig. 5.2-10 Pressure in the alveoli (P) is directly proportional to the surface tension (T) and inversely proportional to the radii (r). Therefore small alveoli tend to become still smaller and large alveoli tend to become still larger.

PULMONARY SURFACTANT

Pulmonary surfactant is a complex mixture of several phospholipids, proteins and ions, and is secreted by type II alveolar epithelial cells (granular pneumocytes). Four unique proteins have been identified in surfactant: SP-A, SP-B, SP-C and SP-D. SP-A and SP-D are hydrophilic proteins, and SP-B and SP-C are strongly hydrophobic proteins. Dipalmitoyl-phosphatidylcholine along with several other phospholipids is responsible for reducing surface tension.

Surface tension with normal alveolar fluid lining without surfactant is about 50 dynes/cm^2 and with surfactant it varies between 5 and 30 dynes/cm^2, depending upon the concentration of surfactant.

Mechanism of action. One portion of each phospholipid molecule is hydrophilic and dissolves in the water lining the alveoli. Lipid hydrophobic portion of the molecule is oriented towards the air. This causes spreading of surfactant molecules over the surface of fluid lining the alveoli. Apoproteins and calcium ions are responsible for uniform and quick spreading of surfactant molecules over the surface. Such a hydrophobic surface exposed to air has one-twelfth to one-half of the surface tension of a pure water surface depending upon the concentrations and orientation of surfactant molecules on the surface.

Functions of pulmonary surfactant

1. Reduces the tendency of alveoli to collapse. The surface tension in a thin walled sphere like alveolus tends to make the sphere smaller to collapse. Surfactant, by reducing surface tension decreases the tendency to collapse.

2. Reduces work of breathing. According to the Laplace's law, due to reduction in the surface tension, the mean alveolar radius is increased. This reduces the transmural pressure required for expanding the alveoli. As alveoli are easily expanded, so work of breathing is reduced. The low surface tension also facilitates the reopening of collapsed airway and alveoli.

3. Prevents pulmonary oedema. Surface tension is retracting force which not only pulls alveolar wall to the centre of alveolus but also pulls fluid from the capillaries into the interstitial space surrounding the alveoli and into the alveoli leading to pulmonary oedema. Surfactant prevents this phenomenon by lowering the surface tension.

4. Alveolar stabilization. Surfactant causes stability of alveoli, i.e. it maintains almost uniform size of alveoli.

Alveolar instability due to surface tension effect is produced as follows:

- In the two alveoli (with unequal size) connected to each other the amount of pressure generated in each according to the Laplace's law will be:

$$\text{Pressure} = \frac{2 \times \text{Surface tension}}{\text{Radius}}$$

- Thus, when the surface tension is constant, the pressure developed in the smaller alveolus will be more than the larger alveolus. This will cause pushing of air from the smaller alveolus (with higher pressure) to the larger alveolus (with lower pressure). As a result, the smaller alveolar sac will become smaller and larger alveolar sac will become larger (Fig. 5.2-10).
- The above cycle of pushing air from smaller to larger sac will continue till the smaller sac totally collapses leading to large distension of the other sac, thereby producing instability of the alveoli.

Pulmonary surfactant causes alveolar stabilization by following mechanisms:

- In the presence of surfactant, the surface tension developed in the alveoli is inversely proportionate to the concentration of surfactant per unit area.
- In the smaller alveolus, the surfactant molecules form a thick layer while in a larger alveolus the surfactant molecules are scattered on the larger surface (as number of molecules is limited) (Fig. 5.2-11).
- Thus, in a smaller alveolus the tendency to develop more pressure due to smaller radius will be neutralized by the tendency to develop less pressure due to more concentration of surfactant per unit area and the reverse will occur in larger alveolus. In this way, there will not be much pressure gradient between the two alveoli helping to maintain the size of alveolar sac constant. Thus, due to the presence of surfactant, stability of alveoli is maintained.

Factors affecting pulmonary surfactant

Factors which decrease the pulmonary surfactant

- Long-term inhalation of 100% O_2,
- Occlusion of main bronchus,

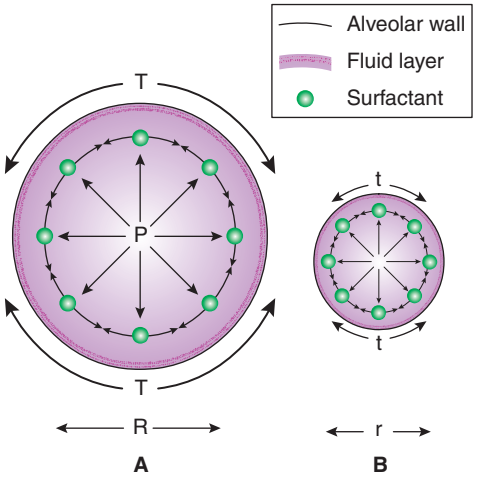


Fig. 5.2-11 Distribution of surfactant in a large (A) and small (B) alveoli.

- Occlusion of one pulmonary artery,
- Cigarette smoking, and
- Cutting both the vagii.

Factors which increase surfactant production

- Thyroid hormones increase the secretion of pulmonary surfactant by increasing the size and number of inclusion bodies in type II alveolar lining epithelial cells.
- Glucocorticoids accelerate the maturation of pulmonary surfactant.

Clinical significance of pulmonary surfactant

1. Respiratory distress syndrome (RDS) of newborn. RDS of newborn is also called hyaline membrane disease due to the formation of hyaline (translucent) membrane by an albuminous fluid in the wall of alveoli and respiratory bronchioles. It occurs in the newborn babies (especially premature) due to inadequate formation of surfactant, resulting in an elevated alveolar surface tension. In this condition, it is extremely difficult to expand the lungs. Respiratory work is greatly increased and there is inadequate exchange of gases due to the alveolar instability, pulmonary oedema, and collapse of alveoli (atelectasis) in many areas. This result into severe respiratory insufficiency and the infant may die. Plasma levels of thyroid hormones and cortisol are low in infants with RDS. *Therapy* of RDS includes administration of exogenous surfactant and application of positive-end expiratory pressure (PEEP).

2. Adult respiratory distress syndrome occurs due to the abnormal surfactant function caused by a variety of severe pulmonary injuries. Clinical trials are now underway using exogenous surfactant in an attempt to improve the outcome in adults.

3. Patchy atelectasis occurs due to surfactant abnormality in patients who have undergone cardiac surgery during which a pump oxygenator is used and the pulmonary circulation is interrupted.

COMPLIANCE

DEFINITION AND NORMAL VALUE

Distensibility or stretchability of lung and thorax is called compliance. Therefore, compliance is defined as the change in lung volume (ΔV) per unit change in transpulmonary pressure (ΔP), i.e.

$$C = \frac{\Delta V}{\Delta P}$$

Transpulmonary pressure is the difference in the pressure between the alveolar pressure and the pleural pressure.

Compliance is expressed under two headings:

(a) Total respiratory compliance or combined compliance of lungs and chest wall, i.e. lungs inside the thoracic cavity. Normal value of total respiratory compliance is 0.13 L/cm H_2O.

(b) Pulmonary compliance, i.e. of lungs only (lungs outside the chest wall). Normal value of compliance for the lungs alone is 0.22 L/cm H_2O.

MEASUREMENT OF TOTAL COMPLIANCE

Total respiratory compliance (combined compliance of chest wall and lungs) can be measured by the pressure–volume curve of respiratory system. Pressure–volume curve of the respiratory system can be obtained in living subjects by using a spirometer (described on page 365) as below:

Procedure

The subject is connected to the spirometer and asked to breathe air through mouth (with nostrils closed). The manometer attached with the spirometer measures airway pressure (Transpulmonary pressure). The subject inhales known volume of air from end-expiratory position and valve of the spirometer is then shut off to close the airway. The subject holds the breath and relaxes the respiratory muscles and change in the airway pressure is measured. The procedure is repeated with actively inhaling or exhaling different air volumes up to the maximum.

The change in the airway pressure is plotted against each air volume and curve obtained is called pressure–volume curve of respiratory system (Fig. 5.2-12).

Observations and inferences

1. Pressure–volume curve of the respiratory system is a sigmoid shape curve which is steepest at the middle

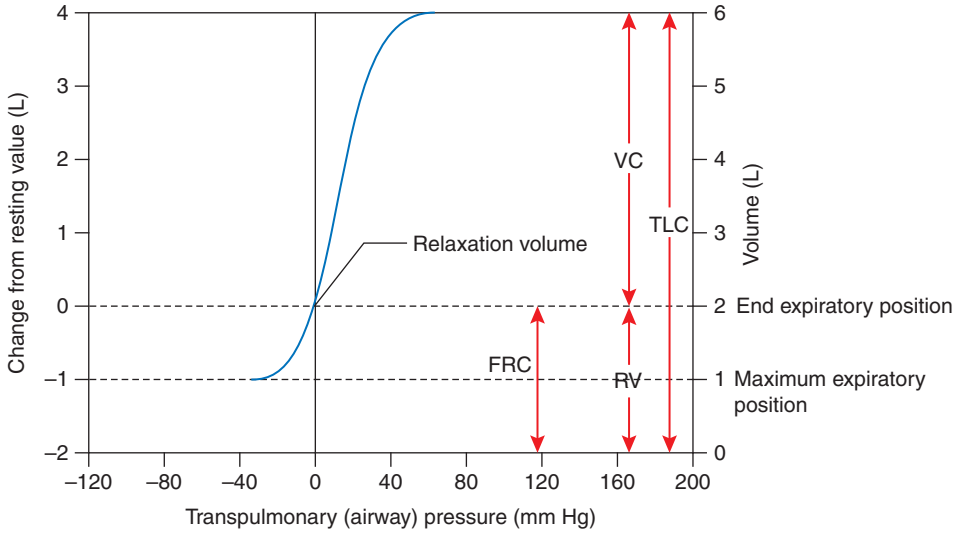


Fig. 5.2-12 Pressure-volume curve of respiratory system.

and almost flat at both ends (i.e. high and low lung volumes).

2. The volume at which air pressure is 0 mm Hg (atmospheric pressure) is called *relaxation volume* which is equal to functional residual capacity (FRC). At relaxation volume, the recoil of chest wall and recoil of lung balance each other.
3. Above relaxation volume, on increasing lung volumes there is increase in the airway pressure and at maximum inspiration pressure rises up to 30 mm Hg. On the other hand, decrease in the lung volumes below relaxation volume airway pressure decreases and at maximum expiration point it decreases up to –30 mm Hg.
4. The residual volume (RV) and total lung capacity (TLC) are limited by decrease compliance of the system as indicated by the reduced slop at both extreme volumes.

Factors affecting compliance of lungs and chest wall

Downward and to right shift of pressure–volume curve indicates decreased total respiratory compliance. The causes of decreased compliance are:

- Pulmonary congestion
- Interstitial pulmonary fibrosis
- Pulmonary oedema

Upward and to left shift of pressure–volume curve indicates increased total respiratory compliance. The causes of increased compliance are:

- Emphysema
- Old age

MEASUREMENT OF PULMONARY COMPLIANCE

- The compliance of lung alone can be measured by measuring the intrapleural pressure at different lung volumes.
- The relationship of lung volume to the intrapleural pressure is plotted on a graph, which gives the *inspiratory compliance curve* (Fig. 5.2-13).
- Similarly, expiration is also done in steps and *expiratory compliance* curve is obtained (Fig. 5.2-13).
- The inspiratory and expiratory compliance curves do not coincide but form a loop called *hysteresis loop.* Fig. 5.2-13 shows that the lung volume at any given pressure is greater during expiration than during inspiration. This is due to difference in the distensibility (stretchability) of the lungs between the inspiratory and expiratory phases. The volume midway between the hysteresis loop (dashed linear line in Fig. 5.2-13) gives the average compliance of the lungs. Had the lungs been a perfectly elastic structure like a spring, the pressure–volume relation line would have been linear (Hook's law).
- The curved lines obtained during inspiration and expiration are because of two resistances:
  - Viscous resistance due to non-elastic tissues in the lungs and
  - Airway resistance.
- Therefore, to calculate the compliance of the lung the point on the graph at the end of inspiration, i.e. when there is no air flow should be considered. At this point, there is no airway resistance and no viscous resistance. The compliance calculated from pressure and volume changes at this point indicates the compliance of lung alone due to elastic tissue only.

STATIC VERSUS SPECIFIC LUNG COMPLIANCE

Static compliance. The compliance measured as described above is the *static compliance.* The static compliance of any system is dependent on its size. Thus the lung compliance depends upon the amount of functional lung tissue. Thus,

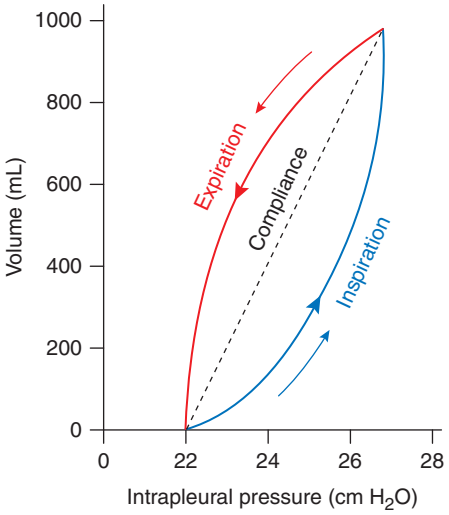


Fig. 5.2-13 Change in volume per unit change in intrapleural pressure during quiet respiration. The dotted line represents the pulmonary compliance.

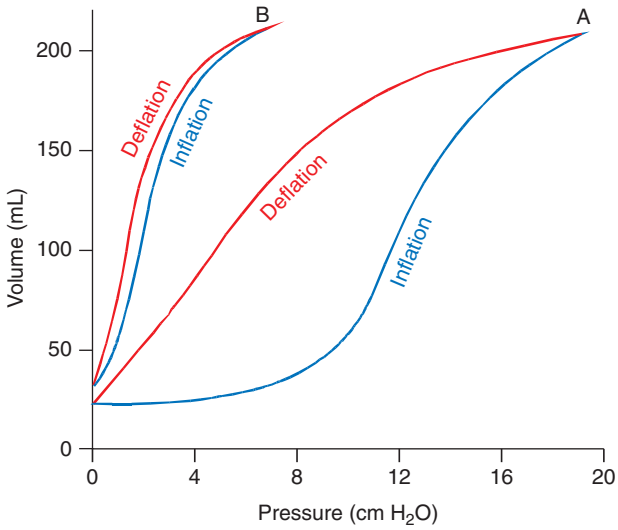


Fig. 5.2-14 Pressure-volume relationship of an isolated lung filled with air, A and with saline, B. The saline filled lung has far greater compliance due to absence of surface tension at water-air interface seen with air filled lung.

static compliance is not a good measure of absolute distensibility.

For example, if a patient's lung compliance is 0.22 L/cm H_2O, then both lungs together are able to expand 0.22 L for each cm H_2O change in the pleural pressure. Assuming equal compliance in both lungs, each lung will take up 0.11 L of gas. If the patient undergoes a pneumonectomy, the compliance measured will be only 0.11 L/cm H_2O, in spite of the fact that distensibility of the remaining lung is normal. Therefore, clinically term specific compliance is used.

Specific compliance is the compliance of the lung at relaxation volume (the point at the end of a tidal expiration), i.e. the functional residual capacity. The specific compliance is expressed per litre of FRC. It is a measure of the absolute distensibility of a structure. For instance, in the above cited example if the FRC is 2.2 L then:

- Specific compliance with both intact lungs will be:

$$= \frac{0.22}{2.2} = 0.1 \text{L/cm } H_2O$$

- Specific compliance after pneumonectomy of one lung will be

$$\frac{0.11}{1.1} = 0.1 \text{L/cm } H_2O$$

FACTORS AFFECTING LUNG COMPLIANCE

Lung compliance is inversely proportional to the lung elastance (elastic recoil force). Therefore, lung compliance is determined on the basis of the following elastic forces:

Elastic forces of the lung tissues due to elastic and collagen fibres contribute smaller amount of elasticity.

Elastic forces caused by surface tension within the alveoli account for about two thirds of total elastic forces in the lungs.

Alveolar surface tension plays a major role in generating elastance forces in the lungs and thus is very important factor affecting lung compliance. It can be demonstrated experimentally by recording the compliance of an isolated lung of an animal by first distending it with air and then with saline. Filling the lungs with saline theoretically eliminates the force of surface tension so that only the elastic forces of the lung tissue produce recoil. Figure 5.2-14 shows the results of such an experiment. Compared with air-filled lungs (Fig. 5.2-14A), the saline-filled lungs (Fig. 5.2-14B) how markedly reduced recoil forces, leading to the conclusion that the elastic forces caused by the alveloar surface tension play a major role in generating elastance forces in the lungs.

CHANGES IN THE LUNG COMPLIANCE

Decreased compliance of the lungs can be caused by lung diseases, (e.g. tuberculosis, silicosis) that produce scarring or fibrosis of the lungs, destruction of the functional lung tissue or both. Reduced compliance produces a condition termed *restrictive lung disease* (RLD). Patients with RLD must generate greater than the normal forces to expand the lungs.

Increased compliance is produced by the pathologic processes that occur in emphysema as well as from the aging process. Alveolar septa which provide some of the retractive forces in lungs are destroyed in both conditions, but emphysema causes a much more extensive loss of septa than the normal aging process.

WORK OF BREATHING

The contraction of respiratory muscles causes the expansion of thoracic cage and thereby causes expansion of lungs and fall in intra-alveolar pressure and allows atmospheric pressure to push air into the lungs during inspiration. Thus, to move the air into the lungs, the respiratory muscles have to do work to overcome the following *resistances.*

RESISTANCE TO BREATHING

Tissue resistance

It is the resistance offered by the tissues as they expand or contract. Tissue resistance comprises:

1. Elastic resistance. It is the sum of forces of elastic recoil exerted by the lung and the chest wall. The elastic recoil of the lungs is due to the presence of elastic fibres in the lungs and due to the alveolar surface tension.

2. Viscous resistance is the resistance offered by the non-elastic tissues in the lungs.

Airway resistance

It is the resistance caused by the friction of gas molecules between themselves and the walls of the airways. Factors affecting airway resistance are:

1. Rate of gas flow. Greater the rate of gas flow, greater is the resistance. Rate of gas flow is highest in the intermediate-sized bronchi as they have highest cross-sectional area; so the highest resistance to flow occurs in this part of tracheobronchial tree.

2. Airway radius. It is the most powerful determinant of resistance. Smaller the radius of airway, greater is the resistance. According to the *Poiseuille–Hagen formula* (see page 227),

$$\text{resistance} \propto \frac{1}{r^4}$$

In other words, if radius decreases by half (keeping the other factors constant), the resistance increases by 16 times.

Airway radius increases when lungs expand (during inspiration) and it decreases when lungs contract (during expiration). Therefore, airway resistance is high during expiration as compared to inspiration. Because of this reason in bronchial asthma patients (where bronchoconstriction develops) inspiration is possible but there is extreme difficulty in expiration.

3. Length of airway. It does not change much during respiration or in lung diseases and therefore, is not an important factor.

4. Type of air flow. The airway resistance is more in a turbulent flow (e.g. during rapid respiration) than in a laminar flow or streamline flow in quiet breathing.

Probability of turbulence in airflow depends on Reynolds (Re) number. Re value increases when there is increase in velocity of airflow, density of air and diameter of bronchi. For details see page 233.

IMPORTANT NOTE

Rise in atmospheric pressure (during deep sea diving) increases the density of the gas which can increase the airway resistance. Therefore, low-density gases like helium are used by the divers.

COMPONENTS OF WORK OF BREATHING

The total work done by the respiratory muscles during quiet breathing may be divided into following components:

- Work done to overcome elastic resistance (65%),
- Work done to overcome viscous resistance (7%) and
- Work done to overcome airway resistance (28%).

CALCULATION OF WORK OF BREATHING

The work of breathing, i.e. the work of inflating the lungs can be calculated by plotting the change in lung volume (ΔV) versus the change in the intrapleural pressure (ΔP) (Fig. 5.2-15). The total area represented by $\Delta P \times \Delta V$ in Fig. 5.2-15 is proportional to the work that must be performed by the respiratory muscles and to the O_2 utilized by them. The work done by the respiratory muscles can be calculated separately for inspiration and expiration.

Work done during inspiration

In Fig. 5.2-15, the area AXBCA denotes the total work done during normal inspiration. The work of inspiration can be divided into three fractions:

Compliance work refers to the work done by respiratory muscles to inflate the lungs against the *elastic resistance* of chest wall and lungs. It is represented by the triangular area

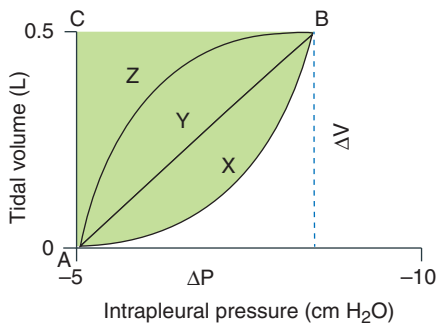


Fig. 5.2-15 Calculation of work done during breathing by plotting the change in volume (ΔV) against change in intrapleural pressure (ΔP) in normal quiet breathing.

AYBCA in Fig. 5.2-15. Thus most of the work done (65%) is used to overcome elastic resistance.

Non-elastic resistance work is done to overcome the non-elastic resistance. It includes the work done to overcome:

- Viscous resistance of lungs (7%) and
- Airway resistance (28%).

It is represented by area AXBYA in Fig. 5.2-15. Thus only a small amount (7%) of the work done is used to overcome the viscosity of the lungs and 28% of the work done is utilized to overcome the resistance of air flow through the respiratory passages.

Work done during expiration

Since in quiet breathing, expiration is a passive process so no work is done during expiration. The triangle AYBCA in Fig. 5.2-15 represents the stored elastic energy that is present at the end of inspiration. This stored energy can compress the alveolar gas and create expiratory flow. When the lungs are recoiling back some energy is required to overcome *non-elastic resistance,* i.e. the airway resistance plus viscous tissue resistance. This is represented by the area AYBZA which in normal quiet expiration falls within the triangle AYBCA (stored energy) and so no extra work is required to overcome this resistance.

Factors affecting total work of breathing

Total work of breathing during quiet respiration under normal circumstances ranges from 0.3 to 0.8 kg m/min. Normally, the work of breathing represents 2–3% of the resting O_2 consumption.

- If either the respiratory resistance increases or the compliance decreases, the respiratory work will increase and therefore the respiratory muscles will use more O_2 to overcome the added load.
- The work of breathing is also increased markedly during muscular exercise (physiological).

APPLIED ASPECTS

- In the presence of restrictive lung disease (RLD) the patient has to overcome significantly higher elastance forces for the normal tidal breathing. So, in RLD more work has to be performed by the respiratory muscles during inspiration to overcome the decreased compliance. Work done in patients with RLD can be minimized by breathing rapidly and shallow. The tidal volume is decreased but the increased respiratory rate ensures adequate ventilation of lungs.
- In patient with obstructive lung disease (OLD), the airway resistance is increased due to narrowing of passages. The patients must generate increased pressure gradients to produce adequate air flow. So, *during inspiration* more work by respiratory muscles is done to overcome the increased non-elastic resistance. In such circumstances active contraction of expiratory muscles is required to accomplish the task of expiration as elastic recoil is not sufficient. The resistance work can be minimized by breathing more slowly and deeply. This prolongs the time of expiration, which reduces the pressure gradient necessary to generate gas flow. The increased tidal volume compensates for the decreased respiratory rate so that normal alveolar ventilation is maintained.

Chapter 5.3

Pulmonary Circulation

FUNCTIONAL ANATOMY

The lungs have three circulations–pulmonary, bronchial and lymphatic.

PULMONARY CIRCULATION

Pulmonary trunk arises from the right ventricle and divides into right and left pulmonary arteries which convey the deoxygenated blood to the right and left lung, respectively. The blood circulates through a capillary plexus intimately related to the walls of alveoli and receives oxygen from the alveolar air. This blood which is now oxygenated is returned to the heart (left atrium) through four pulmonary veins.

BRONCHIAL CIRCULATION

The lungs also receive oxygenated blood like other tissues in the body through *bronchial arteries* (two left and one right), which are branches of the descending thoracic aorta. Bronchial blood flow amounts to about 1–2% of total cardiac output. The oxygenated blood in the bronchial arteries supplies the connective tissues, septa and large and small bronchi of the lungs. Because the bronchial blood empties into the pulmonary veins and bypasses the right heart, the bronchial circulation therefore, constitutes a *physiological shunt,* i.e. a channel that bypasses oxygenation in the lungs.

LYMPHATIC CIRCULATION

Lungs are richly supplied by lymphatics. Lymphatics are present in the walls of the terminal bronchioles and in all the supportive tissues of the lungs. Particulate matter entering the alveoli during inspiration is removed by way of lymphatic channels. Lymphatics also remove the plasma proteins leaking from the lung capillaries and thus help to prevent the pulmonary oedema. The deep lymphatic vessels follow the bronchi and first drain into the *pulmonary nodes* (in the substance of the lungs) and then into the *bronchopulmonary nodes.* The superficial lymph vessels lie near the surface of lungs and converge on to bronchopulmonary nodes. From the bronchopulmonary node, lymph drains into the *tracheobronchial nodes* and from there into the *bronchomediastinal trunk.*

CHARACTERISTIC FEATURES OF PULMONARY CIRCULATION

INTRODUCTION

- The pulmonary arteries and their branches are thin walled and distensible giving the pulmonary arterial tree a large compliance. The distensibility of pulmonary vessels makes the pulmonary circulation a *low-pressure, low-resistance* and *high-capacitance system.*

- The pulmonary capillaries surround the alveoli and are sandwiched between their walls, as a result each alveolus seems to be enclosed in a basket of capillaries (Fig. 5.3-1).

PRESSURES IN THE PULMONARY SYSTEM (FIG. 5.3-2)

Right ventricular pressure during each cardiac cycle reaches a peak value of 25 mm Hg in systole (if 120 mm Hg in the left ventricle) and falls to 0–1 mm in diastole (if 5 mm Hg in left ventricle).

Pulmonary artery pressures vis-a-vis aorta, respectively are:

- Systolic pressure 25 and 120 mm Hg,
- Diastolic pressure 8 and 80 mm Hg,
- Mean arterial pressure 15 and 100 mm Hg and
- Pulse pressure 17 and 40 mm Hg.

Left atrial pressure in major pulmonary veins averages about 5 mm Hg in the recumbent human being. Therefore, the pressure gradient in the pulmonary system (mean pulmonary artery pressure – mean pulmonary vein pressure) is 15−5=10 mm Hg (if 120 mm Hg in the systemic circulation).

In pulmonary capillary pressure, the mean values estimated through indirect means are about 10 mm Hg. Since this pressure is far below the colloid osmotic pressure (25 mm Hg), so a net suction force of 15 mm Hg tends to draw fluid from the alveolar interstitial space into the pulmonary capillaries *which keeps the alveoli dry.* However, if the pulmonary capillary hydrostatic pressure rises above 25 mm Hg, fluid can escape into the interstitial space leading to *pulmonary oedema.* This can happen during exercise, particularly at high altitude, in left heart failure, mitral stenosis and pulmonary fibrosis. This reduces the rate of gas exchange in the lungs. The resultant hypoxia may be fatal.

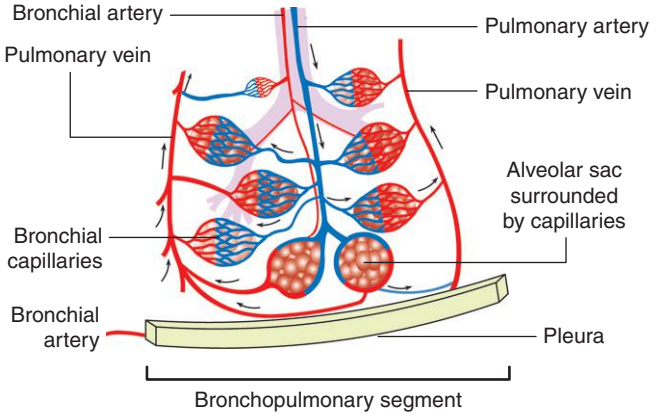


Fig. 5.3-1 Organization of pulmonary circulation.

PULMONARY BLOOD VOLUME

- *Pulmonary vessels contain* about 600 mL of blood at rest. Since the pulmonary vessels act as capacitance vessels their blood content can vary from 200 to 900 mL.
- *Pulmonary blood volume decreases* in the physiological conditions like standing and is shifted to systemic

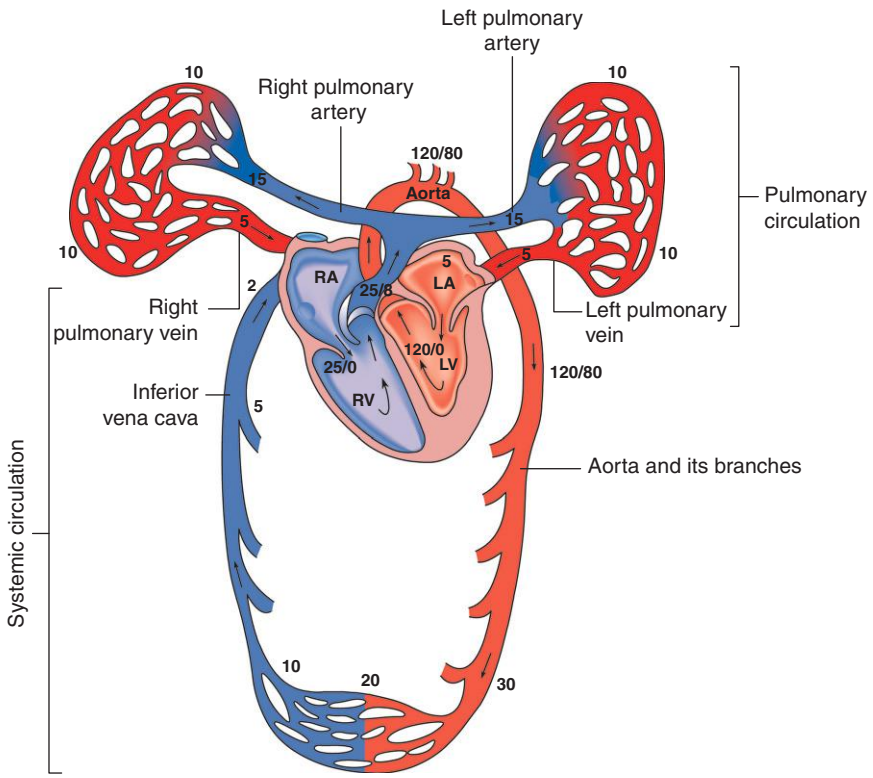


Fig. 5.3-2 Blood pressure (mm Hg) in pulmonary and systemic circulation.

circulation to compensate for the blood pooled in the leg veins due to gravity. In pathological conditions like haemorrhage also the transfer of blood from pulmonary vessels to systemic circulation can partly compensate for the blood loss. Thus, the pulmonary vessels act as a reservoir of blood.

PULMONARY BLOOD FLOW: REGIONAL DISTRIBUTION

Pulmonary blood flow

Pulmonary blood flow is nearly equal to cardiac output, since the right ventricle also pushes the same amount of blood simultaneously into the pulmonary circulation as the left ventricle pushes in the systemic.

Blood flow through the lungs depends upon the relationship between pulmonary arterial pressure (Pa), pulmonary venous pressure (Pv) and alveolar pressure (PA) as:

- The difference between the pulmonary arterial pressure (Pa) and venous pressure (Pv) is the driving pressure and
- The pulmonary capillary pressure must be above the alveolar pressure for the blood flow to continue.

Effect of gravity on regional pulmonary blood flow. In normal adults in supine position, the mean arterial pressure is same all over the lung and so all regions of the lung are uniformly perfused. However, in erect posture the gravity affects the regional distribution of blood through the lung by altering the pulmonary vasculature pressure due to hydrostatic pressure effect. The hydrostatic pressure adds or subtracts from the pressure levels in the supine position if levels are below or above the zero reference plane, respectively. The zero reference plane is at the level of the right atrium, which is approximately at the middle of lung in the region of hilum. Therefore, in standing posture the mean pulmonary arterial pressure in a 30 cm long lung will be:

- In the middle of lung (zero reference level): 15 mm Hg.
- At the apex of lung it will be less by 15 cm H_2O or 11 mm Hg, i.e. about: 4 mm Hg.
- But at the base of lung it will be more by 11 mm Hg, i.e. about: 26 mm Hg.

Perfusion zones of the lung

Depending on the relationship between alveolar pressure (PA), pulmonary arterial pressure (Pa) and pulmonary venous pressure (Pv), the lung can be divided into three zones in the erect posture (Fig. 5.3-3):

Zone 1 refers to the area of zero flow, i.e. any region of the lung that does not receive blood flow. Zone 1 *does not exist in normal lungs.*

Zone 1 is present under following abnormal conditions:

- When the pulmonary arterial pressure is too low as in hypovolaemic shock, pulmonary embolism or
- When the alveolar pressure is too high to allow flow as in severe obstructive lung disorder,

Zone 2 refers to the region of lung that has intermittent blood flow. This occurs during systole when the pulmonary arterial pressure exceeds the alveolar pressure which exceeds the pulmonary venous pressure Pa > PA > Pv, but not during diastole when the arterial pressure is less than the alveolar pressure.

Zone 3 refers to the region of high continuous blood flow where the capillary pressure remains greater than the alveolar

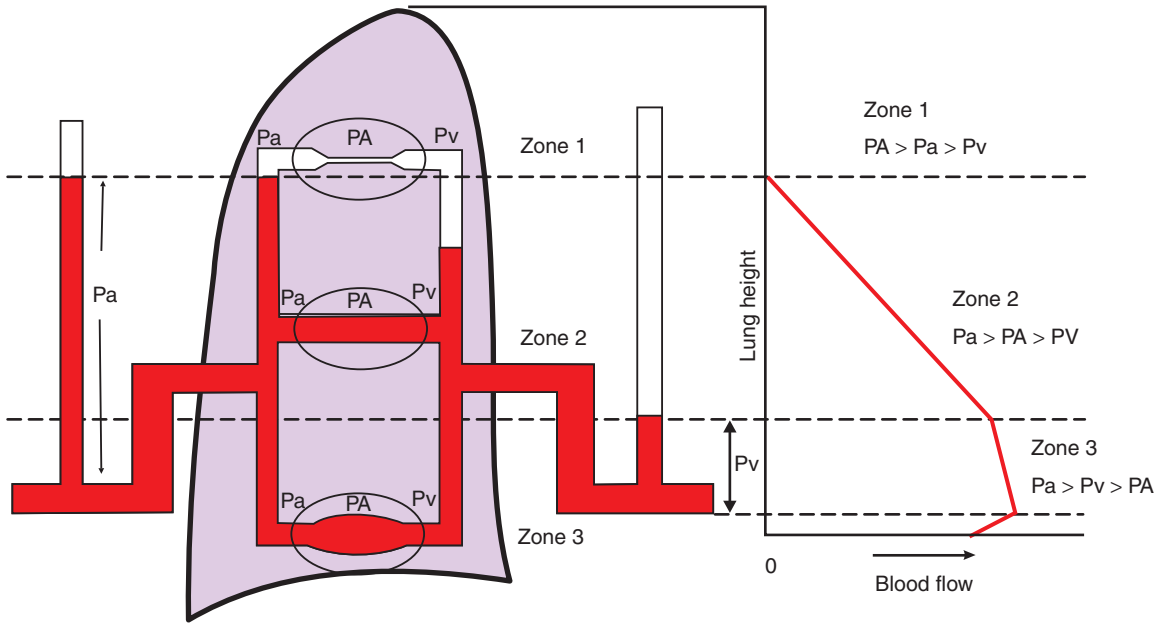


Fig. 5.3-3 Effect of gravity on regional distribution of pulmonary blood flow in standing posture.

pressure during the entire cardiac cycle. So, Zone 3 occurs when the pulmonary artery pressure exceeds the pulmonary venous pressure which exceeds the alveolar pressure (Pa > Pv > PA). This condition occurs near the bottom of the lung.

FUNCTIONS OF PULMONARY CIRCULATION

Respiratory gas exchange is the major function of pulmonary circulation. This function has been discussed in Chapter 5.4 on 'Pulmonary Diffusion'.

Other functions of pulmonary circulation which have been described on page 296 are:

- Reservoir for left ventricle,
- Filter for removal of emboli and other particles from blood,
- Removal of fluid from alveoli,
- Role in absorption of drugs and
- Synthesis of angiotensin converting enzyme.

REGULATION OF PULMONARY BLOOD FLOW

NEURAL CONTROL

Efferent sympathetic vasoconstrictor nerves richly innervate the pulmonary blood vessels. But these nerves have *no resting discharge and tone,* which mean they can only show an increase in activity when stimulated. These participate in the vasomotor reflexes, e.g.

- *Baroreceptor stimulation* produces reflex dilatation of pulmonary vessels, while,
- *Chemoreceptor stimulation* causes pulmonary vasoconstriction.

However, the effect of vasodilatation and vasoconstriction is more on the capacity rather than the resistance of pulmonary vessels. The vasoconstriction produced in the lungs due to sympathetic discharge is considerable and results in transfer of pulmonary blood in the systemic circulation, e.g. during acute haemorrhage.

Afferent control through vagus is mediated through following receptors:

- *Pulmonary baroreceptors (see page 259)*
- *Pulmonary volume receptors (see page 258)*
- *J receptors (see page 340)*

CHEMICAL CONTROL

1. Local hypoxia is responsible for most significant alterations in pulmonary blood flow by producing *vasoconstriction.*

- *In lungs low pO_2* in a region means that the alveoli in that region are not well ventilated. Therefore, blood flow through that region is waste. Since it will not get adequately oxygenated thus, the local low pO_2 induced vasoconstriction in the lungs is a phenomenon that diverts pulmonary blood flow from the alveoli that are poorly ventilated to better ventilated regions so that blood can be properly oxygenated.

2. Hypercapnia and acidosis also produce vasoconstriction. The effects of pCO_2 and acidosis on pulmonary vessels are just opposite to those in the systemic vessels where these stimuli produce vasodilation. The functional significance of this response is same as in the case of local hypoxia.

3. Chronic hypoxia, as occurs in high-altitude dwellers, is associated with a marked increase in pulmonary arterial pressure *(pulmonary hypertension),* which imposes a heavy afterload on the right ventricle that results in right ventricular hypertrophy, right heart failure and pulmonary oedema. This is the reason that:

- Thick pulmonary precapillary vessels develop in high-altitude dwellers and
- Children born and raised at high-altitude show pulmonary hypertension.

Chapter 5.4

Pulmonary Diffusion

INTRODUCTION

Pulmonary diffusion, i.e. transfer of gases from alveoli to capillary blood across the respiratory membrane.

To understand the *intricacies* of diffusion of gases across the respiratory membrane, it is essential to have knowledge about the following related aspects and concepts, most of which form the contents of this chapter:

- *Pulmonary perfusion,* i.e. *pulmonary blood flow,* has been discussed in Chapter 5.3.
- *Physics of gas diffusion and gas partial pressures.* The rate at which the respiratory gases diffuse across the respiratory membrane requires an understanding of the physics of diffusion and gas exchange.
- *Alveolar ventilation,* i.e. the rate at which new air reaches the gas exchange area of the lungs. Alveolar ventilation governs the process of pulmonary diffusion in its own way.
- *Alveolar ventilation–perfusion ratio.* It is the ratio of alveolar ventilation and pulmonary blood flow. This is highly quantitative concept which was developed to help in understanding respiratory exchange when there is an imbalance between alveolar ventilation and alveolar blood flow.
- *Diffusion of gases through the respiratory membrane* is the main process which transfers gases from the alveoli into the capillaries. Various important aspects of this process have also been discussed in this chapter.

PHYSICS OF GAS DIFFUSION AND GAS PARTIAL PRESSURES

It is worthwhile to recapitulate some of the basic principles and laws governing the behaviour of gases for a better understanding of the process of diffusion between alveolar air and pulmonary capillary blood. Some of the important aspects concerning physics of gas diffusion and gas partial pressures are discussed below.

GAS PRESSURE

The gas molecules have a kinetic energy so they are in a continuous random motion. These molecules bounce against

each other and/or against the walls of container and exert a pressure. The gas pressure (P) exerted depends upon the following factors:

Concentration of molecules (n). The pressure of a gas is directly proportional to its concentration, i.e.

$$P \propto n \quad \text{(i)}$$

Volume (V). Unlike liquids, gases expand to fill the volume available to them. Therefore, at a constant temperature, the pressure (P) of a given mass of gas is inversely proportional to its volume (V), i.e.

$$P \propto \frac{1}{V} \quad \text{(ii)}$$

This is called *Boyle's law* of gases.

Absolute temperature (T). According to the *Charles' law,* at a constant pressure the volume of gas is directly proportional to its absolute temperature, i.e. V α T, and as mentioned above according to the Boyle's law.

$$P \propto \frac{1}{V} \quad \text{(iii)}$$

Therefore, it can be derived that

$$P \propto T$$

From equation (i)–(iii) it can be derived that

$$p = \frac{nRT}{V},$$

where

P = Pressure of gas,
n = Number of molecules of gas,
T = Absolute temperature,
V = Volume of gas and
R = Gas constant.

PARTIAL PRESSURE

According to the *Dalton's law* of partial pressure, the total pressure exerted by a mixture of gases is equal to the sum of the partial pressure of all gases present in the mixture. Thus, the partial pressure (p) refers to the pressure exerted by any one gas present in a mixture of gases. It is equal to the total pressure exerted by the mixture of gases times the fraction of the total amount of mixture of gases it represents. Hence, the partial pressure (p) of a gas can be calculated by multiplying its fractional concentration by the total pressure. For example, environmental air (which has atmospheric pressure (at sea level) of about 760 mm Hg) is a mixture of 21% oxygen (O_2) and 79% nitrogen (N_2). Therefore, the partial pressure (p) of O_2 and N_2, respectively will be:

$$pO_2 = 760 \times \frac{21}{100} = 160 \text{ mm Hg and}$$

$$pN_2 = 760 \times \frac{79}{100} = 600 \text{ mm Hg.}$$

Table 5.4-1 Solubility coefficient of important gases to respiratory physiology

Temperature (°C)	Solubility coefficient (mL gas/mL saline/mm Hg gas tension)			
	O_2	N_2	CO_2	CO
0	0.049	0.024	1.71	0.035
20	0.032	0.016	0.90	0.023
37	0.024	0.012	0.58	0.019

PARTIAL PRESSURE OF GASES IN WATER AND TISSUES

It is important to have knowledge about partial pressure of gases in water and tissues because the respiratory gases to cross the respiratory membrane must first dissolve in the tissues and then diffuse into the plasma of pulmonary capillaries. The pressure (p) of a gas in a solution is determined not only by its concentration, but also by its solubility coefficient. According to the *Henry's law,* when temperature is kept constant, the content of gas (n) dissolved in any solution is directly proportional to the partial pressure of a gas, i.e.

$$n = pc,$$

where

n = concentration (amount) of a gas in a solution,
p = partial pressure of gas and
c = solubility coefficient of the gas.

Solubility coefficient is a constant for each gas that equates gas content and partial pressure. Increasing the temperature reduces the solubility coefficient of gases (Table 5.4-1). It is important to note that *solubility coefficient of carbon dioxide (CO_2) is about 24 times more and that of nitrogen (N_2) is one half that of oxygen (O_2).* Solubility coefficient of gases (important to respiratory physiology) is given in Table 5.4-1.

WATER VAPOUR PRESSURE

The atmospheric air entering the respiratory passages during inspiration is humidified by the water vapours from the conducting passages. By the time, the atmospheric air reaches the alveoli it is saturated with water vapours. Thus, in the alveolar air, besides O_2 and N_2, water vapours also exert its partial pressure. Vapour pressure of water is dependent upon its temperature. At body temperature (37°C), the vapour pressure of water in alveolar air is 47 mm Hg.

ALVEOLAR VENTILATION

Alveolar ventilation is the volume of the fresh air which reaches the gas exchange area of the lung each minute.

During inspiration, some of the air inhaled never reaches the gas exchange areas but instead fills the non-gas exchange areas (conducting zone) of the respiratory tract called the *dead space,* which is equal to about 150 mL.

During expiration, out of 500 mL of tidal volume 150 mL of the alveolar expired air remains in the conducting passages. Therefore, of 500 mL air entering the lungs only 350 mL/breath is the fresh air which contributes to the alveolar ventilation. Thus alveolar ventilation can be calculated as:

Alveolar ventilation (VA) = Respiratory rate × (tidal volume – Dead space volume) with a normal tidal volume of 500 mL, a normal dead space of 150 mL and a respiratory rate of 12 breaths/min, alveolar ventilation equals 12 × (500 – 350), or 4200 mL/min.

PHYSIOLOGICAL SIGNIFICANCE OF ALVEOLAR VENTILATION

The physiological significance of alveolar ventilation can be understood by comparing the alveolar ventilation of two subjects with following parameters:

Subject A, having normal breathing with a tidal volume of 500 mL and respiratory rate of 12 breaths/min will have:

- Pulmonary ventilation = 12 × 500 = 6 L/min and
- Alveolar ventilation = 12 × (500 – 150) = 4.2 L/min.

Subject B, having rapid shallow breathing with a tidal volume of 200 mL and respiratory rate 30/min will have:

- Pulmonary ventilation = 30 × 200 = 6 L/min and
- Alveolar ventilation = 30 × (200 – 150) = 1.5 L/min.

On comparison, we see that both the subjects A and B have similar amounts of pulmonary ventilation (6 L/min), but the subject B has the alveolar ventilation (1.5 L/min) which is much less than that of subject A (4.2 L/min). Consequently, the subject B is likely to suffer from hypoxia and hypercapnia.

DEAD SPACE AIR

Dead space air is the portion of minute ventilation that does not take part in the exchange of gases. There are three types of dead spaces (Fig. 5.4-1):

1. Anatomical dead space refers to the volume of air present in the conducting zone of the respiratory passage, i.e. from nose to terminal bronchioles. As mentioned earlier, the anatomical dead space contains approximately 150 mL of air.

2. Alveolar dead space air refers to the volume of air present in those alveoli which do not take part in gas exchange. Normally, all the alveoli take part in the gas exchange, but in some lung diseases, some alveoli do not take part in the gas exchange.

3. Physiological dead space refers to the total dead space which includes both the anatomical and alveolar dead spaces. In a normal healthy person, physiological dead space nearly equals the anatomical dead space. However, in certain respiratory disorders with many non-functioning alveoli, the physiological dead space may be as much as ten times the anatomical dead space.

Measurement of anatomical dead space

Single breath oxygen technique. This technique is also called Single breath Nitrogen washout test. In this test, nitrogen contents in the expired air are used as an indicator for determining dead space.

Procedure. The subject is asked to take a deep breath of pure oxygen (100% O_2). Then steadily exhales into the nitrometer, which continuously measures N_2 contents in the expired air. The anatomical dead space can be measured by the analysis of single breath nitrogen curve (Fig. 5.4-2). This curve has four phases labelled by roman letters (I, II, III and IV).

The curve shows that:

- During inspiration, N_2 contents are nil (zero %) as subject has inspired pure O_2.
- During initial phase of expiration (Phase I), N_2 contents are nil (zero %) as the expired air is from the dead space (which is filled with pure O_2).
- Subsequently (Phase II), there is rise in N_2 contents in expired air because exhaled air contains mixture of dead space air and alveolar air.
- In phase III, the N_2 contents reach to a plateau (60%) and phase III of single breath nitrogen curve ends at closing volume (CV) and followed by the phase IV.
- Phase IV. In this phase, N_2 contents of the expired air are further increased.

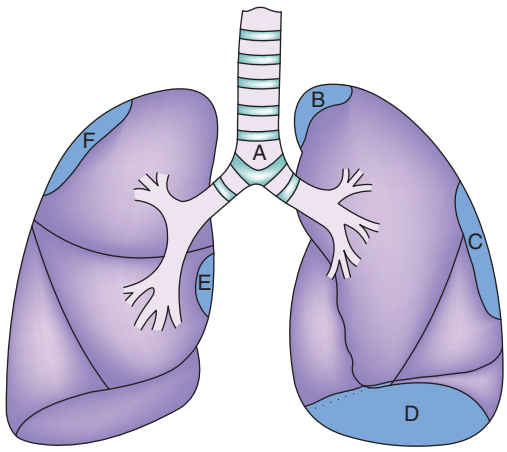


Fig. 5.4-1 Dead space: anatomical (A), alveolar (B+C+D+E+F) and physiological (A+B+C+D+E+F).

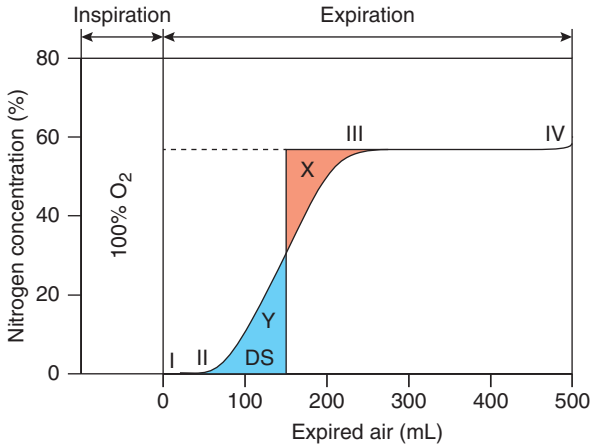


Fig. 5.4-2 Single breath nitrogen curve for measuring dead space air. The changes in N_2 concentration of expired air during expiration are indicated by various phases (I–IV). (DS = Dead space.)

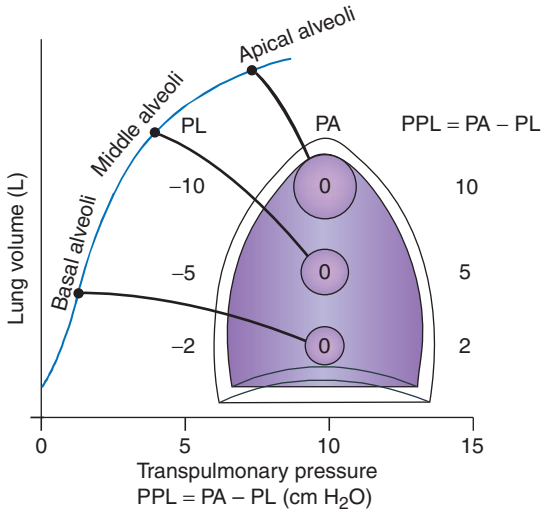


Fig. 5.4-3 Correlation of transpulmonary pressure (PPL = PA – PL) and lung inflation in different regions during erect posture. (PA = Alveolar pressure; PL = intrapleural pressure; PPL = transpulmonary pressure.)

- The volume of anatomical dead space is measured by placing a vertical line on the record from mid-portion of phase II of expiration (red area X = blue area Y).

Measurement of physiological dead space

Bohr's equation is used to measure the physiological (total) dead space by determining CO_2 tensions in the expired and alveolar gas. Bohr's equation is based on the fact that inspired air contains negligible quantity of CO_2 (almost zero). Therefore, all the CO_2 in expired air is derived from the functional alveoli. The equation is

$(VT - VD) \times pACO_2$ + ($V_D \times pCO_2$ in inspired air, i.e. zero) $= VT \times pECO_2$

$$\text{or} \quad VT - VD = \frac{VT \times pECO_2}{pACO_2}$$

$$\text{or} \quad VD = VT - \{(VT \times pECO_2) / pACO_2\},$$

where

VD = Physiological (total) dead space air,
VT = Tidal volume,
$pECO_2$ = Carbon dioxide tension in mixed expired air and is represented by average pCO_2 and of the total expired air and
$pACO_2$ = Carbon dioxide tension in alveolar air and is theoretically represented by the pCO_2 of end-tidal samples of expired gases.

For example, if

$pECO_2$ = 28 mm Hg
$pACO_2$ = 40 mm Hg
VT = 500 mL

then, VD = 150 mL

EFFECT OF GRAVITY ON ALVEOLAR VENTILATION

Alveolar ventilation is more or less evenly distributed in the supine position because hydrostatic effect on the intrapleural pressure is reduced. However, in a vertical lung the alveolar ventilation is unevenly distributed because of variation in compliance in different regions of the lungs as explained (Fig. 5.4-3):

- The alveolar pressure is zero throughout the lung under static conditions.
- The intrapleural pressure shows a gradient of about 8 cm H_2O between apex (–10 cm H_2O) and base (–2 cm H_2O).
- So, transpulmonary pressure (intrapleural pressure – alveolar pressure) also varies from –10 cm H_2O at apex to –2 cm H_2O at the base.
- Consequently, the lung compliance (change in lung volume per unit change in transpulmonary pressure) also shows corresponding gradient between apex and base.
- Because of more negative intrapleural pressure at apex (–10 cm H_2O), the apical alveoli are larger but poorly ventilated. While the basal alveoli because of less negative (–2 cm H_2O) intrapleural pressure is smaller but better ventilated.
- There is a linear reduction in the regional alveolar ventilation from base to apex in an erect position (Fig. 5.4-2).

ALVEOLAR VENTILATION–PERFUSION RATIO

Alveolar ventilation–perfusion ratio (VA/Q) is the ratio of alveolar ventilation per minute to quantity of blood flow to alveoli per minute. Normally, alveolar ventilation (VA) is

4.2–5.0 L/min and the pulmonary blood flow (equal to cardiac output) is approximately 5 L/min. So, the normal VA/Q is about 0.84–0.9. At this ratio maximum oxygenation occurs.

Effect of gravity on VA/Q

- Because of the effect of gravity, the basal alveoli are overperfused and apical alveoli are under perfused. There is almost of a linear reduction in the blood flow from the base to apex (Fig. 5.4-4).
- The alveolar ventilation also reduces linearly from the base to apex (Fig. 5.4-4) and thus the basal alveoli are overventilated and apical alveoli are under ventilated.
- However, gravity affects perfusion much more than it affects ventilation. Hence, as shown in Fig. 5.4-5, the apical alveoli are more underperfused than underventilated. Because of this relationship, the VA/Q is more than one.

APPLIED ASPECTS

Because of high VA/Q ratio, the apical alveolar air has low pCO_2 and high pO_2. Since high alveolar pO_2 provides favourable environment for the growth of *Mycobacterium tuberculosis* so the apices of lungs are more predisposed to tuberculosis.

Effects of alterations in VA/Q ratio

1. Normal VA/Q ratio implies that there is both normal alveolar ventilation and normal alveolar perfusion. The exchange of gases is optimal and the alveolar pO_2 is about 104 mm Hg and pCO_2 is about 40 mm Hg.

2. Increased VA/Q ratio means that the alveolar ventilation is more than the perfusion. As a result, the whole of the alveolar air is not utilized for gaseous exchange. The extra air in the alveoli which goes waste forms the so-called alveolar dead space air. There will also be a change in the composition of alveolar air (Fig. 5.4-5). *When VA/Q ratio increases to infinity,* i.e. when alveolar perfusion becomes zero, no exchange of gases can occur. Under such circumstances, the composition of alveolar air becomes equal to the humidified inspired air, which has pO_2 of 149 mm Hg and a pCO_2 of 0 mm Hg (Fig. 5.4-5).

3. Decreased VA/Q ratio occurs when the rate of blood *flow is more than the rate* of alveolar ventilation. Since the alveolar ventilation is not enough to provide oxygen, a fraction of venous blood passes through the pulmonary capillaries without becoming oxygenated. This fraction is called *shunted blood.* This shunted blood along with the additional deoxygenated blood from the bronchial veins to the pulmonary vein (about 2% of cardiac output) forms the so-called *physiological shunt.* The greater the physiological shunt, the greater is the amount of blood that fails to be oxygenated as it passes through the lungs. *When VA/Q becomes zero,* there is no

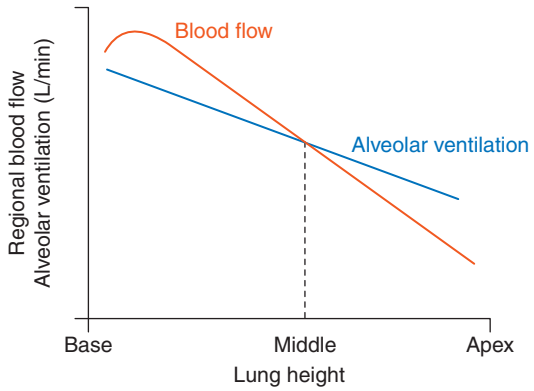


Fig. 5.4-4 Distribution of alveolar ventilation and pulmonary blood flow in different regions of lung during standing.

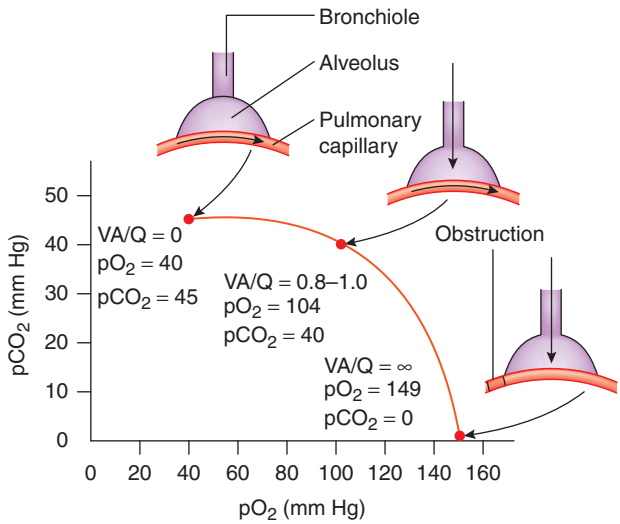


Fig. 5.4-5 Relationship of alveolar ventilation–perfusion ratio (VA/Q) with alveolar air pO_2 and pCO_2.

alveolar ventilation, so that the air in the alveolus comes to equilibrium with O_2 and CO_2 in the venous blood flowing through the pulmonary capillaries. So, alveolar air will have a pO_2 of 40 mm Hg and pCO_2 of 45 mm Hg.

Causes of alteration in VA/Q ratio

Obviously, the factors altering the alveolar ventilation or/and pulmonary perfusion will alter the VA/Q ratio.

Causes of uneven alveolar ventilation include:
- Bronchial asthma,
- Emphysema,
- Pulmonary fibrosis,
- Pneumothorax and
- Congestive heart failure.

Causes of uneven pulmonary perfusion are:
- Anatomical shunts, e.g. Fallot's tetralogy,
- Pulmonary embolism,

- Regional decrease in pulmonary vascular bed in emphysema and
- Increased pulmonary resistance in conditions like pulmonary fibrosis, pneumothorax and congestive heart failure.

ALVEOLAR AIR

Volume of air which is available for the exchange of gases in the alveoli per breath is called alveolar air, which is equivalent to tidal volume minus dead space, i.e. (500 – 150) or 350 mL.

COMPOSITION OF ALVEOLAR AIR

Composition of alveolar air can be studied by an *alveolar air sampling* that involves analysis of the last few millilitres of air that issues from the lungs during expiration. Alveolar air composition is considerably different than that of atmospheric air (Table 5.4-2) because of the following reasons:

- Water vapours dilute the other gases in the inspired air.
- Alveolar air is renewed very slowly by the atmospheric air.
- Oxygen is constantly being absorbed from the alveolar air.
- Carbon dioxide is constantly diffusing from the pulmonary blood to the alveoli.

COMPOSITION OF EXPIRED AIR

As shown in Table 5.4-2, the composition of expired air is different than that of the alveolar air. This is because of the fact that the expired air is a combination of dead space air and alveolar air. The expired air can be split into three portions:

Table 5.4-2 Composition and partial pressure of gases in atmospheric air, humidified air, alveolar air and expired air

Gas	Partial pressure (mm Hg) and concentration (percentage) of various gases			
	Atmospheric air	Humidified air	Alveolar air	Expired air
N_2	597.0 (78.62%)	563.4 (74.09%)	569.0 (74.9%)	566.0 (74.5%)
O_2	159.0 (20.84%)	149.3 (19.67%)	104.0 (13.6%)	120.0 (15.7%)
CO_2	0.3 (0.04%)	0.3 (0.04%)	40.0 (5.3%)	27.0 (3.6%)
H_2O	3.7 (0.5%)	47.0 (6.20%)	47.0 (6.2%)	47.0 (6.2%)
Total	**760.0 100%**	**760.0 100%**	**760.0 100%**	**760.0 100%**

- *First portion* of the expired air represents the dead space air and its composition is similar to typical humidified air.
- *Middle portion* of the expired air is a mixture of more and more alveolar air with the dead space air until the dead space air is washed out.
- *Last portion* of the expired air represents the alveolar air and so is used to study the alveolar air composition *(alveolar air sampling).*

DIFFUSION OF GASES THROUGH THE RESPIRATORY MEMBRANE

RESPIRATORY UNIT AND RESPIRATORY MEMBRANE

Each respiratory unit is composed of a respiratory bronchiole, alveolar ducts, atria and alveoli. There are about 300 million respiratory units in the two lungs. Gas exchange occurs through the membranes of all the structures forming a respiratory unit, not merely in the alveoli themselves.

Respiratory membrane or pulmonary membrane or the alveolocapillary membrane is the name given to the tissues which separate the capillary blood from the alveolar air. The exchange of gases between the capillary blood and alveolar air requires diffusion through this membrane.

Structure of respiratory membrane. It consists of the following layers (Fig. 5.4-6):

- Layer of pulmonary surfactant and fluid lining the alveolus (I).

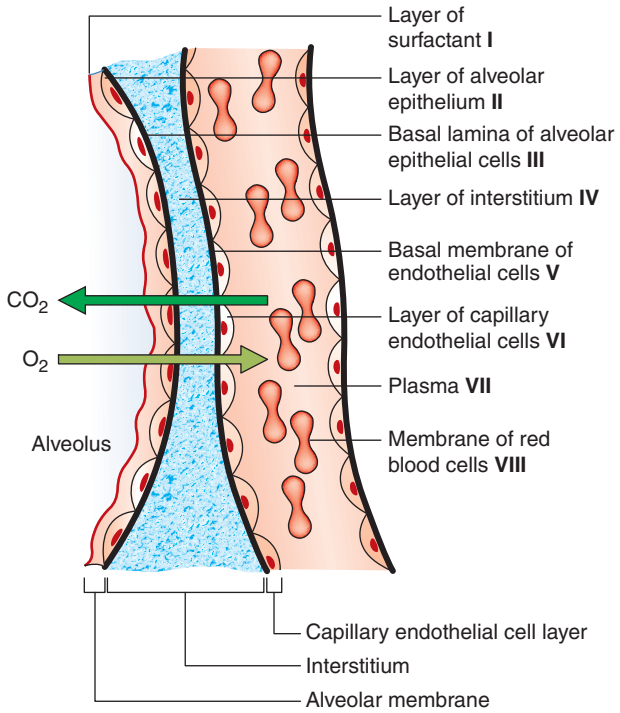


Fig. 5.4-6 Layers of respiratory membrane.

- Layer of alveolar epithelial cells (II).
- Basal lamina of the alveolar epithelial cells (III).
- A very thin interstitial space between the epithelial and endothelial cells (IV).
- Basement membrane of capillary endothelial cells and
- Layer of capillary endothelial cells (VI).

Characteristic features of respiratory membrane which optimize for the gas exchange are:

- *Thickness* of the respiratory membrane, despite the large number of layers forming, it averages about 0.6 µm.
- *Surface area* of the total respiratory membrane is about 70 m^2 in the normal adult.

FACTORS AFFECTING DIFFUSION ACROSS RESPIRATORY MEMBRANE

The diffusion of gases across the respiratory membrane is affected by following factors:

1. Thickness of respiratory membrane. The rate of diffusion through the membrane is inversely proportional to the membrane thickness (diffusion distance, denoted by d), i.e.

$$\text{V gas (volume of gas diffused)} \propto \frac{1}{d} \qquad \text{(i)}$$

Any factor which increases the thickness will therefore significantly decrease the gaseous exchange. Examples are:

- Pulmonary oedema, i.e. collection of fluid in the interstitial space and alveoli,
- Pulmonary fibrosis occurring in certain lung diseases increases the thickness of respiratory membrane.

2. Surface area of respiratory membrane. Normally, the total surface area of the respiratory membrane is about 70 m^2. The rate of diffusion is directly proportional to the surface area (A), i.e. with the decrease in total surface area, the rate of diffusion of gases decreases,

$$\text{V gas} \propto \text{A} \qquad \text{(ii)}$$

Some *causes of decrease in the surface area* are:

- Pulmonectomy, i.e. removal of one complete diseased lung reduces surface area to half the normal.
- In emphysema, many of the alveoli coalesce with dissolution of alveolar walls; this often causes the total surface to decrease by as much as five fold.

3. Diffusion coefficient. The rate of diffusion is directly proportional to the diffusion coefficient (D) of the gas, i.e.

$$\text{V gas} \propto \text{D} \qquad \text{(iii)}$$

- The diffusion coefficient of CO_2 is about 20 times that of O_2 through water (and therefore through fluid of respiratory membrane). Therefore, CO_2 diffuses much more easily through the respiratory membrane.
- The diffusion coefficient of a gas is a function of molecular weight (diffusion is inversely related to the square root of the molecular weight), solubility in a particular solvent and absolute temperature.

$$\text{D} = \frac{\text{Solubility of the gas}}{\sqrt{\text{Molecular weight of the gas}}}$$

4. Pressure gradient across respiratory membrane. The rate of diffusion across the respiratory membrane is directly proportional to the pressure difference between the partial pressure of a gas in alveoli (pA) and in pulmonary capillary (pC), i.e.

$$\text{V gas} \propto (\text{pC} - \text{pA}) \qquad \text{(iv)}$$

From the equations (i)–(iv) it can be derived that

$$\text{V gas} = (\text{pC} - \text{pA}) \frac{\text{D.A.}}{d} \qquad \text{(iv)}$$

This is called *Fick's law of diffusion.* Thus, it can be concluded that the rate of pulmonary gas diffusion, i.e. the volume of gas that crosses the respiratory membrane per minute is determined by several factors as defined by Fick's law of diffusion.

DIFFUSION AND EQUILIBRATION OF GASES ACROSS THE RESPIRATORY MEMBRANE

Diffusion of O_2. The normal alveolar pO_2 is 104 mm Hg, whereas the blood entering the pulmonary capillary normally has pO_2 of 40 mm Hg. Pressure gradient therefore is 64 mm Hg in the beginning. After dissolving in the respiratory membrane, the O_2 molecules diffuse into the blood. As O_2 diffuses from the alveoli to the blood, the pO_2 of blood becomes the same as in alveolar air (104 mm Hg), the gradient becomes zero and no diffusion occurs (Fig. 5.4-7). By the time blood passes to one-third of distance in capillary, the pO_2 of blood equals that of alveoli. This means that the pressure gradient is there only for one-third of blood flow in the capillary. This time integrated average pressure gradient is about 11 mm Hg.

Equilibration time. The blood remains for about 0.75 s in the capillary (*transit time).* As mentioned above, normally, enough O_2 diffuses across the respiratory membrane so that the blood pO_2 and alveolar pO_2 equalize in one-third of the transit time, i.e. about 0.25 s; then no further gas transfer normally takes place for the rest 0.50 s of transit time. This time provides a *safety margin* that ensures adequate O_2 uptake during the periods of stress (e.g. exercise, exposure to high altitude) or impaired diffusion. When the normal diffusing capacity of the lung for O_2 is diminished, the equilibrium time is prolonged or never reached.

Diffusion of CO_2 occurs from the blood to the alveoli because pCO_2 is higher in blood than in the alveolar air. The average pCO_2 in the pulmonary capillary blood is 46 mm Hg,

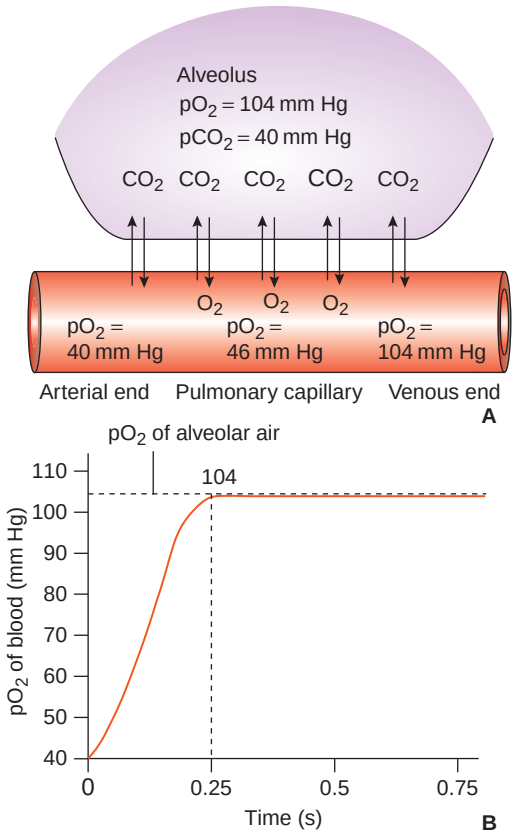


Fig. 5.4-7 A, diffusion of oxygen across respiratory membrane and B, leading to progressive increases in capillary blood pO_2.

as opposed to 40 mm Hg in the alveoli. Therefore, pressure gradient in the beginning is 6 mm Hg and time-integrated pressure gradient calculated for CO_2 (in a manner similar to O_2) across the respiratory membrane is only 1 mm Hg. Although the pressure gradient for CO_2 is only one-tenth of the O_2 diffusion gradient, CO_2 diffuses almost 20 times more rapidly than O_2 because of higher diffusion coefficient.

DIFFUSION CAPACITY OF LUNG

Diffusion capacity (DL) of the lung is a quantitative expression of the ability of the respiratory membrane to exchange a gas between the alveoli and the pulmonary blood. It is defined as the volume of gas (V gas) that diffuses through the respiratory membrane of lung each minute for a pressure gradient of 1 mm Hg.

Diffusion capacity of lungs for O_2

- *At rest,* the diffusing capacity of lungs for O_2 is about 20–25 mL/min/mm Hg. As the mean oxygen pressure gradient across the respiratory membrane is about 11 mm Hg, so at rest about 250 mL of O_2 diffuses through the lungs per minute.
- *During exercise,* the diffusing capacity of lungs for O_2 is increased. It may reach up to 65 mL/min/mm Hg during strenuous exercise (three times of diffusing capacity for O_2 at rest).

Diffusion capacity of lungs for CO_2

has never been measured because carbon dioxide diffuses across the respiratory membrane so rapidly that the difference between the average pCO_2 of the capillary blood and pCO_2 of alveolar air is only 1 mm Hg. Such a small difference cannot be detected by any available technique. However, from the available knowledge about diffusion coefficient of CO_2, the diffusion capacity of lungs for CO_2 is estimated.

- *At rest,* the diffusing capacity of lungs for CO_2 is about 20 times that for O_2, i.e. about 400 mL–500 mL/min/mm Hg.
- *During strenuous exercise,* the diffusing capacity for CO_2 is increased to 1200 mL–1300 mL/min/mm Hg.

Clinical significance of higher diffusion capacity of CO_2. When diffusing capacity of respiratory membrane is markedly decreased due to certain diseases, one expects retention of CO_2 and lack of O_2. But due to vast difference in the diffusing capacities for O_2 and CO_2, a serious impairment of diffusion of O_2 causes significant lack of O_2 with little signs of CO_2 retention.

Equilibration time. It is estimated that the time required for the blood pCO_2 and alveolar air pCO_2 to equalize is also approximately 0.25 s.

Measurement of diffusion capacity of lungs

Diffusion capacity of lungs for different gases can be measured using the *Fick's law,* according to which diffusion capacity is given by

$$DL = \frac{V}{(pA - pC)},$$

where

DL = Diffusion capacity of lungs for a given gas.
V = Volume of the gas uptake in 1 min and (increase in the gas content of blood in 1 min).
pA – pC = Partial pressure gradient between the alveolar air and pulmonary capillary blood.

- Thus, diffusion capacity for O_2 (DLO_2) is

$$DLO_2 = \frac{O_2 \text{ consumption/min}}{(pAO_2 - pO_2)}$$

- It is easy to measure O_2 consumption/min and pAO_2, but it is difficult to measure pO_2 in pulmonary capillary blood (because collection of sample is extremely difficult).
- The diffusion capacity for O_2 is measured from the diffusion capacity for carbon monoxide.
- Carbon monoxide is preferred for measuring the diffusion capacity of lung because of two reasons:
  - After entering the blood, the CO very rapidly reacts with haemoglobin to form carboxyhaemoglobin and so do not allow the partial pressure of CO to build up in plasma. In this way, pCO in pulmonary capillary blood is almost zero.

– Diffusion of CO across respiratory membrane is diffusion limited. Therefore, the amount of CO transferred to the blood is the correct estimate of the diffusion capacity.

Procedure of single-breath carbon monoxide technique of measuring diffusion capacity of lung

- The subject is made to take a single breath of gas mixture containing dilute concentration (0.01%) of carbon monoxide (CO). He is asked to hold the breath for 10 s to allow the diffusion.
- The CO uptake/min is calculated from the difference between the inspired air and expired air CO concentration, and FRC. The pCO of alveolar air is estimated from the end-expiratory sample. The pCO of pulmonary capillary blood is considered zero and so the diffusion capacity for CO (DL CO) is calculated as

$$\text{DL CO} = \frac{\text{CO uptake/min}}{\text{pCO in alveolar air}}$$

- The diffusion capacity measured for CO at rest is about 17 mL/min/mm Hg.
- Diffusion capacity for O_2 is determined by multiplying the diffusion capacity for CO by 1.27, because diffusion coefficient of O_2 is about 1.2 times that of CO: $DLO_2 = 17 \times 1.2$, or about 20 mL/min/mm Hg.

Chapter 5.5

Transport of Gases

TRANSPORT OF OXYGEN

- **Uptake of oxygen by pulmonary blood**
- **Transport of oxygen in arterial blood**
  - Oxygen transport in dissolved form
  - Oxygen transport in combination with haemoglobin
    - Oxygenation of haemoglobin
    - Oxygen carrying capacity of haemoglobin
    - Oxygen–haemoglobin dissociation curve
    - Shifts in O_2–Hb dissociation curve
    - Concept of p_{50} and its significance
    - O_2–Hb dissociation curve of fetal haemoglobin
    - Effect of carbon monoxide on O_2 transport
    - Oxygen dissociation curve for myoglobin
- **Release of oxygen in tissues**
  - Vehicles for O_2 transport

TRANSPORT OF CARBON DIOXIDE

- **Diffusion of CO_2 in the blood**
- **Transport of CO_2 in the blood**
  - In dissolved form
  - In bicarbonate form
  - In carbamino form
  - Carbon dioxide dissociation curve
- **Release of CO_2 in the lungs**
- **Other facts about CO_2 transport**
  - Vehicles for CO_2 transport
  - Rate of total CO_2 transport
  - Changes in blood pH during CO_2 transport
  - Respiratory quotient

TRANSPORT OF OXYGEN

Transport of oxygen from the lungs to the tissues occurs due to constant *circulation of blood* and *diffusion* of O_2 that occurs in the direction of concentration gradient which is represented by O_2 tension (pO_2) differences given:

- Alveolar air pO_2: 104 mm Hg
- Arterial blood pO_2: 95 mm Hg
- Venous blood pO_2: 40 mm Hg
- Tissue interstitial fluid pO_2: 40 mm Hg

From the above, it is clear that:

- Oxygen from the alveolar air is taken up by the pulmonary capillary blood along a pressure gradient of 104–40 or 64 mm Hg and
- Oxygen from the arterial blood is released into the tissues by a pressure gradient of 95–40 or 55 mm Hg.

Transport of oxygen from the lungs to the tissues can be described as under:

A. Uptake of oxygen in the lungs by pulmonary blood,
B. Transport of oxygen in arterial blood and
C. Release (diffusion) of oxygen from blood to the interstitial fluid in tissues.

UPTAKE OF OXYGEN BY PULMONARY BLOOD

- As mentioned above, pO_2 of pulmonary arterial blood is about 40 mm Hg and that of alveolar air is 104 mm Hg. Therefore, due to this great concentration gradient oxygen readily diffuses from the alveoli into the blood.
- The process of diffusion across the respiratory membrane has been described in Chapter 5.4.

TRANSPORT OF OXYGEN IN ARTERIAL BLOOD

PO_2 in pulmonary venous blood is 104 mm Hg, however, by the time blood reaches the aorta pO_2 falls to about 100 mm Hg. This happens due to ***venous admixture,*** i.e. mixing of the venous blood to the arterial blood. The venous blood which mixes with the arterial blood is:

- Blood present in the bronchial veins (which forms about 2% of cardiac output),
- Part of the coronary venous blood flows into the chambers of left side of heart through thebesian veins.

Arterial blood contains about 20 mL and venous blood about 15 mL of O_2 per 100 mL. Thus, about 5 mL of O_2 is transported per 100 mL of blood from the lungs to the tissue cells.

Oxygen is transported in the blood in two forms: as dissolved form and in combination with haemoglobin.

OXYGEN TRANSPORT IN DISSOLVED FORM

- The solubility of O_2 in water (plasma) is so little that at pO_2 value of 100 mm Hg, out of the 20 mL of O_2 present in 100 mL of blood, only 0.3 mL is in dissolved form and rest is combined with haemoglobin (as oxyhaemoglobin).
- The dissolved oxygen obeys the Henry's law, i.e. amount dissolved is proportional to the pO_2. Thus, there is no limit to the amount of O_2 that can be carried in dissolved form, provided the pO_2 is sufficiently high. This is a distinct advantage over O_2 transport as oxyhaemoglobin which cannot exceed a certain limit.
- Therefore, dissolved O_2 at high pO_2 (hyperbaric oxygen) is utilized in clinical practice for the oxygenation of tissues when the haemoglobin gets denatured in certain types of poisoning, e.g. carbon monoxide poisoning.

OXYGEN TRANSPORT IN COMBINATION WITH HAEMOGLOBIN

Oxygenation of haemoglobin

After entering the blood from the alveolar air, most of the oxygen combines with haemoglobin to form a *loose and reversible* combination. This process is called *oxygenation* (not oxidation) and converts deoxyhaemoglobin into oxyhaemoglobin. The reaction is very rapid requiring <0.01 s. The *driving force* for this reaction is O_2 tension in the pulmonary capillaries.

One molecule of oxygen combines with one iron ion of the haem molecule. The O_2 molecule occupies the sixth *co-ordination* position of the iron atom. Since haemoglobin contains four molecules of haem, so each molecule of haemoglobin can combine with as many *as four O_2* molecules. The reaction proceeds in four steps:

$$Hb_4 + O_2 \rightleftharpoons Hb_4O_2$$
$$Hb_4O_2 + O_2 \rightleftharpoons Hb_4O_4$$
$$Hb_4O_4 + O_2 \rightleftharpoons Hb_4O_6$$
$$Hb_4O_6 + O_2 \rightleftharpoons Hb_4O_8$$

After first step, affinity of haemoglobin for O_2 increases with each next step. This happens so, because the insertion of O_2 molecule ruptures all the salt links and changes the quaternary structure of haemoglobin from the tense or *T-state* to relaxed or *R-state* which favours O_2 binding. Subsequent molecules of oxygen therefore, find it easier to bind to haem moieties. This phenomenon is termed as *co-operative binding kinetics.* This is the reason for the sigmoid nature of oxygen haemoglobin association or dissociation curve.

Oxygen carrying capacity of haemoglobin

One gram of haemoglobin can bind with maximum of 1.34 mL of O_2. Thus 100 mL of blood with haemoglobin level of 15 g/dL can carry 1.34 × 15, or 20.1 mL of oxygen. However, due to the presence of various physiological shunts (venous admixture) only 95% of the haemoglobin is available for carrying the oxygen. Therefore practically 100 mL of the arterial blood carries about 19.8 mL of oxygen out of which about 19.5 mL as oxyhaemoglobin and 0.3 mL as dissolved form in the plasma.

Thus, under normal circumstances, the haemoglobin carries most of the bulk of oxygen present in the blood. The only *disadvantage* is that the amount of oxygen that can be carried by the haemoglobin depends upon the concentration of haemoglobin in the blood.

Haemoglobin saturation is the percentage of haemoglobin that is combined with oxygen. When all the four sites on haemoglobin are occupied by O_2, then that molecule of haemoglobin is 100% saturated.

Oxygen–haemoglobin dissociation curve

Oxygen–haemoglobin curve refers to the curve obtained when the relation between the pO_2 and the percentage of haemoglobin saturation is plotted (Fig. 5.5-1). The O_2–Hb dissociation curve shows that percentage saturation of haemoglobin increases with the increase in pO_2 of the arterial blood. However, the relation is not linear but *sigmoid or S-shaped* (because of the reason explained above), but it has several physiological advantages (vide infra).

Two distinct zones of the O_2–Hb dissociation curve are recognized:

- *Loading (association) zone* refers to the upper flat part (plateau) of the curve, which is related to the process of O_2 *uptake in the lungs.* The curve shows that at pO_2 values of 100 mm Hg or above, the haemoglobin is 100% saturated. More important to note is the fact that even if pO_2 falls to 60 mm Hg, the Hb saturation is still 90%. Thus the loading zone provides a *margin of safety,* because it ensures fairly high uptake of O_2 by pulmonary blood even when alveolar pO_2 is moderately decreased in situations like:
  - Climbing mountains to moderate altitude and
  - Pulmonary diseases.
- *Unloading (dissociation) zone* of the O_2–Hb dissociation curve refers to the *steep portion* of the curve that occurs at pO_2 below 60 mm Hg. The steep part of the curve is concerned with O_2 delivery in the tissues and shows that large amounts of oxygen can be liberated from the blood with relatively minor fall of O_2 tension. This property keeps the O_2 tension in the capillary blood relatively high so that the diffusion gradient for O_2 is maintained. At pO_2 value of 40 mm Hg, Hb is 75% saturated with O_2 (Fig. 5.5-1). Thus, each 100 mL of blood can hold only 15 mL of O_2 as compared to 20 mL at pO_2 of 100 mm Hg

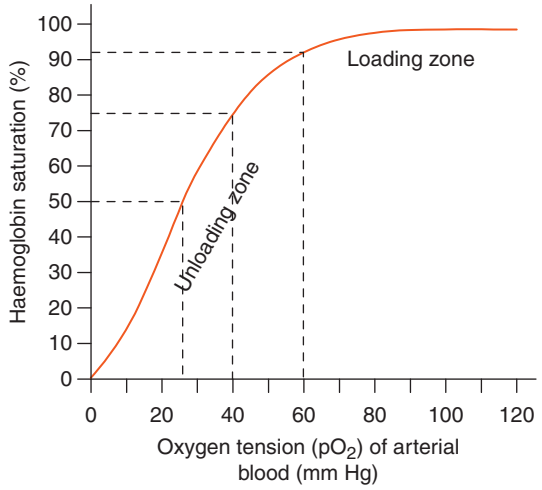


Fig. 5.5-1 Normal oxygen–haemoglobin dissociation curve.

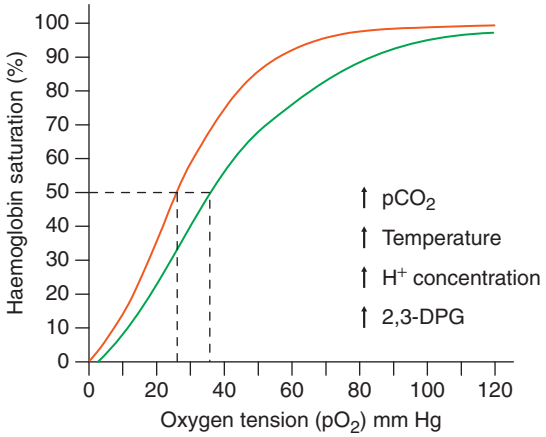


Fig. 5.5-3 Right shift of oxygen–haemoglobin dissociation curve which occurs due to increase in pCO_2, H^+ concentration, temperature and 2,3-DPG. Note that the p_{50} value is increased.

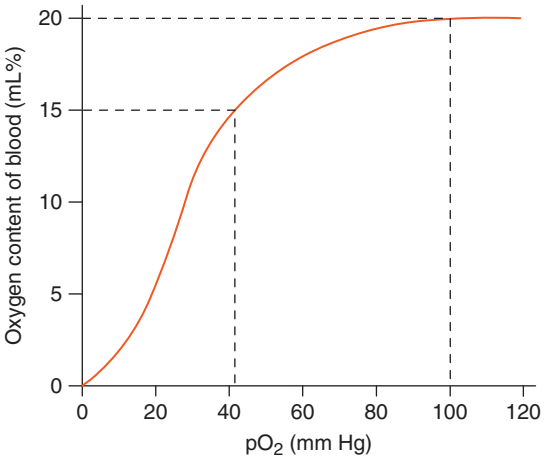


Fig. 5.5-2 The relationship between pO_2 and oxygen contents of the blood.

(Fig. 5.5-2). Thus 5 mL of O_2 is extracted by the tissues at rest. At values of pO_2 lower than 40 mm Hg still larger volume of O_2 would be offloaded and become available to the tissue (e.g. during exercise). The greater release of O_2 with slight decrease in tissue pO_2 (due to increased O_2 consumption) minimizes decrease of tissue pO_2 that would otherwise take place. Thus, it will regulate tissue pO_2 (buffering effect).

Physiological advantages of S-shaped O_2–Hb dissociation curve, which can be summarized from the above discussion, are:

- It allows greater uptake of O_2 at lungs despite great variation in the alveolar air pO_2.
- Tissues are supplied with O_2 according to their needs.
- Haemoglobin acts as a buffer and maintains tissue pO_2 at about 40 mm Hg.

Shifts in O_2–Hb dissociation curve

Several factors affect the affinity of haemoglobin for O_2 and thus shift the O_2–Hb dissociation curve to either right or left.

Shift to right

Shift to right of the O_2–Hb dissociation curve signifies decreased affinity of haemoglobin for O_2 (Fig. 5.5-3). Thus at every level of pO_2, the oxygen saturation of haemoglobin is somewhat lower than the normal curve leading to more offloading of O_2.

Causes. The factors causing right shift are:

- **An increase in pCO_2** shifts the curve to the right, this phenomenon is known as *Bohr's effect* (Fig. 5.5-3). When the arterial blood reaches the tissues, it is exposed to not only low tissue pO_2 (40 mm Hg) but also higher pCO_2 (45 mm Hg); So, normally in tissues the curve shifts to the right and due to the Bohr's effect for a given decrease in pO_2 larger volume of O_2 is shed off (about 2% more).
- **A decrease in pH of blood**, as occurs in the tissues, also shifts the O_2–Hb dissociation curve to the right.
- **An increase in the temperature** of blood shifts the curve to the right.
- **Effect of diphosphoglycerate.** The red blood cells are rich in 2,3-diphosphoglycerate (2,3-DPG) which is formed from 3-phosphoglyceraldehyde produced during glycolysis via *Embden-Meyerhof pathway.*

  Thus, an increase in the concentration of 2,3-DPG decreases the affinity of Hb for O_2 and shifts the normal O_2–Hb dissociation curve to the right. Factors affecting 2,3-DPG concentrations in RBC are depicted in Table 5.5-1.

Table 5.5-1 Factors affecting 2,3-DPG concentration in RBC

Increase	Decrease
• Exposure to chronic hypoxia at high altitude and certain pulmonary diseases	• Decrease in blood pH (acidosis)
• Anaemia	• Stored blood (acid citrated buffer used for storage inhibits glycolysis in RBC leading to decreased 2,3-DPG)
• Exercise	
• Hormones: Thyroxin, androgen and growth hormone	
• Rise of body temperature	

IMPORTANT NOTE

During *exercise* the demand of O_2 is increased in the skeletal muscles. The factors which facilitate the delivery of the required larger amounts of O_2 (by causing right shift in the curve) are:

- Increased temperature in the skeletal muscles due to more heat production,
- Increased pCO_2 due to accumulation of CO_2 resulting from rapid metabolism,
- Decreased pO_2 due to rapid consumption and
- Decreased pH due to accumulation of lactic acid produced during muscular exercise.

Advantages versus disadvantages of right shift

- Right shift is advantageous in the tissues where greater O_2 is released from the haemoglobin (at the same pO_2).
- However, right shift is disadvantageous in the lungs because (at the same pO_2) blood takes up less oxygen.

Shift to left

Shift to left of the O_2–Hb dissociation curve signifies increased affinity of haemoglobin for O_2 (Fig. 5.5-4). Thus, at every pO_2 level the oxygen saturation of haemoglobin is somewhat greater than the normal curve.

Causes for left shift of the curve are:
- Decreased pCO_2 of blood
- Increased pH of blood
- Decreased temperature
- Fetal haemoglobin

Advantages versus disadvantages of left shift. Left shift of the curve has limited advantage because though it allows greater uptake of O_2 at lungs (at same pO_2) but it decreases release of O_2 to the tissues (at same pO_2).

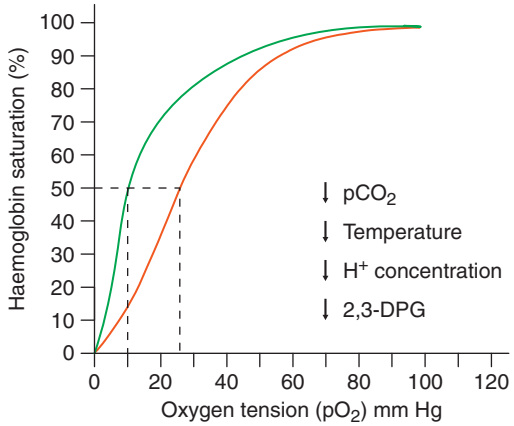


Fig. 5.5-4 Left shift of oxygen–haemoglobin dissociation curve which occurs due to decrease in pCO_2, H^+ concentration, temperature and 2,3-DPG. Note that the p_{50} value is decreased.

Concept of p_{50} and its significance

p_{50} refers to the partial pressure of O_2 that produces a 50% saturation of the haemoglobin with O_2; Normal p_{50} for the arterial blood in an adult at 37°C body temperature and at pCO_2 of blood 40 mm Hg is 25–27 mm Hg (Fig. 5.5-1).

Haemoglobin affinity for O_2 is inversely related to the p_{50} value, therefore:

- *Decreased p_{50}* (Hb gets 50% saturated at lower pO_2) indicates increased affinity of haemoglobin for O_2. Thus, decreased p_{50} is equivalent to shift of O_2–Hb dissociation curve to the left (Fig. 5.5-4). Fetal haemoglobin and myoglobin have lower p_{50} value than the adult haemoglobin.
- *Increased p_{50}* (Hb gets 50% saturated at higher pO_2) indicates decreased affinity of haemoglobin for O_2. Thus, increased p_{50} is equivalent to the right shift of O_2–Hb dissociation curve (Fig. 5.5-3).

O_2–Hb dissociation curve of fetal haemoglobin

- As shown in Fig. 5.5-5, the O_2–Hb dissociation curve of fetal haemoglobin (HbF) is shifted to the left in comparison with the O_2–Hb dissociation curve of adult haemoglobin (HbA).
- The O_2–Hb dissociation curve of HbF is shifted to left because its affinity for 2,3-DPG is considerably less than that of HbA. This is because two gamma (γ) chains present in HbF have very little affinity for 2,3-DPG as compared to the beta (β) chains of HbA.
- Thus, affinity of HbF to combine with O_2 is more than that of HbA. This property of HbF helps it to take up

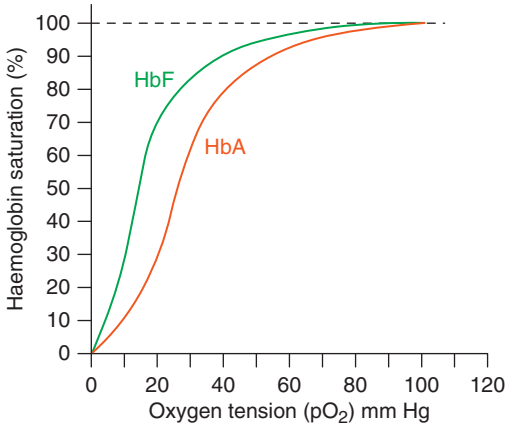


Fig. 5.5-5 Oxygen–haemoglobin dissociation curves of adult haemoglobin (HbA) and fetal haemoglobin (HbF).

normal volumes of O_2 in spite of the fact that the fetal blood is exposed to rather low pO_2 values of the maternal blood in placenta. As shown in Fig. 5.5-5, at pO_2 of 20 mm Hg where HbA is only 35% saturated the HbF is 70% saturated that is why HbF can store more O_2.

Effect of carbon monoxide on O_2 transport

- Carbon monoxide (CO) interferes with O_2 transport because it has about 200 times the affinity (of oxygen) for haemoglobin.
- CO combines with Hb at the same site on its molecule as O_2 and forms the *carboxyhaemoglobin* and thus decreases the functional haemoglobin concentration.
- Because of the extreme affinity of CO for haemoglobin, the carboxyhaemoglobin curve lies along the Y-axis and a CO tension of only 0.5 mm Hg inactivates about 50% of the haemoglobin, i.e. p_{50} for CO is only 0.5 mm Hg (Fig. 5.5-6).
- CO lowers the tissue O_2 tension by decreasing the O_2 content and the p_{50} for O_2.

Oxygen dissociation curve for myoglobin

Myoglobin is present in higher quantities in the muscles specialized for sustained contraction, e.g. muscles of legs and heart. The characteristic features of O_2 dissociation curve of myoglobin *vis-a-vis* haemoglobin are (Fig. 5.5-7):

- Dissociation curve of myoglobin is *rectangular hyperbola* rather than sigmoid in shape, because it takes up O_2 at low pressure much readily, i.e. rate of association of myoglobin with O_2 is very fast.
- Myoglobin does not show the Bohr's effect.
- At pO_2 of 40 mm Hg the myoglobin is 95% saturated while haemoglobin is 75% saturated. Even at pO_2 of 5 mm Hg myoglobin is saturated by slightly less than 60%. Thus, it acts as a temporary store house for O_2 in the muscles.

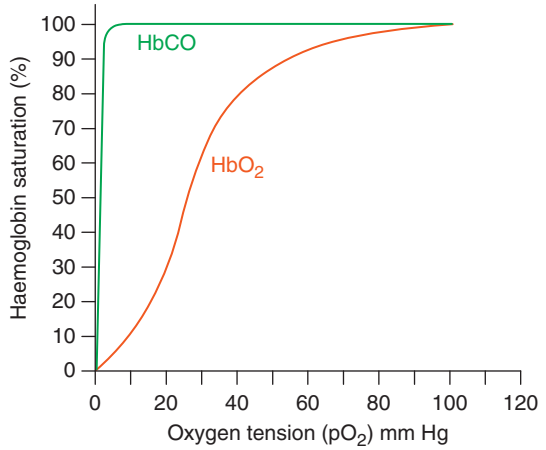


Fig. 5.5-6 Haemoglobin saturation curves for both O_2 (HbO_2) and carbon monoxide (HbCO). Note that HbCO curve lies along the Y axis which indicates extreme high affinity of haemoglobin for carbon monoxide. Note that the p_{50} value for CO is only 0.5 mm Hg.

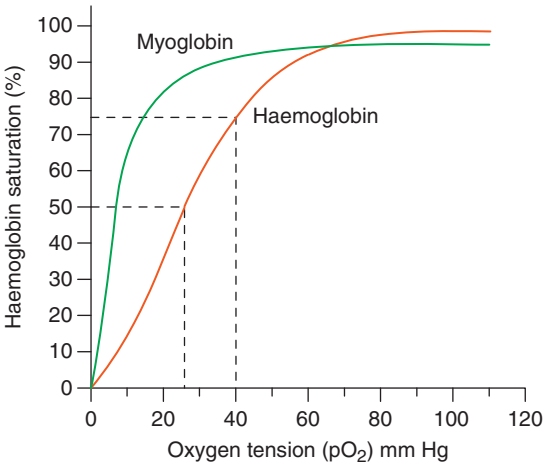


Fig. 5.5-7 Oxygen dissociation curve of myoglobin versus haemoglobin.

RELEASE OF OXYGEN IN TISSUES

OXYGEN RELEASE AT REST

Oxygen delivery represents the amount of O_2 that is presented to body cells per minute and is equal to the arterial O_2 content multiplied by a cardiac output. Since 100 mL of arterial blood at pO_2 of about 100 mm Hg contains about 20 mL of oxygen, thus with a cardiac output of about 5 L/min, the normal oxygen delivery to the entire body is about 1 L/min. The oxygen delivery to the tissues decreases with either decrease in the arterial O_2 content or decrease in the cardiac output.

Oxygen consumption. When the arterial blood with approximate pO_2 of 100 mm Hg, reaches the tissues with tissue fluid pO_2 of 40 mm Hg; because of pressure gradient,

about 5 mL of O_2 diffuses from the tissue capillaries to the interstitial fluid out of 100 mL of blood (containing ~19 mL O_2) every minute. Thus, oxygen consumption of the whole body at rest with a cardiac output of 5 L/min about

$$= \frac{5 \times 5000}{100}, \text{ or } 250 \text{ mL of } O_2 \text{ per minute.}$$

Utilization coefficient. Utilization coefficient refers to the percentage of oxygen consumed out of oxygen delivered to the tissue, i.e.

Coefficient of utilization

$$= \frac{\text{Oxygen consumed/min}}{\text{Oxygen delivered/min}} \times 100$$

So, at rest coefficient of utilization of whole body

$$= \frac{250 \text{ mL/min}}{1000 \text{ mL/min}} \times 100 = 25\%$$

VEHICLES FOR TRANSPORT OF OXYGEN: A COMPARISON OF PLASMA, HAEMOGLOBIN AND WHOLE BLOOD

The amount of oxygen that can be loaded and unloaded by different transport vehicles is depicted in Table 5.5-2.

From the lungs at pO_2 of about 100 mm Hg, the amount of O_2 loaded by 100 mL of transport vehicle is:

- Whole blood : 19.8 mL
- Haemoglobin solution : 19.5 mL
- Plasma solution : 0.3 mL only

In the tissue at pO_2 of 40 mm Hg, the amount of oxygen released by 100 mL of transport vehicle is:

- Whole blood : 5 mL,
- Haemoglobin solution : 1.5 mL and
- Plasma : 0.18 mL

At maximum haemoglobin saturation, the whole blood can release:

- 5 mL of O_2 at rest when tissue pO_2 is about 40 mm Hg,
- 13 mL of O_2 during moderate exercise when tissue pO_2 is about 25 mm Hg and
- 15–16 mL of O_2 during severe exercise, when tissue pO_2 is about 15 mm Hg.

From the above, it is quite clear that the whole blood is an ideal vehicle for the transport of O_2 to load itself in the lungs with O_2 and to release O_2 in the tissues as per requirement.

Table 5.5-2 Amount of oxygen held by different vehicles

Vehicle	Oxygen in arterial blood at pO_2 at 100 mm Hg/dL	Oxygen in venous blood at pO_2 at 40 mm Hg/dL	Oxygen released to tissues/dL
Plasma	0.3 mL	0.12 mL	0.18 mL
Haemoglobin solution	19.0–19.5 mL	18.0 mL	1.0–1.5 mL
Whole blood	19.8 mL	14.0 mL	5.0 mL

TRANSPORT OF CARBON DIOXIDE

Transport of carbon dioxide from the tissue cells to the lungs occurs due to *constant circulation* of blood and diffusion of CO_2 that occurs at various sites in the direction of concentration gradient which is represented by CO_2 tension differences as given:

- Intracellular pCO_2 : 46 mm Hg
- Interstitial fluid pCO_2 : 45 mm Hg
- Arterial blood pCO_2 : 40 mm Hg (in tissue capillaries)
- Venous blood pCO_2 : 45 mm Hg
- Alveolar air pCO_2 : 40 mm Hg.

From the above pCO_2 levels, it is clear that:

- CO_2 from the cells diffuses into the interstitial fluid along a tension gradient of 1 mm Hg,
- From the interstitial fluid, the CO_2 diffuses into the capillaries at a tension gradient of 5 mm Hg and
- From the venous blood that is supplied to the pulmonary capillaries the CO_2 diffuses across the respiratory membrane into the alveoli along a tension gradient of 5 mm Hg.

Transport of carbon dioxide from the tissue cells to the lungs can be described as under:

A. Diffusion of CO_2 in the blood,
B. Transport of CO_2 in the blood and
C. Release of CO_2 in the lungs.

DIFFUSION OF CO_2 IN THE BLOOD

Tissue cells constantly form CO_2 inside the cells due to metabolism. As mentioned above, intracellular pCO_2 is 46 mm Hg and that of interstitial fluid surrounding the cells is 45 mm Hg. Though the cells are continuously forming CO_2 but still the CO_2 tension gradient between inside and outside of the cells is only 1 mm Hg. This is owing to the rapid diffusion of CO_2 (20 times that of O_2) out of the cells into the interstitial fluid. Since pCO_2 of the arterial blood flowing in the tissue capillaries is lower (40 mm Hg) than that of the interstitial fluid, so CO_2 diffuses inside the capillary blood which flows in the systemic venous system.

TRANSPORT OF CO_2 IN THE BLOOD

Venous blood contains about 52 mL and the arterial blood about 48 mL of CO_2 per 100 mL. Thus, 4 mL of CO_2 is transported per 100 mL of blood from tissue cells to the lungs. Thus, total CO_2 transported from the whole body tissue cells at rest with a cardiac output of 5 L/min is about 4 × 5000/100, or 200 mL/min.

Carbon dioxide is transported in the blood in three forms (Fig. 5.5-8):

- In dissolved state (70%),
- In bicarbonate form (70%) and
- In carbamino compound form (23%).

CARBON DIOXIDE TRANSPORT IN DISSOLVED FORM

- The venous blood, with pCO_2 of 45 mm Hg contains about 2.7 mL/100 mL of CO_2 in dissolved state.
- The arterial blood with pCO_2 of 40 mm Hg contains about 2.4 mL/100 mL of CO_2 in dissolved state.
- Thus only 0.3 mL of CO_2 is transported in a dissolved state per 100 mL of blood from tissues to the lungs. This represents about 7% of all CO_2 that is transported.

CARBON DIOXIDE TRANSPORT IN BICARBONATE FORM

Approximately 70% of the carbon dioxide is transported in the form of *plasma bicarbonate ions.* However, these bicarbonate ions are formed in the RBCs and then diffuse into the plasma as explained:

- After entering the blood, most of the CO_2 enters the RBCs, wherein in the presence of *carbonic anhydrase,* it rapidly reacts with water to form carbonic acid (H_2CO_3).
- Carbonic acid (H_2CO_3) dissociates into bicarbonate ions (HCO_3^-) and hydrogen ions (H^+).
- *The bicarbonate ions diffuse* out of the RBCs into plasma and are transported as sodium bicarbonate (alkali reserve of blood). Thus, although HCO_3^- is formed within the RBCs, most of the CO_2 is carried in the plasma as HCO_3^- and some in the RBCs as well.
- *The H^+ are buffered* by deoxygenated haemoglobin, which is a weaker acid than oxyhaemoglobin. This enables the reaction to proceed unabated in the forward direction.
- *In chloride shift* (Hamburger phenomenon), the HCO_3^- diffuses out of the RBCs into the plasma, the inside of the cells become less negatively charged. Because the RBC membrane is relatively impermeable to cations, so in order to neutralize this effect, negatively charged chloride ions (Cl^-) diffuse from the plasma into the RBCs to replace the HCO_3^-. The movement of chloride ions into the RBCs is called *chloride shift* or Hamburger phenomenon. This process is mediated by *Band 3*, a major ion exchange membrane protein. The chloride shift occurs very rapidly and essentially completed within 1 s.

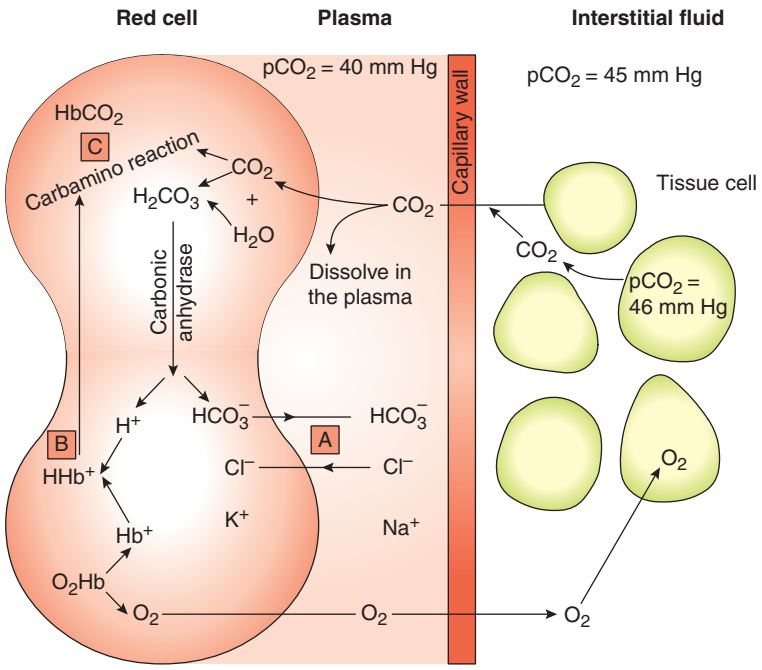


Fig. 5.5-8 A, Transport of carbon dioxide (CO_2) in blood demonstrating formation of HCO_3^- and chloride shift phenomenon; B, H^+ buffering by haemoglobin and C, formation of carbaminohaemoglobin.

- As a result of chloride shift, the total number of ions inside the RBCs increase, so the osmotic pressure inside the RBCs becomes higher than that of plasma. This causes osmotic absorption of fluid into the RBCs. Thus, venous RBCs contain the greater quantity of fluid as compared to the arterial blood RBCs. Because of this:
- Packed cell volume of venous blood is slightly higher (~3%) than that of arterial blood and
- Venous RBCs are more fragile than the arterial RBCs.

TRANSPORT OF CARBON DIOXIDE IN CARBAMINO FORM

Approximately 23% of the total CO_2 is transported in the blood in the form of carbamino compounds. After entering the blood, some of the CO_2 combines with the amino group ($-NH_2$) of proteins to form carbamino compounds.

In the plasma, CO_2 combines with an amino group of plasma proteins ($PrNH_2$) to form carbamino proteins:

$$CO_2 + PrNH_2 \rightarrow Pr.NH.COOH$$

- This reaction is much less significant because quantity of these proteins is one-fourth that of haemoglobin.

In the RBCs, CO_2 combines with an amino group of haemoglobin ($HbNH_2$) to form a compound called carbamino haemoglobin.

$$CO_2 + HbNH_2 \rightarrow Hb.NH.COOH$$

- This combination of CO_2 with haemoglobin is a reversible reaction that occurs with a loose bond, so that the CO_2 is easily released into the alveoli where pCO_2 is lower than that in the tissue capillaries.
- This reaction of CO_2 with Hb is much slower than the reaction of CO_2 with water in RBCs. This is because, more CO_2 (70%) is transported as bicarbonates and less as carbamino compounds (23%).
- This reaction of CO_2 with Hb is further decreased to a great extent when 2,3-DPG concentration is more, because both 2,3-DPG and CO_2 compete for the same sites on Hb.

CARBON DIOXIDE DISSOCIATION CURVE

- Carbon dioxide dissociation curve is obtained by plotting the relationship between pCO_2 and total CO_2 content of the blood (Fig. 5.5-9).
- The graph shows that relationship between the two is nearly linear over wider range of pCO_2 (if compared with O_2–Hb, dissociation curve, which is sigmoid shaped).
- It is important to note that practically in the body the pCO_2 value of arterial and venous blood varies within a narrow range of 40–45 mm Hg (in contrast the corresponding values of pO_2 vary from 100 to 40 mm Hg).

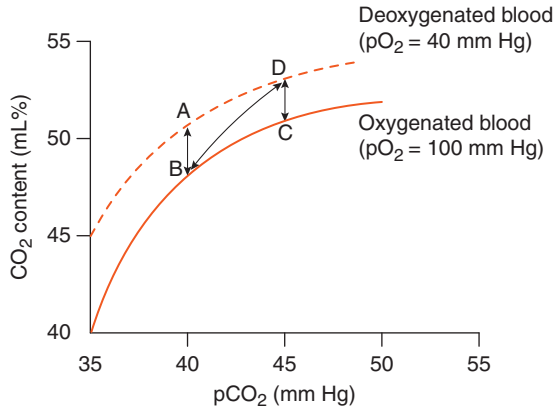


Fig. 5.5-9 Carbon dioxide dissociation curve for oxygenated (solid line) and for deoxygenated blood (dotted line) to demonstrate Haldane's effect.

Therefore, the full range of CO_2 dissociation curve shown in Fig. 5.5-9 is an experimental theoretical phenomenon and does not operate in the body practically. Physiologically, the curve operates only within A–V range (Fig. 5.5-9).

Factors affecting CO_2 dissociation curve

1. Oxygen. Deoxyhaemoglobin present in the tissue capillaries is capable of loading more CO_2 than the oxyhaemoglobin, and also the oxygenation of haemoglobin in the lungs increases the CO_2 unloading. This effect is called *Haldane's effect.* Haldane's effect on the transport of CO_2 depicts that:

- Blood with pCO_2 of 40 mm Hg reaching the tissues is capable of drawing CO_2 from the tissues more at pO_2 of 40 mm Hg (point A) than at pO_2 of 100 mm Hg (point B). Thus, because of the Haldane's effect, the CO_2 dissociation curve shifts to the left when blood flows to the tissues.
- Blood with pCO_2 of 45 mm Hg reaching the lungs is capable of retaining less or in other words releasing more CO_2 in the lungs at pO_2 of 100 mm Hg (point C) than at pO_2 of 40 mm Hg (point D). Thus, because of the Haldane's effect, the CO_2 dissociation curve shifts to the right when blood flows in the pulmonary capillaries.
- Figure 5.5-9 also depicts that the Haldane's effect is beneficial because it almost doubles the quantity of CO_2 to be carried from the tissues to lungs and also doubles the amount of CO_2 to be excreted by the lungs.

2. 2,3-DPG. 2,3-DPG decreases the formation of carbaminohaemoglobin because it competes with CO_2 for the same sites on Hb especially in case of reduced blood. Thus, 2,3-DPG shifts the CO_2 dissociation curve to the right meaning thereby that CO_2 carrying capacity is decreased.

3. Increase in body temperature causes release of O_2 from blood, which causes left shift of the CO_2 dissociation curve, i.e. larger amount of CO_2 can be taken off at a given pCO_2.

RELEASE OF CO_2 IN THE LUNGS

When the venous blood with pCO_2 of 45 mm Hg, CO_2 content of 52 mL/100 mL and pO_2 of 40 mm Hg reaches the pulmonary capillaries, it is separated from the alveolar air having pCO_2 of 40 mm Hg. CO_2 content of 48 mL/100 mL and pO_2 of 104 mm Hg by the respiratory membrane; following changes occur which lead to diffusion of 4 mL of CO_2 in 100 mL of blood/min in the alveoli (Fig. 5.5-10):

Release of CO_2 from carbaminohaemoglobin into plasma (Fig. 5.5-10A)

- O_2 diffuses into the capillary blood with a concentration gradient of 104–40 mm Hg or 64 mm Hg.
- The O_2 enters the RBCs and converts the deoxyhaemoglobin into oxyhaemoglobin which has very low affinity for CO_2. Therefore, this CO_2 is released from the carbaminohaemoglobin which diffuses into the plasma.

Release of CO_2 from bicarbonate into plasma (Fig. 5.5-10B)

- The oxyhaemoglobin so formed is a strong acid. Therefore, increased acidity of the blood results in increased H^+ concentration. To neutralize it the bicarbonate (HCO_3^-) ions diffuse into the RBCs where H^+ and HCO_3^- react to form H_2CO_3 (carbonic acid) which dissociates to form H_2O and CO_2. This whole reaction occurs in the presence of enzyme carbonic anhydrase. CO_2 so released diffuses into the plasma.
- With the movement of HCO_3^- inside the RBCs, inside of the cell becomes more negative. To neutralize them, either the positive charged cations should move in or negative charge anions should flow out of the cell. Since RBC membrane is relatively impermeable to cations, so the Cl^- anions return in the plasma from the RBC. This whole reaction catalyzed by carbonic anhydrase inhibitor is called *reversal of chloride shift.*

Diffusion of CO_2 from plasma to alveoli

The CO_2 dissolved in plasma plus that released from the carbaminohaemoglobin and bicarbonates combinedly exerts pCO_2 of 45 mm Hg. Since the pCO_2 of alveolar air is 40 mm Hg, so because of pressure gradient CO_2 diffuses from the blood to the alveoli. Due to constant ventilation, CO_2 from alveoli is transported to atmosphere.

OTHER FACTS ABOUT CO_2 TRANSPORT

VEHICLES FOR CO_2 TRANSPORT

The amount of CO_2 that can be loaded and unloaded by different transport vehicles from the tissues at pCO_2 of 45 mm Hg (Table 5.5-3) reveal that:

- *Plasma is* not a good transport vehicle as very small amount of CO_2 (only 0.2 mL) can be taken from tissues/100 mL of blood/min.

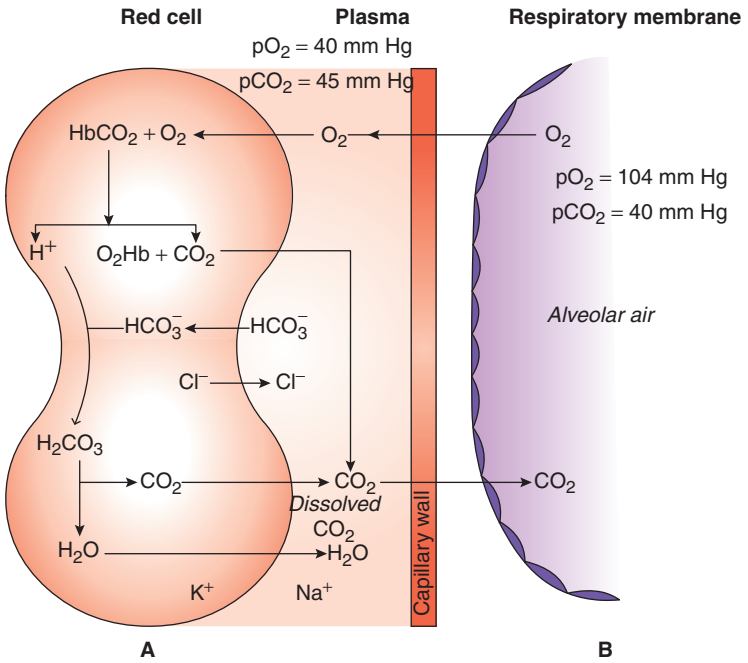


Fig. 5.5-10 A, release of carbon dioxide in the plasma from carbaminohaemoglobin and from bicarbonate ions and B, diffusion of CO_2 from the plasma into the alveoli through a respiratory membrane.

Table 5.5-3 Amount of CO_2 that can be loaded and unloaded by different transport vehicles

Vehicle	Content of CO_2/100 mL/min		
	In venous blood at pCO_2 of 45 mm Hg	In arterial blood at pCO_2 of 40 mm Hg	Loaded by the blood
Plasma	1.8 mL	1.6 mL	0.2 mL
Bicarbonate solution	48 mL	48 mL	Nil
Whole blood	52 mL	48 mL	4 mL

- *Bicarbonate solution* is also not a good transport vehicle because beyond pO_2 of 40 mm Hg there is no further transport of CO_2.
- *Whole blood* is an ideal vehicle for transport of CO_2 to load itself in tissues with CO_2 and to release CO_2 in the lung.

RATE OF TOTAL CO_2 TRANSPORT

In resting conditions, each 100 mL of blood transports about 4 mL of CO_2 from the tissues to the lungs. Thus, with an average cardiac output of 5 L/min, a total of (4 × 5000)/100 or 200 mL of CO_2 is transported/min.

During exercise, the amount of CO_2 transported increases depending upon the severity of exercise. In severe exercise, as much as 4 L of CO_2 may be transported per minute. Because of greater solubility and transport in different forms, the transport of such large amounts of CO_2 occurs without any difficulty. The conversion of most of CO_2 into bicarbonate ions prevents any significant change in the pH of blood even when such a large volume of CO_2 enters the circulation.

CHANGES IN BLOOD pH DURING TRANSPORT OF CO_2

Normally, the pH of arterial blood is 7.4. As it passes through the tissues, it acquires CO_2 and the pH of blood falls due to formation of carbonic acid (H_2CO_3) in the venous blood. The pH may fall by about 0.4. However, during exercise the fall in pH may become to the tune of 0.5 and even more. Nevertheless, most of it is neutralized by the blood buffers.

RESPIRATORY QUOTIENT

Definition. Respiratory quotient (RQ) refers to the ratio of the rate of CO_2 excretion and rate of O_2 consumption per minute. It is also called *respiratory exchange ratio.*

Normal value. Normally, the rate of CO_2 excretion is 4 mL/100 mL/min and rate of O_2 consumption is 5 mL/100 mL/min. So, respiratory quotient = 4/5 or 0.8.

Chapter 5.6

Regulation of Respiration

INTRODUCTION

Respiration is regulated by a complex integration of neural control mechanisms which are modified by certain respiratory reflexes and chemical control mechanisms.

Neural control mechanisms include:

- *A system for automatic control* of respiration as an involuntary function. The involuntary control system of respiration is located in the medullary and pontine centres of the brainstem.
- *A system for voluntary control* of respiration is located in the cerebral cortex.

Thus, respiration enjoys the distinction of being an involuntary function which can be influenced voluntarily. This *dual control* has great functional significance, i.e.

- Involuntary control which allows human to breathe without conscious efforts under all circumstances including sleep and is thus essential for life.
- The voluntary control system facilitates acts like talking, singing, swimming, breath holding and voluntary hyperventilation.

Respiratory reflexes which modify the effects of neural mechanisms are those initiated by stimulation of stretch receptors, irritant receptors, J-receptors and chest wall receptors.

Chemical control mechanisms are influenced by alterations in arterial pO_2, pCO_2 and H^+ concentration. The chemical control mechanisms are initiated by stimulation of the chemoreceptors (central and peripheral).

Functions of respiratory regulatory mechanisms include:

- *Genesis of normal respiratory spontaneous rhythm.* This is the function of medullary and pontine centres of neural mechanism.
- *Control of rate and depth of respiration,* i.e. adjustment of total ventilation to match metabolic needs of the body so that arterial oxygen tension (pO_2), carbon dioxide

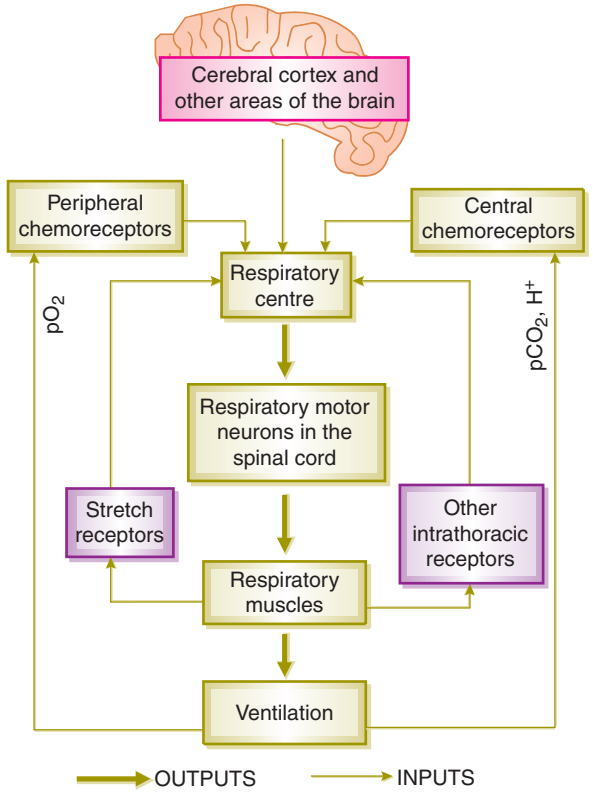


Fig. 5.6-1 Diagrammatic representation of the inputs to the respiratory centre that generate the output to the respiratory muscles.

tension (pCO_2) and H^+ concentration (pH) are almost maintained constant whether it be during quiet breathing, sleep or muscular exercise. This function is accomplished by all the respiratory control mechanisms acting in unison (Fig. 5.6-1):

- *Sensors,* i.e. chemoreceptors and other receptors (e.g. stretch, irritant, chest wall and J-receptors), perceive the respiratory needs of the body and convey via afferent nerves to the central controller.
- *Central controllers,* i.e. medullary, pontine and other parts of the brain, adjust the efferent outputs as per the body needs and convey to the effectors.
- *Effectors* are the respiratory muscles which perform their activity as per the neural discharges received.

NEURAL REGULATION OF RESPIRATION

The neural mechanisms regulating respiration can be described under two headings:

- Automatic control system and
- Afferent impulses to respiratory centres.

AUTOMATIC CONTROL SYSTEM

NEURAL GENESIS OF RESPIRATORY RHYTHM

The involuntary neural control system regulates respiration by several groups of neurons situated bilaterally in medulla and pons, which include medullary respiratory centres, pontine respiratory centres and reticular activating system (RAS).

I. MEDULLARY RESPIRATORY CENTRES

The medullary respiratory centres include two groups of neurons: the dorsal respiratory group (DRG) and ventral respiratory group (VRG), which generate the basic respiratory rhythm.

The respiratory control pattern generator. The respiratory control pattern generator, which is responsible for automatic respiration, is located in the medulla. A group of neurons called pacemaker cells form the pre-Botzinger complex, which is situated between nucleus ambiguus and lateral reticular nucleus. The rhythmic activity is initiated by these synaptically coupled neurons. These neurons discharge rhythmically and generate rhythmic motor activity in phrenic nerve, hypoglossal nerve and intercostal nerves.

1. Dorsal respiratory group neurons

Most of the neurons are located within the nucleus of tractus solitarius (NTS) and some in the adjacent reticular substance (Fig. 5.6-2). The neurons of DRG are of three types:

(i) Inspiratory neurons. The DRG mainly contains inspiratory cells called *I-neurons* that discharge during inspiration only. The axons of I-neurons cross the midline and descend on the contralateral side of spinal cord to make contact with the spinal motor neurons of the inspiratory muscles, namely the diaphragm (supplied by the phrenic nerve arising from C_3 to C_5) and the external intercostal muscles (supplied by the intercostal nerves). In other words, the neurons in the DRG are the upper motor neurons of respiratory muscles.

From *I-neurons,* nerve signals pass to the muscles of inspiration. The signal is not instantaneous but is a ramp signal, i.e. it is weak in the beginning and it steadily increases in a ramp manner for about 2 s and is thus called *inspiratory ramp.* This leads to a steady increase in the lung volume during inspiration rather than the respiratory gasps (abrupt distension). Ramp signal then abruptly ceases for approximately next 3 s.

(ii) Inspiratory off-switch (IOS) neurons. Inspiratory off-switch neurons refer to a group of neurons that are responsible for terminating the activity of I-neurons and causes turning off of excitation of muscles of inspiration (diaphragm and external intercostal muscles). These muscles, therefore relax allowing elastic recoil of the chest wall and the lungs

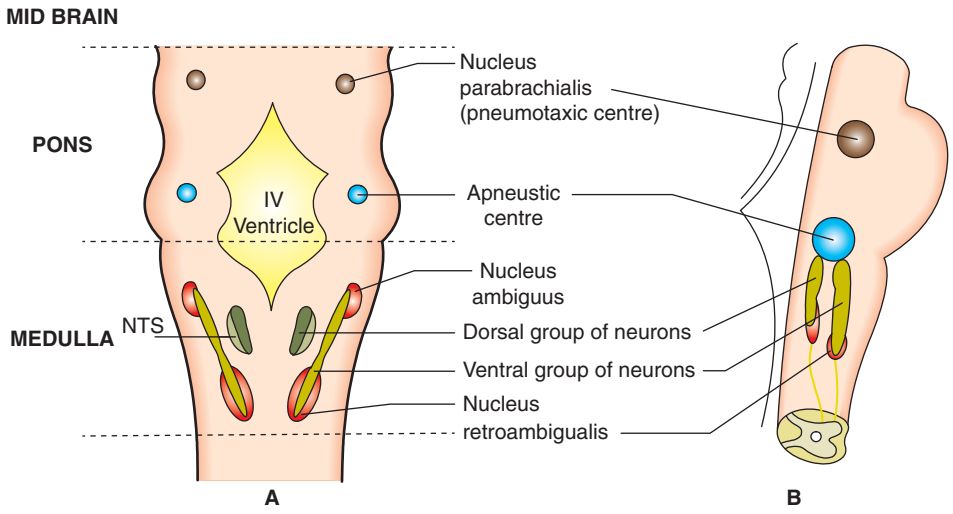


Fig. 5.6-2 Medullary and pontine respiratory centres: A, front view and B, lateral view.

to cause expiration. After expiration again there is signal for starting another cycle.

(iii) Integrator neurons. Integrator neurons are other type of neurons of DRG present near the I-neurons and are stimulated by them. They subserve following integrating functions:

- The integrator neurons when depolarize to a critical level lead to firing of the so-called *IOS* neurons which are responsible for terminating the inspiratory ramp. The *I-neurons* trigger the IOS neurons indirectly through integrators and thereby bring about the termination of their own discharge. This forms the basic circuity of an automatic respiratory rhythm (Fig. 5.6-3). In this way, cycle of inspiration/expiration goes on continuously to cause tidal respiration.
- The *integrator neurons receive both excitatory and inhibitory inputs* (Fig. 5.6-3) and thus integrate the activity of *I-neurons* and IOS neurons accordingly.
- *The excitatory inputs* to integrator neurons come from:
  - Cerebral cortex
  - Pneumotaxic centre
  - Vagal afferents from stretch receptors
- *The inhibitory inputs* to integrator neurons come from the medullary inhibitory neurons which form the so-called *apneustic centre.*

2. Ventral respiratory group neurons

Ventral respiratory group neurons contain both inspiratory cells called I-neurons and expiratory cells called E-neurons (cf DRG which contain only I-neurons). Their axons cross the midline and descend on the contralateral side to make contact with the motor neuron pool for the muscles of expiration, i.e. internal intercostal muscles and abdominal muscles.

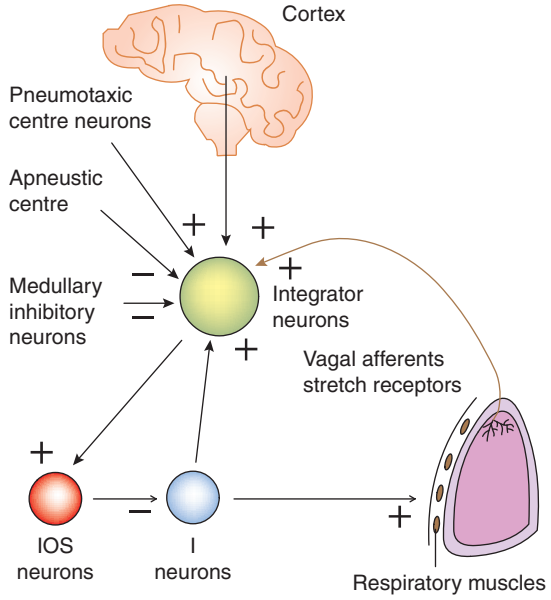


Fig. 5.6-3 Basic circuity for generation of respiratory rhythm.

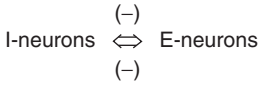


Fig. 5.6-4 Reciprocal innervation of I- and E-neurons.

Interaction of I- and E-neurons

The I- and E-neurons have inhibitory connections to each other, i.e. there exists *reciprocal innervation* between the two. Therefore, the motor neurons to the expiratory muscles are inhibited when those supplying the inspiratory muscles are active and vice versa (Fig. 5.6-4). This reciprocal innervation is mediated via collaterals from excitatory pathway that synapses on inhibitory interneurons. Thus, an

impulse that stimulates the one will inhibit the other and vice versa.

Role of VRG neurons

The VRG neurons normally remain totally inactive during quiet breathing. The VRG neurons become active during inspiration (role of I-neurons) as well as expiration (role of E-neurons) during forceful respiration. This area is especially important in providing powerful expiratory signals to expiratory muscles. Thus, VRG operates when high levels of pulmonary ventilation is required, for example, during exercise.

II. PONTINE RESPIRATORY CENTRES

The pontine centres include the apneustic centre (APN) and pneumotaxic centre (PNC), both of which modify the activity of medullary respiratory centres.

1. Apneustic centre

Apneustic centre refers to a group of inhibitory neurons located bilaterally in the lower part of pons (Fig. 5.6-2). It sends signals to the integrator neurons of DRG that affect the IOS neurons and prevent the switch-off of the inspiratory ramp signals from the central inspiratory neurons (Fig. 5.6-3). This increases the tidal volume and duration of inspiration, resulting in a deeper and more prolonged inspiratory effort termed as *apneusis.* However, normally the apneustic centre is inhibited by impulses carried by the vagus nerves and also by the activity of the pneumotaxic centre. Either of these two influences, pneumotaxic centre or the vagii, seems to be adequate to keep apneusis in check. The existence and functions of APN are based on following experimental observations (Fig. 5.6-5):

- Sectioning of brain between the medulla and pons (level 1 in Fig. 5.6-5) leaves the basic rhythm intact indicating thereby that medullary respiratory centres are working normally.
- Sectioning the brain at mid-pontine level, i.e. between PNC and APN, (i.e. at level 2 in Fig. 5.6-5) along with bilateral vagotomy leads to prolonged periods of inspiration, i.e. apneusis or apneustic breathing. This indicates that removal of two inhibitor influences (PNC and vagii) on the APN allows APN to exert its influence on the medullary centres producing apneusis.
- Sectioning the brain rostral to the pons (level 3 in Fig. 5.6-5) leaves the respiratory rhythm intact even if combined with bilateral vagotomy. This indicates that mere presence of check by pneumotaxic centre is sufficient to control the apneustic effect of APN on the medullary centres.

2. Pneumotaxic centre

The pneumotaxic centres are located bilaterally in the upper pons.

Functions. As described above, the PNC inhibits the APN. Therefore, stimulation of PNC shortens inspiration, leading to shallow and more rapid respiratory pattern.

Thus, though rhythm of respiration resides in the DRG neurons in medulla, PNC and APN control these neurons to regulate the depth and rate of respiration.

III. RETICULAR ACTIVATING SYSTEM

The RAS stimulates the respiratory centres to increase the respiratory drive. During sleep, RAS activity diminishes, decreasing respiratory drive, which diminishes alveolar ventilation and results in a slight elevation of arterial CO_2 tension.

AFFERENT IMPULSES TO RESPIRATORY CENTRES

The respiratory centres generate the respiratory rhythm and execute their effects through the efferent nerves supplying the respiratory muscles. The activity of respiratory centres in turn is influenced by the afferent impulses from the lungs and various other parts of the body.

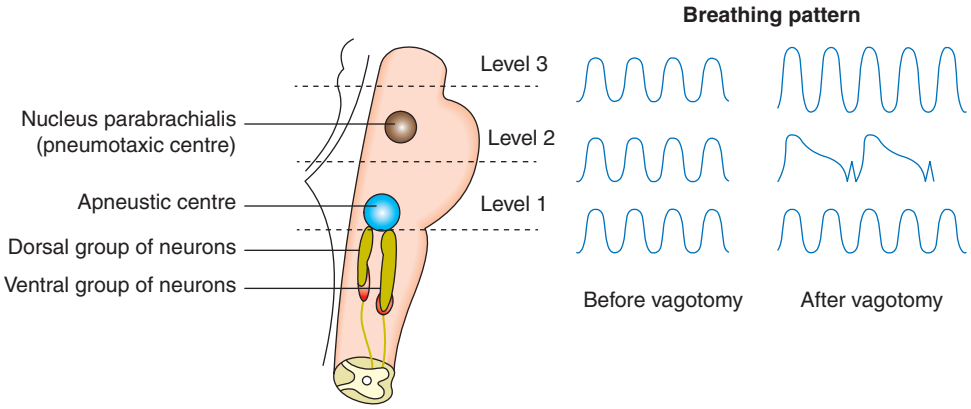


Fig. 5.6-5 Sectioning of brain at different levels to demonstrate the activity of different respiratory centres.

Various afferent impulses to the respiratory centres can be grouped as (Fig. 5.6-6):

- Afferent impulses from higher centres.
- Afferent impulses from non-chemical receptors, which may constitute *non-chemical regulation of respiration.*
- Afferent impulses from the chemical receptors, which constitute the 'chemical regulation of respiration' and hence has been described separately.

AFFERENT IMPULSES FROM HIGHER CENTRES

The afferent impulses from the higher centres which influence the involuntary activity of respiratory centres, mainly include voluntary control system and limbic control system.

1. Voluntary control system

As described above, normally, breathing is an involuntary effort and goes on automatically. However, the respiratory muscles are typical skeletal muscle and can also be controlled voluntarily. The voluntary control of respiration is mediated by a pathway which originates from the neocortex, bypasses the medullary respiratory centres to project directly on the spinal respiratory neurons. The voluntary control of breathing is exercised during activities like talking, singing, swimming and breath-holding, etc.

Breath holding or voluntary apnoea. Breathing can be stopped voluntarily for about 50–60 s (breath-holding time). But, after this time the chemical drive overrides the voluntary inhibition and the person has uncontrollable desire to breathe and ultimately breathing is resumed involuntarily. That is why it is impossible to commit suicide by holding the breath voluntarily. The decrease in arterial pO_2 and increase in the arterial pCO_2 seem to be the chief causes for the end of breath-holding.

Breath-holding can be prolonged by 15–20 seconds by an initial hyperventilation which lowers the arterial pCO_2. As a result it takes longer duration of breath-holding to increase the arterial pCO_2 to the critical level. Breath-holding may also be increased by prior inhalation of pure O_2. Some mechanical or reflex factors originating from the chest wall also seem to be involved in limiting the duration of breath-holding.

Voluntary hyperventilation, i.e. voluntary overbreathing can be done for sometime only similar to breath-holding.

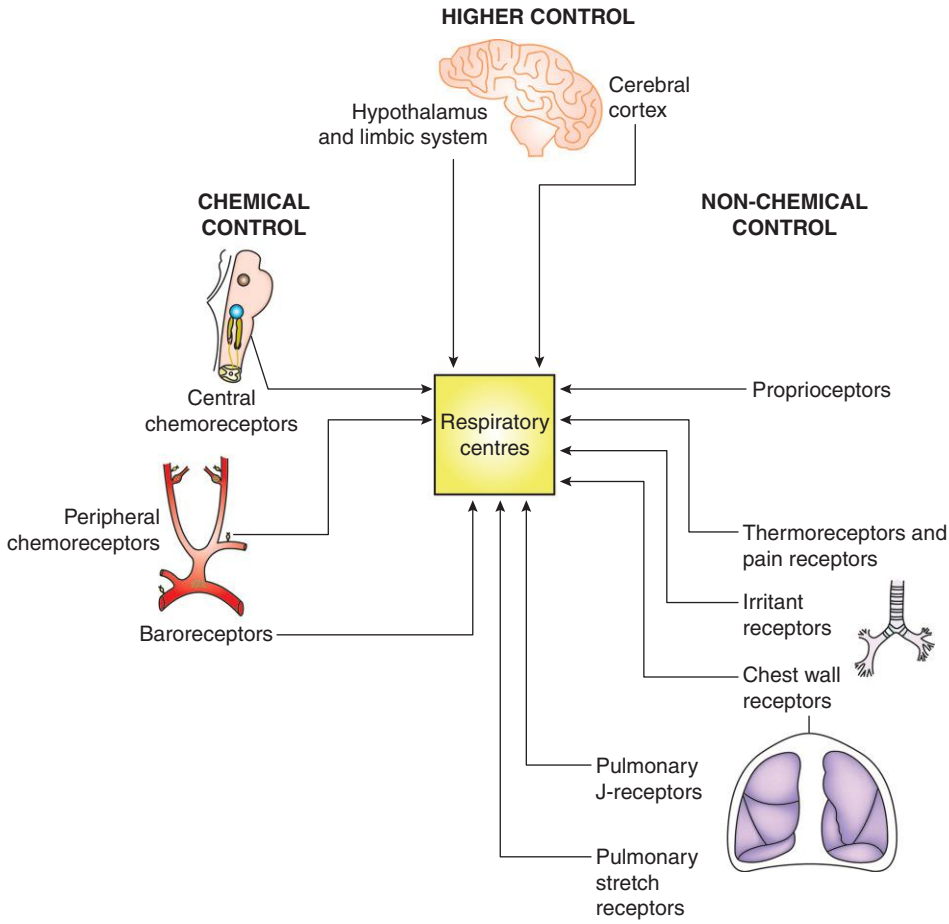


Fig. 5.6-6 Afferent impulses to respiratory centres.

Effects of hyperventilation are discussed in chemical control of respiration (see page 346).

2. Limbic control system

Pain and emotional stimuli influence the rate and depth of breathing. It indicates the presence of afferents from the limbic system to the pontomedullary respiratory neurons. Experimentally also marked changes in respiration are observed on electrical stimulation of various regions of hypothalamus. The influence of hypothalamus and the other parts of the limbic system on respiration is only to be expected in view of respiratory changes being a part of emotional expression.

IMPORTANT NOTE

Changes in the breathing pattern are the basis for part of polygraph test used as a lie detector.

APPLIED ASPECTS

As respiration has two separate controls, voluntary and automatic, sometimes automatic control is disrupted whereas voluntary control remains intact. Clinically, this condition is known as *Ondine curse*. In this state person would stay alive only if he is awake and remembers to breathe. Ondine was a water nymph cursed by the king and all his automatic functions were withdrawn. Due to exhaustion he fell asleep and died because of stoppage of breathing. This condition usually occurs in the patients suffering with bulbar poliomyelitis or conditions which compress the medulla.

AFFERENT IMPULSES FROM NON-CHEMICAL RECEPTORS

Afferent impulses from the receptors other than the chemoreceptors, i.e. from non-chemical receptors include the following:

1. Afferent impulses from pulmonary stretch receptors (Hering–Breuer reflex)

Hering–Breuer reflex is one of the first examples of negative feedback. In 1868, Hering and Breuer found that lung inflation inhibits output of the phrenic motor neurons, thereby protecting lung from overinflation.

The Hering–Breuer inspiratory inhibitory reflex is initiated when the stretch receptors located in the smooth muscles of the bronchi and bronchioles are stimulated by inflation of the lungs. The afferents of Hering–Breuer reflex are carried through *vagii* to the pontomedullary respiratory centres to inhibit respiration. This reflex has an important role in controlling tidal volume during eupnoea in human infants. In adults this reflex is weakest, therefore, it does not play any regulatory role in tidal respiration. The reflex is initiated only when the tidal volume is more than 1–1.5 L, thus, the reflex tends to limit the tidal volume.

2. Afferent impulses from J-receptors

Afferent impulses from J-receptors constitute the J-reflex. The J-receptors were discovered by an Indian physiologist A. S. Paintal in 1954. The name J-receptors (juxtapulmonary capillary receptors) was given to them because of their location very close to the pulmonary capillaries (Fig. 5.6-7). Important features of J-receptors are:

- J-receptors are basically unmyelinated vagal afferent nerve endings (type C fibres).
- These receptors are primarily sensitive to increase in the content of interstitial fluid between the capillary endothelium and alveolar epithelium, therefore they are stimulated in conditions like pulmonary congestion, pulmonary oedema, pneumonia, hyperinflation of lungs and microembolism in pulmonary capillaries.
- The J-reflex response is characterized by apnoea followed by tachypnoea (rapid and shallow breathing), bradycardia, hypotension and weakness of skeletal muscles.
- Physiological role of J-receptors has been postulated during exercise especially at high altitude when some fluid is entrapped in the alveolar interstitial space which stimulates the J-receptors producing dyspnoea and reduction of the skeletal muscle tone. This effect would discourage exercise, thereby taking away the trigger for pulmonary congestion.

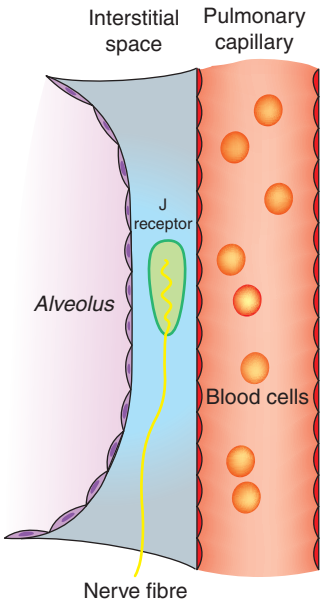


Fig. 5.6-7 Location of J-receptors.

3. Afferent impulses from the irritant receptors in the respiratory tract

Irritant receptors are located below the mucosa of whole respiratory tract. These are stimulated by a variety of chemical stimuli. These agents include serotonin, prostaglandins, bradykinin, ammonia, smoke, noxious gases, particulate matter in the inspired air and in a number of other conditions. An important function of these receptors may be to detect pathophysiological processes in the airway, such as chemical irritation, inflammation, congestion, etc. These receptors also detect histamine which produces bronchoconstriction in asthma. These receptors initiate following reflexes:

(i) Cough reflex. This is a protective reflex caused by stimulation of irritant receptors in the pharynx, larynx, trachea and bronchi (conducting zone of respiratory tract). Cough begins with a deep inspiration followed by a forced expiration with closed glottis. So, intrapleural pressure rises above 100 mm Hg. The glottis is then suddenly opened producing an explosive outflow of air. The velocity of the air flow may reach 960 km/h. By this endeavour the irritants may be expelled out of the respiratory tract.

(ii) Sneezing reflex is also a protective reflex produced on stimulation of irritant receptors of the nasal mucosa. The sneezing begins with deep inspiration followed by forceful expiration with opened glottis (in cough reflex where glottis is closed).

(iii) Hering–Breuer deflation reflex is produced on stimulation of irritant receptors located in the bronchial epithelium due to distortion of bronchial epithelium caused by large deflations of the lungs as seen in pneumothorax and lung collapses (atelectasis). This reflex may also be responsible for the *sighs* or *yawning*. The reflex helps in opening up the collapsed alveoli again.

(iv) Deglutition reflex refers to a temporary apnoea produced during pharyngeal phase of swallowing of food. It is a protective reflex which prevents the entry of food particles into the respiratory tract (see page 350).

4. Afferent impulses from proprioceptors

Proprioceptors are the receptors present in the muscles, tendons and joints and are stimulated during change in the position of different parts of the body.

IMPORTANT NOTE

- This reflex helps in increasing ventilation during exercise.
- The paediatricians employ this reflex for initiating first breath in the newborn by slapping it.

5. Afferent impulses from chest wall stretch receptors

Chest wall stretch receptors are nothing but the *muscle spindles* present in the intercostal muscles. Stretching of the intercostal muscles produce a stretch reflex due to the stimulation of muscle spindles that is characterized by contraction of intercostal muscles. The muscle spindles present in the respiratory muscles help to co-ordinate breathing during change in posture or during speech. They play a special role in maintaining normal tidal volume when breathing is impeded by an increase in airway resistance or a decrease in pulmonary compliance.

When the mechanical load on the respiratory system is increased, intercostal muscles are stretched and their muscle spindles are stimulated leading to increased strength of contraction of the intercostal muscles. It has been observed that increase in the intercostal nerve afferent activity leads to contraction of neighbouring intercostal muscles.

6. Afferent impulses from baroreceptors

Baroreceptors or pressure receptors located in the carotid sinus and aortic arch (see details on page 256) are stimulated by an increase in the arterial blood pressure. Though they play a primary role in regulation of blood pressure, but the impulses do travel to respiratory centres and cause inhibition of respiration; in physiological conditions the baroreceptors play an insignificant role in regulation of respiration. The *adrenaline apnoea* observed on injection of high doses of adrenaline causes a large rise in the arterial pressure which in turn inhibits respiration by afferent impulses from the baroreceptors to the respiratory centres.

7. Afferent impulses from thermoreceptors

Thermoreceptors are those receptors which are stimulated by a change in the body temperature. When warm receptors are stimulated, the impulses are conveyed to cerebral cortex via somatic afferent nerves. Cerebral cortex in turn stimulates the respiratory centres to produce hyperventilation.

Respiration helps to maintain body temperature, as some amount of heat is lost in the expired air. In dogs, panting is one of the major mechanisms of thermoregulation.

CHEMICAL REGULATION OF RESPIRATION

The chemical factors regulating respiration are pCO_2, pO_2 and pH of blood. These factors influence respiration in such a way that their own blood levels are maintained constant. The chemical mechanism of regulation operates through the chemoreceptors.

CHEMORECEPTORS

Chemoreceptors are the sensory nerve endings, which are highly sensitive to changes in pCO_2, pO_2 and pH of blood. These are of three types:

- Peripheral chemoreceptors
- Central chemoreceptors
- Pulmonary and myocardial chemoreceptors

PERIPHERAL CHEMORECEPTORS

Location. Peripheral chemoreceptors include the carotid and aortic bodies (Fig. 5.6-8).

- *Carotid body* is located on either side near the bifurcation of common carotid artery.
- *Aortic bodies,* two or more in number, are located near the arch of aorta.

Structure. Each carotid and aortic body consists of:

Capsule, surrounding each carotid and aortic body, is very thin.

Sinusoidal large capillaries present below the capsule surround the main mass of each body.

Epithelial cells. The main mass of the body consists of islands of epithelial cells, which are of two types: type I and type II (Fig. 5.6-9):

- *Type I or glomus cells.* These cells have dense-core granules containing catecholamines (probably dopamine). Unmyelinated nerve endings are closely applied to these cells; these nerve endings are cup shaped and have dopamine receptors (D_2) on them. When exposed to hypoxia, the type 1 cells release catecholamine which stimulates the D_2 receptors.
- *Type II cells,* which are probably glial cells, are also closely applied to the type 1 cells.
- *Nerve fibres.* Outside the capsule of each body, the nerve fibres acquire myelin sheath, they are only 2–5 µm in diameter and conduct at relatively low rate of 7–12 m/s. Afferent fibres from the carotid body join the sinus nerve, a branch of glossopharyngeal (IX) nerve and ultimately ascends to the medulla. Those from the aortic body join the aortic nerve branch of vagus (Xth cranial) nerve and ascend to the medulla.

Blood flow to each carotid and aortic body is highest in the body (2000 mL/100 g/min). Therefore, the O_2 needs of these cells can be met largely by dissolved O_2 only.

Functions. The peripheral chemoreceptors respond to lowered pO_2, increased pCO_2 and increased H^+ concentration in the arterial blood. The afferent impulses from the chemoreceptors stimulate the DRG neurons, which lead to an increased rate and depth of respiration called *hyperventilation.* Salient points of their functions are:

- The peripheral chemoreceptors are the only sites that detect changes in pO_2.
- Carotid bodies are seven times more effective than the aortic bodies in stimulating respiration.
- Carotid bodies increase both rate and depth of respiration, while the aortic bodies increase only the frequency of respiration with small increase in the ventilation.

Mechanism of chemoreceptors stimulation by hypoxia and oxygen transduction in glomus cells. The peripheral chemoreceptors are stimulated by the release of neurotransmitter by the glomus cells. Oxygen transduction is the process by which changes in the arterial pO_2 results in

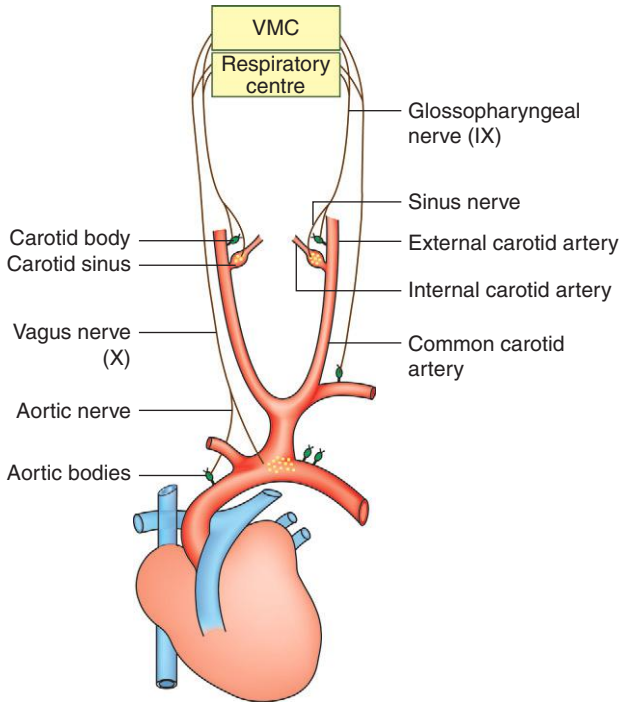


Fig. 5.6-8 Location of carotid and aortic bodies (peripheral chemoreceptors).

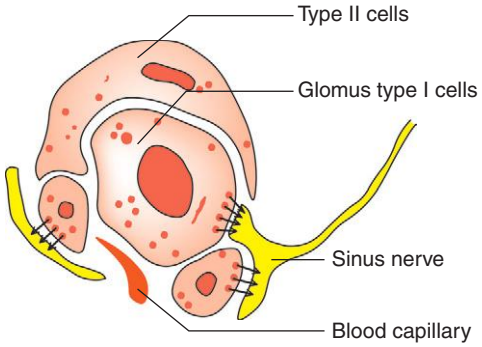


Fig. 5.6-9 Histological structure of carotid body.

proportionate changes in the frequency of action potential discharge. The sequence of events is:

- Hypoxia leads to a decrease in activity of oxygen-sensitive K^+ channels present in the cell membrane of glomus cells leading to decrease in the K^+ efflux depending upon the level of pO_2.
- Thus, the glomus cells get depolarized in proportion to the fall in arterial pO_2.
- Depolarization of the glomus cells opens up the L-type Ca^{2+} channels in the glomus cell membrane leading to an increase in the Ca^{2+} influx.
- The Ca^{2+} influx triggers the release of neurotransmitter which stimulates the afferent nerve endings.
- *Drugs,* such as cyanide, nicotine and lobeline, prevent O_2 utilization at the tissue level and stimulate peripheral chemoreceptors.

Factors affecting peripheral chemoreceptors stimulation

1. ***O_2 tension versus O_2 content.*** The peripheral chemoreceptors monitor the dissolved O_2, i.e. pO_2, rather than its total content. They are stimulated when pO_2 falls below 60 mm Hg. Therefore, they respond to various types of hypoxia differently as:
   - *Hypoxic hypoxia* in which arterial pO_2 is reduced stimulates peripheral chemoreceptors.
   - *Histotoxic hypoxia* in which there is reduced utilization of O_2 by the tissue cells including glomus cells also stimulates chemoreceptors.
   - *Anaemic hypoxia,* methaemoglobinaemia, or carbon monoxide poisoning do not stimulate the peripheral chemoreceptors, because in these conditions, though the total content of O_2 may be low, but the O_2 tension, which is determined by the amount of dissolved O_2 remains normal.
   - *Vascular stasis,* in which the amount of O_2 delivered to receptors per unit of time is decreased leads to chemoreceptor stimulation.
2. ***Elevated pCO_2.*** Elevated pCO_2 (by 10 mm Hg) also stimulates the peripheral chemoreceptors, but the major effect of CO_2 is on the central chemoreceptors.
3. ***H^+ concentration*** when increased in the blood (decreased pH by 0.1 unit) stimulates the peripheral chemoreceptors.
4. ***Increase in plasma K^+ levels*** may stimulate the peripheral chemoreceptors even in the absence of hypoxia. Increase in plasma K^+ levels during exercise contributes to exercise-induced hyperventilation.
5. ***Asphyxia,*** i.e. combination of O_2 lack plus CO_2 excess in the blood stimulates the peripheral chemoreceptors.

Effects of stimulation of peripheral chemoreceptors

- They regulate the respiration from breath to breath and their stimulation increases the rate and depth of respiration.
- They also cause an increase in the blood pressure and tachycardia.
- About 15–20% of resting respiratory drive is due to the stimulatory effect of CO_2 on the peripheral chemoreceptors.

CENTRAL CHEMORECEPTORS

Location. Central chemoreceptors are the cells (neurons) that lie just beneath the ventral surface of the medulla oblongata and are therefore also called medullary receptors (Fig. 5.6-10).

Innervation. The neurons forming central chemoreceptors project directly over to the respiratory centres which are located slightly deeper to the central chemoreceptors.

Stimulation characteristics of central chemoreceptors are:

- They respond to H^+ concentration in the surrounding interstitial fluid and cerebrospinal fluid (CSF).
- The magnitude of stimulation is directly proportional to the local H^+ concentration, which in turn parallels arterial pCO_2.
- *Mechanism by which an increase in CO_2* concentration affects central chemoreceptors. CO_2 readily crosses the

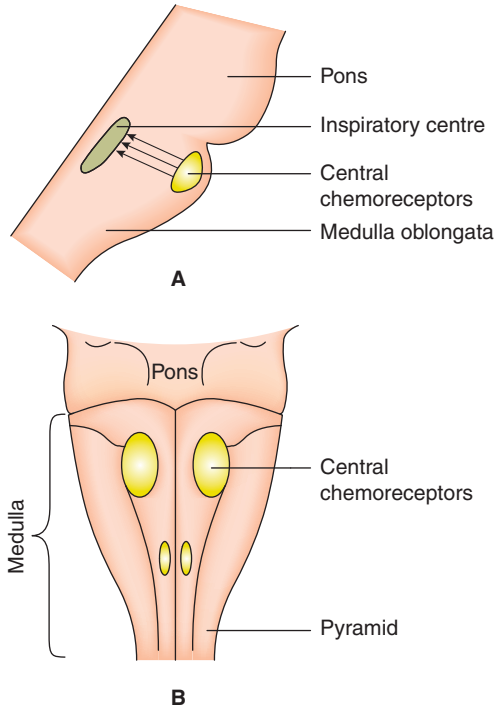


Fig. 5.6-10 Location of central chemoreceptors in medulla: A, lateral view and B, front view.

blood–brain barrier, because it is a small, very soluble, uncharged molecule. In the CSF, CO_2 combines with water to form H_2CO_3 which dissociates into H^+ and HCO_3^- ions. The increase in H^+ concentration of CSF and interstitial fluid stimulates the central chemoreceptors, whereas a decrease in the H^+ concentration inhibits respiration. It is important to note that the blood–brain barrier does not allow the charged ions (e.g. H^+, HCO_3^- etc.) to cross through readily. Because of this reason if the arterial pCO_2 is kept constant experimentally a decrease in the arterial pH (raised H^+ concentration) fails to stimulate central chemoreceptors.

- Central chemoreceptors are not stimulated by hypoxia, rather like any other cell, they are depressed by hypoxia.
- Central chemoreceptors are also inhibited by anaesthesia, cyanide and during sleep.

Effects of stimulation of central chemoreceptors are:

- The central chemoreceptors regulate the respiration from minute-to-minute. Their stimulation leads to an increase in rate and depth of respiration.
- It is important to note that about 80–85% of the resting respiratory drive is due to the stimulatory effect of CO_2 on the central chemoreceptors. While peripheral chemoreceptors provide only 15–20% of initial drive to increase respiration.

PULMONARY AND MYOCARDIAL CHEMORECEPTORS

Location. Pulmonary and myocardial chemoreceptors are located in the pulmonary and coronary blood vessels, respectively.

Innervation. These are innervated by vagus (Xth cranial) nerve.

Stimulation characteristics and effects of these receptors are:

- *Pulmonary chemoreceptors* are stimulated by the injection of veratridine or nicotine into the pulmonary circulation and produce the so-called *pulmonary chemoreceptor reflex,* which is characterized by bradycardia, hypotension and apnoea followed by tachypnoea (rapid shallow breathing). Physiological role of this reflex is not established. It occurs in the pathological states like pulmonary congestion or embolism.
- *Myocardial chemoreceptors* are similarly stimulated when these agents are injected into coronaries supplying the left ventricle and produce the so-called *coronary chemoreflex* or *Bezold–Jarisch reflex* having features similar to the pulmonary chemoreflex. *Physiological role* of this reflex is not established. It is known to occur after myocardial infarction.

EFFECT OF pO_2, pCO_2 AND H^+ CONCENTRATION ON RESPIRATION

EFFECT OF HYPOXIA ON RESPIRATION

The normal arterial pO_2 is 100 mm Hg, which may fall in many conditions (see page 353), producing the so-called hypoxic hypoxia. A decrease in arterial pO_2 is the most potent stimulus for the peripheral chemoreceptors, consequently the rate of discharge in the peripheral chemoreceptors begins to increase.

When the arterial pO_2 levels falls to between 100 and 60 mm Hg, not much effect is produced on ventilation. However, a marked increase in the pulmonary ventilation occurs when the pO_2 falls below 60 mm Hg (Fig. 5.6-11).

At pO_2 levels from 100 to 60 mm Hg, pulmonary ventilation does not increase significantly because of following two reasons:

- *Breaking effect of CO_2.* When decrease in pO_2 of the arterial blood stimulates ventilation, increased ventilation causes washing out of CO_2. This leads to decrease in pCO_2 of blood which inhibits the respiration through its effect on the central chemoreceptors. It opposes and neutralizes the effect of decreased pO_2 and thus there is no marked effect on ventilation. This phenomenon is called breaking effect of CO_2.
- Due to hypoxia, the amount of deoxyhaemoglobin is increased which is a weaker acid as compared to oxyhaemoglobin. This results in mild decrease in H^+ concentration of blood which tends to nullify the hypoxic drive on pulmonary ventilation.
- Hypoxia stimulating respiration through peripheral chemoreceptors can be proved experimentally by denervating them. Under such circumstances, hypoxia cannot increase pulmonary ventilation, rather it causes direct depression of the respiratory centre.

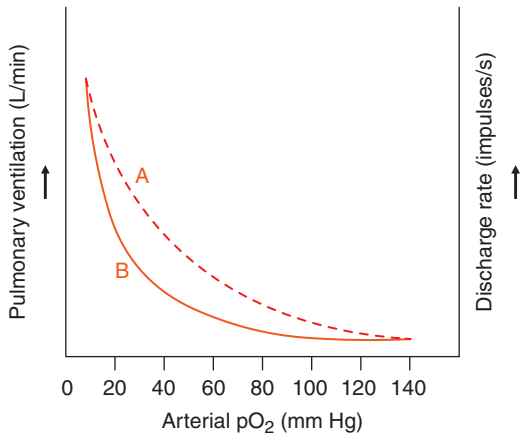


Fig. 5.6-11 Relationship of arterial pO_2 with sinus nerve discharge rate (A) and pulmonary ventilation (B).

EFFECT OF HYPERCAPNIA ON RESPIRATION

Normal arterial pCO_2 is 40 mm Hg, which is kept constant by chemical regulation of respiration. ***Hypercapnia,*** i.e. rise in pCO_2 rarely occurs due to an increase in CO_2 production. Clinically, it may occur in restrictive lung disorders (see page 304).

An increase in arterial pCO_2 causes a prompt increase in the pulmonary ventilation resulting in CO_2 washout and a near restoration of arterial pCO_2 to normal level (40 mm Hg). There exists a linear relation between increase in arterial pCO_2 and increase in pulmonary ventilation.

CO_2 increases pulmonary ventilation mainly by stimulating the central chemoreceptors. This can be demonstrated experimentally by removing the peripheral chemoreceptors and then making the person to breath from a bag of air with different concentration of CO_2. The same effect is obtained as above (Fig. 5.6-12).

CO_2 is capable of increasing the pulmonary ventilation by stimulating the peripheral chemoreceptors as well. When central chemoreceptors are depressed by anaesthesia, then CO_2 increases respiration through stimulation of peripheral chemoreceptors.

CO_2 acts as a main regulator of respiration because of following facts:

- It has a direct effect on respiratory centre through the central chemoreceptors,
- It can cross the blood–brain or blood–CSF barrier easily, therefore CO_2 concentration in the CSF and in the interstitial fluid of brain increases soon after the increase in concentration of CO_2 in the blood.
- CO_2 has a very strong breaking effect on the action of either decreased pO_2 or pH,
- pO_2 or pH does not have a very strong breaking effect on the action of increased CO_2 on ventilation (Fig. 5.6-12).

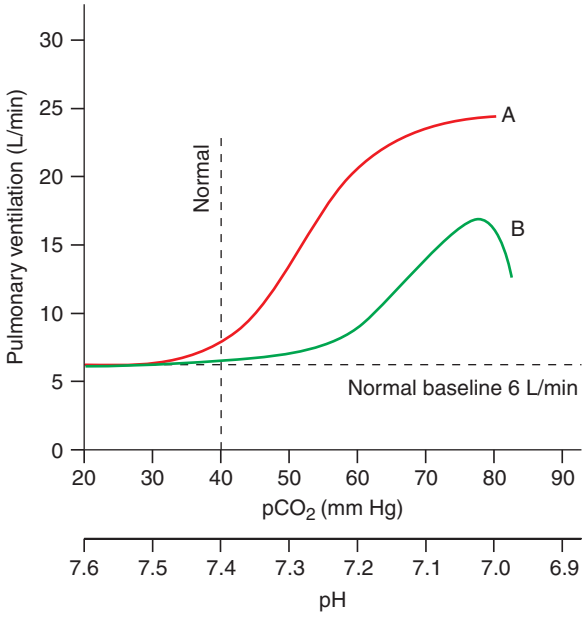


Fig. 5.6-12 Effect of increase in arterial pCO_2 (A), and decreased pH; (B), on pulmonary ventilation.

Carbon dioxide narcosis. It develops when arterial pCO_2 increases above 50 mm Hg. Accumulation of such a large amount of CO_2 (hypercapnia) in the body depresses the CNS, including respiratory centres producing headache, confusion, convulsions and finally coma and death may occur.

- *Causes.* Carbon dioxide narcosis may occur in patients with prolonged severe emphysema or due to accidental inhalation of CO_2 (in breweries, refrigeration plants, etc). Experimentally, it can be produced by making the person to inhale the air containing more than 7% CO_2. When the inspired air pCO_2 approaches close to the alveolar pCO_2, as a result elimination of CO_2 becomes difficult which causes alveolar and arterial pCO_2 to rise abruptly in spite of the hyperventilation.

APPLIED ASPECTS

Whenever CO_2 is to be used to stimulate respiration in a comatosed patient with respiratory depression it is always advisable to estimate the CO_2 content of the blood to avoid occurrence of death from carbon dioxide narcosis.

EFFECT OF ARTERIAL pH ON RESPIRATION

(i) Increased H^+ concentration (metabolic acidosis) produces prolonged respiratory centre stimulation via peripheral chemoreceptors, leading to a decrease in the arterial pCO_2 by elimination of larger amounts of CO_2 producing compensatory fall in the blood H^+ concentration. The related aspects of renal correction of acid–base balance are discussed in Chapter 6.5, page 428.

Causes of metabolic acidosis, i.e. decrease in HCO_3^- concentration in blood secondary to increase in H^+ concentration of blood are:

- Diabetic ketoacidosis. Hyperventilation occurring in this condition is called *Kussmaul breathing.*
- Renal failure (when kidney fails to excrete their normal quota of H^+).
- Due to accumulation of lactic acid in severe muscular exercise.
- Ketoacidosis in starvation.
- Infantile diarrhoea associated with loss of $NaHCO_3$.

(ii) Decreased H^+ concentration (metabolic alkalosis) depresses respiratory centre via peripheral chemoreceptors,

leading to retention of CO_2 and an increase in the arterial pCO_2. The secondary changes in the arterial pCO_2 compensate for the primary metabolic defects and help to restore H^+ concentration of blood.

Common causes of metabolic alkalosis, i.e. increase in HCO_3^- concentration in blood secondary to decreased H^+ concentration in blood: excessive vomiting with loss of HCl from the body.

Respiratory acidosis and alkalosis. It may be added here that primary changes in the pulmonary ventilation also affect the pH of blood causing respiratory acidosis or alkalosis.

(iii) Primary pulmonary hypoventilation may lead to elevation of arterial pCO_2 (hypercapnia) producing the so-called *respiratory acidosis.*

(iv) Primary pulmonary hyperventilation may cause a decrease in the arterial pCO_2 producing the so-called *respiratory alkalosis.*

INTERACTION OF pO_2, pCO_2 AND pH IN REGULATION OF RESPIRATION

In the above discussion we have seen that each of hypoxia, increased pCO_2 and acidosis individually cause an increase in the respiration. In many physiological or clinical situations more than one factor may be present. Their interaction is summarized here:

1. Interaction of pCO_2 and pO_2

Hypoxia sensitizes the respiratory mechanism to excess of CO_2 or H^+ concentration, therefore increased pCO_2 and H^+ concentration produce a much greater effect.

When pCO_2 is held constant at a level 2–3 mm Hg above normal (i.e. in the presence of hypercapnic drive), there is an inverse relationship between ventilation and alveolar pO_2, even in the 90–110 mm Hg range (Fig. 5.6-13).

When pCO_2 is held constant at a level 2–3 mm Hg below normal, (i.e. in the absence of CO_2 related drive) a fall in pO_2 level between 110 and 60 mm Hg does not produce any effect on ventilation. However, a marked increase in pulmonary ventilation occurs when pO_2 falls below 60 mm Hg (Fig. 5.6-13).

2. Interaction of pH and CO_2 response

The stimulatory effect of H^+ concentration and CO_2 on respiration is *additive,* i.e. a fall in pH (acidosis) shifts the CO_2 response curve to left without change in slope. In other words, the same amount of respiratory stimulation is produced by lower arterial pCO_2 levels.

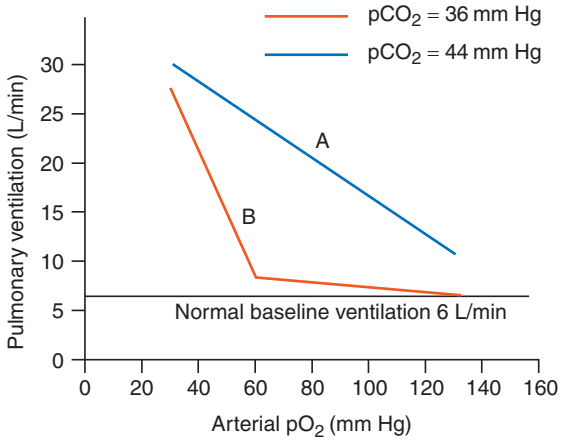


Fig. 5.6-13 Effect of hypoxia (pO_2) on pulmonary ventilation with arterial pCO_2 being kept 2–3 mm above normal (A) and 2–3 mm below normal (B).

3. Interaction of CO_2 and body temperature

The effect of CO_2 on respiration increases with an increase in body temperature.

SOME OTHER ASPECTS RELATED TO CHEMICAL REGULATION OF RESPIRATION

EFFECTS OF HYPERVENTILATION

Effect of short lasting severe hyperventilation

The voluntary hyperventilation may show following effects (Fig. 5.6-14):

(i) Effects on respiration. After a period of hyperventilation, any of the following pattern of respiration may be seen for a small period before normal respiration is restored:

- *Hypoventilation* for a prolonged period is seen in most of the individuals.
- *Apnoea,* i.e. complete cessation of breathing for 1–2 min may occur in some individuals.
- *Periodic breathing (Cheyne–Stokes breathing),* i.e. alternate phases of apnoea and breathing may occur for sometime in a few individuals.

(ii) Effects on arterial pCO_2 and pO_2 and their correlation with effect on respiration (Fig. 5.6-14)

- Arterial pO_2 may go as high as 150 mm Hg and pCO_2 as low as 15 mm Hg after a period of voluntary hyperventilation.
- Apnoea occurring in some individuals at the end of hyperventilation seems to be related to lack of CO_2.
- During phase of apnoea, due to metabolism of body, there occurs a decline in the arterial pO_2 and an increase

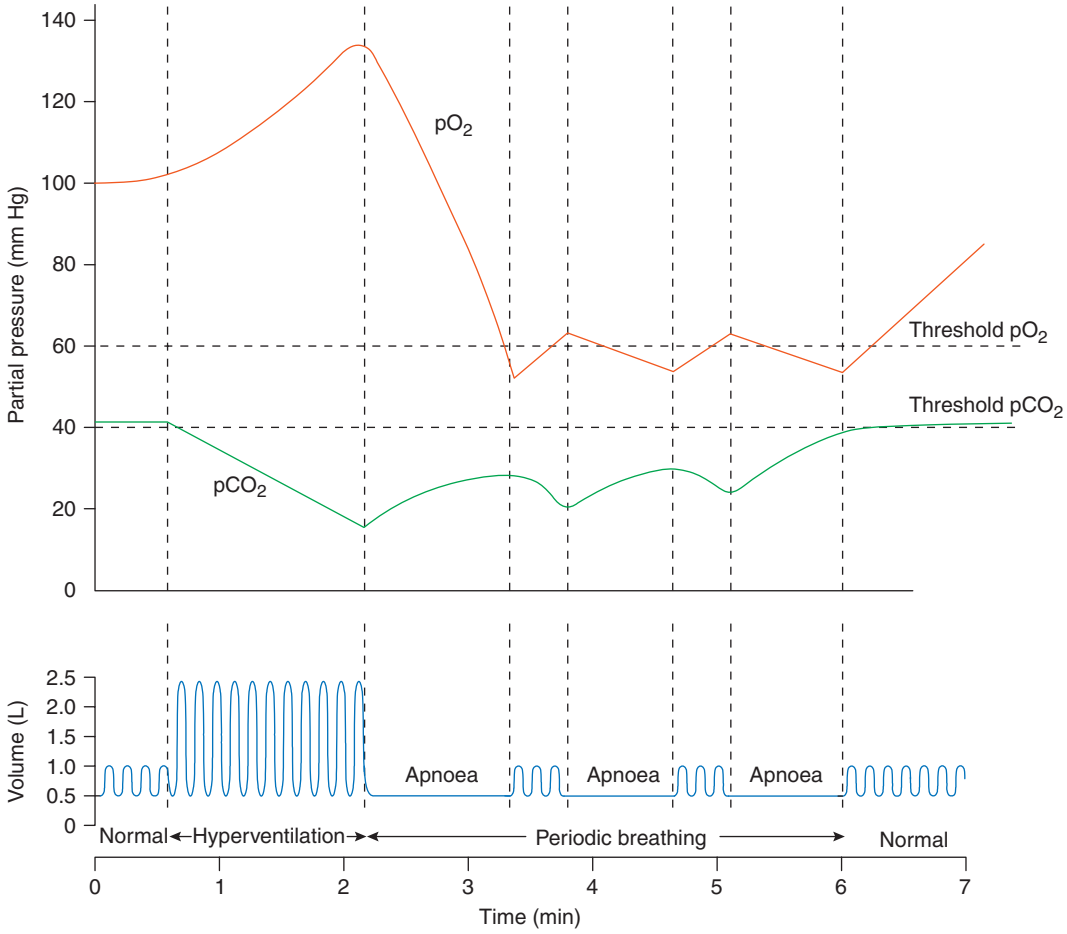


Fig. 5.6-14 Effect of hyperventilation on arterial pO_2, pCO_2 and respiration. Note the correlation between pCO_2, pO_2 and periods of hyperventilation; apnoea and periodic breathing and normal breathing.

in pCO_2. Depending upon the interaction between levels of pO_2 and pCO_2 attained, following effects can occur on respiration:

- If the arterial pCO_2 is reached at threshold level (40 mm Hg), then a normal breathing is resumed
- If the hypoxic stimulus (decreased pO_2) becomes strong before the pCO_2 reaches a threshold level, then *periodic breathing* (Cheyne–Stokes breathing) may result, i.e. a few breaths eliminate hypoxic drive, then breathing stops and restarts when again hypoxic drive stimulates it. Such cycles are repeated till pCO_2 reaches threshold level to normalize breathing.

Effects of prolonged moderate hyperventilation

Prolonged moderate hyperventilation, i.e. two to five fold increase in the pulmonary ventilation, may occur under following circumstances:

- In residents of high altitude.
- In clinically hypoxic patients due to some pulmonary disease.
- It may also be maintained even voluntarily.

The prolonged moderate hyperventilation is associated with a decrease in alveolar and arterial pCO_2. The low arterial pCO_2 which lasts for many days may result in the following complication in the body:

Respiratory alkalosis. Low levels of pCO_2, reduce the formation of H^+ and HCO_3^- in the blood causing an increase in its pH to 7.55 or even 7.6, a condition called respiratory alkalosis.

Renal changes. The respiratory alkalosis produced due to decreased arterial pCO_2 interferes with H^+ secretory mechanism in the kidney. There occurs failure of proximal tubular reabsorption of HCO_3^-, which results in excretion of alkaline urine containing HCO_3^-.

Neurological changes may occur due to:

- *Respiratory alkalosis-induced hypocalcaemic tetany,* which include numbness and tingling in the extremities and carpopedal spasm and
- *Low arterial pCO_2* induced constriction of cerebral vessels may produce symptoms like dizziness and light headache. The consciousness may be dulled or even lost.

Cardiovascular changes may occur due to moderately increased cardiac output owing to muscular efforts involved in the production of hyperventilation.

EFFECT OF SLEEP ON RESPIRATION

It has been reported that due to inhibition of central chemoreceptors during sleep the sensitivity of respiratory centre neurons to arterial pCO_2 is decreased. It may cause following effects:

- *Apnoea for brief period (10s duration)* is of common occurrence during sleep in normal individuals.
- *Sleep apnoea syndrome* is a serious clinical problem which may occur in some individuals.

Causes. The disorders in which ventilation ceases are:

- Deeper stage of sleep [rapid eye movement (REM) sleep].
- Due to lack of central automatic drive of respiration (Ondine curse).
- Collapse of airway with sleep (obstructive sleep apnoea).
- Decrease airway tone during sleep in obese persons.

Sleep apnoea syndrome may occur in two forms:

- Obstructive sleep apnoea occurs when inspiration is prevented by a transient blockage of the airway due to the collapse of hypopharynx as a result of loss of tone of the pharyngeal muscles which prevent airflow though strong contractions of the inspiratory muscles occur. Partial airway obstruction causes snoring. The association of sleep apnoea with extreme obesity is referred to as the *pickwickian syndrome*.
- Non-obstructive (central) sleep apnoea refers to the complete stoppage of rhythmic activity from the respiratory centres. Obviously, during apnoea there is no respiratory muscle contraction. It is supposed to result from decreased chemoreceptor sensitivity to O_2 and CO_2. Central sleep apnoea has been proposed as one of the many possible causes of *sudden infant death syndrome.*

Chapter 5.7

Respiration: Applied Aspects

INTRODUCTION

Applied respiratory physiology forms a link between the basics of respiration and clinical manifestations of respiratory diseases. This chapter is concerned with some of the important applied aspects of respiration which include:

- Respiratory adjustments to stresses in health,
- Disturbances of respiration,
- Artificial respiration and
- Pulmonary function tests.

RESPIRATORY ADJUSTMENTS TO STRESSES IN HEALTH

Respiratory adjustments to stresses in health illustrate the integrated operation of the respiratory regulatory mechanisms. The stresses faced by respiration requiring adjustments in day-to-day life include:

1. Respiratory adjustments during exercise. Exercise is the most frequently faced stress in day-to-day life. Since during exercise, many complex adjustments of muscular blood flow,

metabolism, respiration, circulation and temperature are required, so they have been comprehensively discussed in Chapter 5.8 on 'Physiology of Exercise' (page 367).

2. Respiratory adjustments at high altitude. At high altitude, barometric pressure is low and so the partial pressure of O_2 is also low; however, the amount of O_2 in the atmosphere is same as it is at the sea level. When a person is exposed to high altitude particularly by a rapid ascent, the different systems of the body cannot cope with the lowered O_2 tension and the effects of hypoxia start. Respiratory adjustments are thus a part of changes in the body at high attitude, so these have been discussed comprehensively under the title *'Physiology of high altitude'* (see page 357).

3. Respiratory adjustments to high atmospheric pressure form a part of the physiological problems faced by the body while going under the sea and have been discussed comprehensively under the title *Deep sea physiology* (see page 361).

4. Respiratory adjustments on exposure to cold and heat have been discussed under the title *Effects of exposure to heat and cold on the body* (see page 959).

5. Respiratory adjustments at birth. Birth is the most traumatic event that the respiratory system must withstand during the entire life span of an individual (see page 972).

DISTURBANCES OF RESPIRATION

From the physiological viewpoint, disturbances of respiration can be discussed under the following headings:

- Abnormal respiratory patterns and
- Disturbances related to respiratory gases.

ABNORMAL RESPIRATORY PATTERNS

Eupnoea refers to the normal respiratory pattern, which implies a normal rate, rhythm and depth of respiration. Various abnormal respiratory patterns (Fig. 5.7-1) can be produced by the changes in the environment or diseases affecting the respiratory system, cardiovascular system or brain. The terms used for the altered pattern of respiration are:

- *Tachypnoea* refers to an increase in the rate of respiration.
- *Bradypnoea* means decrease in the rate of respiration.
- *Polypnoea* is used to denote the rapid but shallow breathing resembling panting in dogs. In this, the rate of respiration is increased but the force does not change significantly.
- *Apnoea* refers to the temporary cessation of breathing.
- *Hypoventilation* term is used to describe a decrease in the rate and force of respiration.
- *Hyperventilation* refers to an increase in the rate as well as force of respiration.
- *Hyperpnoea* signifies a marked increase in the pulmonary ventilation due to an increase in rate and/or force of respiration.
- *Dyspnoea.* When hyperpnoea involves four to five fold increase in the pulmonary ventilation, an unpleasant sensation or discomfort is felt. This type of respiration is called dyspnoea.

Periodic breathing refers to a respiratory pattern characterized by alternate periods of respiratory activity and apnoea. Some of the abnormal respiratory patterns are discussed in detail.

APNOEA

Apnoea refers to a temporary cessation of breathing. Depending upon the cause, apnoea may be of following types:

1. Voluntary apnoea refers to a temporary arrest of breathing due to the voluntary control of respiration. It is also called breath-holding. The breath-holding time or apnoea time during which breathing can be withheld voluntarily is about 40–60 s in a normal person, after a deep inspiration (for details see page 339).

Breaking point is the point at which breathing can no longer be voluntarily inhibited. At this point, chemical regulation overcomes the neural regulation. The breaking point is due to an increased arterial pCO_2 and a decreased pO_2.

2. Apnoea after hyperventilation occurs due to the reduced stimulation of respiratory centre owing to CO_2 wash caused by hyperventilation (for details see page 346).

3. Deglutition apnoea occurs reflexly during swallowing (about 0.5 s). During pharyngeal stage of swallowing, the fluid or food stimulates the sensory nerve endings (5th, 9th and 10th cranial nerves) around the pharynx. Nerve impulses from these irritant receptors, via the swallowing centres specifically inhibit the respiratory centre, stopping the breathing at any point of the cycle (deglutition apnoea). Simultaneously, there is closure of glottis (the opening

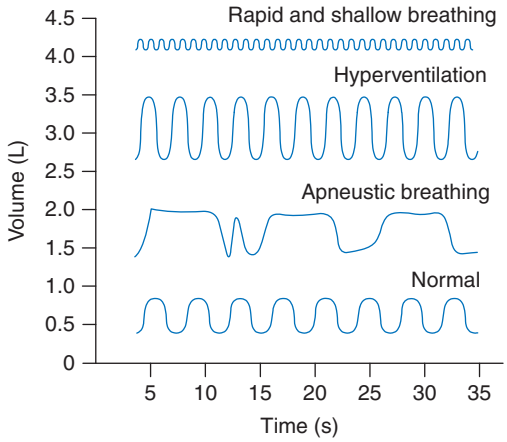


Fig. 5.7-1 Various abnormal respiratory patterns.

between vocal cords). Both these effects prevent aspiration of fluid or food into the lungs (also see page 461).

4. Breath-holding attacks are attacks of brief period of apnoea which occur in infants and young children and are generally precipitated by an emotional distress.

5. Adrenaline apnoea occurs after injection of high doses of adrenaline (see page 341).

6. Sleep apnoea refers to the cessation of breathing for a brief period (10 s) during sleep in a normal individual (see page 348).

HYPOVENTILATION

Hypoventilation is used to describe a decrease in the rate and force of respiration. Thus, in hypoventilation, the amount of air moving in and out of lungs is reduced.

Causes of hypoventilation are:
- Depression of respiratory centres by some drugs and
- Partial paralysis of respiratory muscles.

Effects. Hypoventilation leads on to hypoxia and hypercapnia (respiratory acidosis), which result in an increase in rate and force of respiration and the patient may develop dyspnoea.

HYPERVENTILATION

Hyperventilation refers to an increase in the rate as well as force of respiration. Thus, in hyperventilation, the amount of air moving in and out of lungs is increased.

Causes of hyperventilation are:
- During exercise due to stimulation of respiratory centres by increased pCO_2,
- Voluntary hyperventilation and
- Secondary to hypoxia.

Effects of hyperventilation on respiration are described on page 346.

DYSPNOEA

Dyspnoea literally means distressed breathing. Increased respiration without discomfort is called *hyperpnoea.* One is not aware of one's respiration till resting pulmonary ventilation becomes more than double. When hyperpnoea involves four to five fold increase in pulmonary ventilation, an unpleasant sensation or discomfort is felt. This type of respiration is called dyspnoea. The word 'air-hunger' is used as synonym to dyspnoea in general language.

Dyspnoea point refers to the height of hyperpnoea at which dyspnoea appears.

Predisposing factors for dyspnoea include:

1. *Low vital capacity.*
2. *Maximum ventilatory volume (MVV).* Patients with reduced MVV, (Normal value is 120 L/min) are more predisposed to get dyspnoea.
3. *Breathing reserve (BR)* is the difference between MVV and respiratory minute volume (RMV). RMV is the volume of air that is taken in or given out per minute (Normal $500 \times 12 = 6$ L/min). Individuals with increased RMV (also called pulmonary ventilation) by four to five times get dyspnoea. Individuals with less breathing reserve are more prone to get dyspnoea.

$$BR = MVV - RMV = 114 \text{ L/min}$$

Dyspnoeic index (DI) refers to the breathing reserve percentage of MVV, i.e.

- $DI = \frac{BR \times 100}{MVV} = \frac{100 \times 100}{120} = 95\%$
- Normal value of DI range from 70% to 95%
- Dyspnoea occurs when DI is <60%.

Causes of dyspnoea are:

- *Physiologically dyspnoea occurs* in severe muscular exercise.
- *Pathological causes* include:
  - *Respiratory disorders,* such as bronchial asthma, emphysema, pneumonia, pulmonary oedema and penumothorax, and
  - *Cardiac failure* (Fig. 5.7-2).
  - *Metabolic disorders* causing dyspnoea are diabetic acidosis, uraemia and increased H^+ concentration. Metabolic acidosis causes dyspnoea by increasing the pulmonary ventilation.

IMPORTANT NOTE

It is important to note that patients with cardiac failure prefer to sit rather than lie down, because in lying position pulmonary congestion is increased which causes dyspnoea. Dyspnoea occurring in lying down position is called *orthopnoea.*

PERIODIC BREATHING

Periodic breathing is characterized by the alternate periods of respiratory activity and apnoea.

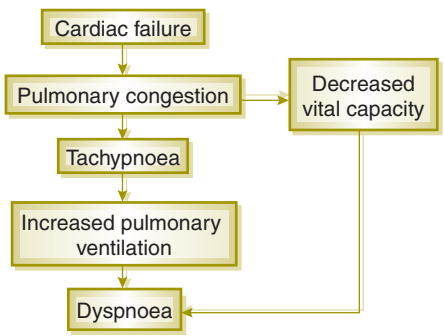


Fig. 5.7-2 Cardiac failure causing dyspnoea: mechanism.

Cheyne–Stokes respiration

The Cheyne–Stokes respiration is a periodic type of breathing in which the alternate periods of respiratory activity and apnoea occur at regular intervals and during the period of respiratory activity there is waxing and waning of tidal volume. The duration of one cycle is about 1 min. The arterial pO_2 and pCO_2 fluctuate during each cycle. The pO_2 is lowest and the pCO_2 is highest at the end of apnoea (Fig. 5.7-3).

Causes of Cheyne–Stokes respiration are:

1. Physiological causes include:
- Voluntary hyperventilation,
- High altitude and
- During sleep in some normal individuals especially infants.

2. Pathological causes are:
- Chronic heart failure,
- Brain damage,
- Uraemia and
- Poisoning by narcotics.

Mechanism of development of Cheyne–Stokes respiration in three most important conditions is described

Voluntary hyperventilation. Mechanism of development of Cheyne–Stokes breathing in hyperventilation has been described on page 347.

Heart failure. Mechanism of development of Cheyne–Stokes breathing is summarized as:

- Left ventricular failure → Pulmonary congestion → Hypoxia → Stimulation of respiratory centres → Increased ventilation → Increased alveolar pO_2 and decreased pCO_2 → Decreased arterial pCO_2.

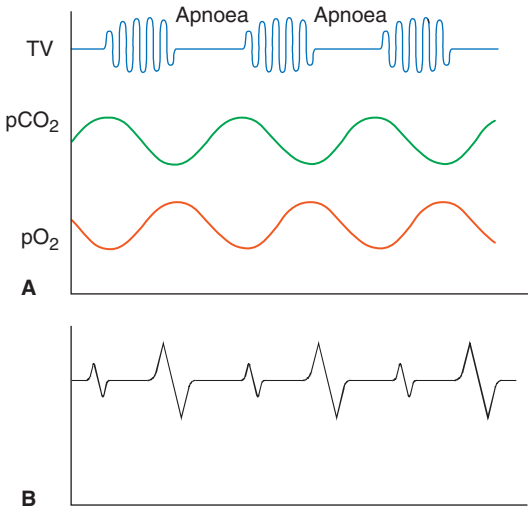


Fig. 5.7-3 Periodic breathing: A, Cheyne–Stokes breathing and B, Biot's breathing.

- As in heart failure *circulation time* is prolonged, so it takes longer than normal time for the blood with low pCO_2 to reach the brain and cause apnoea by inhibiting respiratory centre.
- Since in heart failure the pulmonary congestion is continuously present, so hypoxia is maintained and the above described cycle of apnoea followed respiratory activity that keeps on repeating till the heart failure is treated or alveolar pCO_2 comes back to normal.

Brain damage. In brain damage, when the supramedullary inhibitory pathway is damaged, the medullary (central) chemoreceptors become more sensitive to the action of CO_2 and produces Cheyne–Stokes breathing as:

Increased sensitivity of central chemoreceptors to CO_2 → Hyperventilation → CO_2 washout → Apnoea → Accumulation of CO_2 → Increased pCO_2 → Hyperventilation → Cycle of respiratory activity and apnoea continues.

Biot's breathing

- Biot's breathing also known as *ataxic breathing* is a type of periodic breathing showing alternate periods of respiratory activity and apnoea. It differs from the Cheyne–Stokes breathing in following aspects:
- It occurs at irregular intervals,
- There is no waxing and waning of tidal volume during the period of respiratory activity and
- It can never occur physiologically.

Causes. Biot's breathing indicates a disruption of the normal medullary rhythmicity of respiration. It may occur when medulla is involved in disorders, such as meningitis, head injury, medullary compressions like pontine haematomas or cerebellopontine herniation. *Central medullary lesions* are the most common cause of Biot's breathing. So, it is rare in cerebral ischaemia, which has to be bilateral to infarct the central medulla.

DISTURBANCES RELATED TO RESPIRATORY GASES

Respiratory disturbances related to respiratory gases include:

- Hypoxia,
- Hypercapnia,
- Hypocapnia,
- Asphyxia and
- Carbon monoxide poisoning.

HYPOXIA

The term hypoxia is used to denote deficiency of oxygen supply at the tissue level. It has almost replaced the term anoxia (complete absence of oxygen), which rarely occurs practically.

Causes and types

Causes. Hypoxia can occur because of any one or more of the following defects:

- Decreased oxygen tension (pO_2) of the arterial blood,
- Decreased oxygen carrying capacity of the blood,
- Decreased rate of blood flow to the tissue or
- Decreased utilization of oxygen by the tissue cells.

Types. Depending upon the mechanism of occurrence, there are four types of hypoxia:

- Hypoxic hypoxia,
- Anaemic hypoxia,
- Stagnant hypoxia and
- Histotoxic hypoxia.

Characteristic features of four types of hypoxia are summarized in Table 5.7-1.

Symptoms of hypoxia

Symptoms of hypoxia depend upon:

- Rapidity of development of hypoxia,
- Severity of hypoxia and
- Effectiveness of the body's compensatory mechanisms.

Based on the above, the hypoxia may be fulminant, acute or chronic.

1. Fulminant hypoxia refers to a severe hypoxia developing very fast, i.e. which occurs within seconds after exposure to an arterial O_2 tension of less than 20 mm Hg. It results in:

- *Unconsciousness* within 15–20 s due to lack of O_2 supply to brain and
- *Brain death* may follow in 4–5 min.

2. Acute hypoxia is produced by exposure to arterial O_2 tensions of 25–40 mm Hg (e.g. as would occur at altitudes of 18,000–25,000 ft). Symptoms of acute hypoxia are very similar to the effects of ethyl alcohol and include:

- Lack of co-ordination,
- Slowed reflexes,
- Slurring of speech,
- Overconfidence and eventually,
- Unconsciousness,

Table 5.7-1 Characteristic features of different types of hypoxia

Features	Hypoxic hypoxia	Anaemic hypoxia	Stagnant hypoxia	Histotoxic hypoxia
Pathophysiology	Occurs due to decreased O_2 tension (decreased arterial pO_2)	Occurs due to low O_2 carrying capacity of blood	Occurs due to decreased blood flow to tissue	Occurs due to decreased ability of the tissue to utilise O_2
Causes	• Low O_2 tension (low pO_2 in inspired air) • Hypoventilation • ↓ Diffusion of O_2 across respiratory membrane • Physiological shunt • Anatomical shunt	• ↓ RBC count • ↓ Hb content of blood • Altered Hb	• Shock • Circulatory failure	• Cyanide poisoning
Arterial pO_2	Decreased	Normal	Normal	Normal
Arterial O_2 contents	Decreased	Markedly Decreased	Normal	Normal
Arterial Hb contents	Normal	Reduced	Normal	Normal
% O_2 saturation (in arterial blood)	Decreased	Decreased	Normal	Normal
O_2 carrying capacity of arterial blood	Normal	Decreased	Normal	Normal
A–V (arterial–venous) pO_2 difference	Decreased	Normal	More than normal	Less than normal (nil)
Cyanosis	Present	Absent	Present	Absent
Peripheral chemoreceptor stimulation	Present (because dissolved oxygen in plasma is reduced)	Absent (because dissolved oxygen in plasma is sufficient)	Present (because arterial pCO_2 increases and pO_2 decreases)	Present (cyanide decreases oxygen utilisation at tissue level)
Tachypnoea	Present	Absent	Absent	Absent

- Coma and death can occur in minutes to hours if the compensatory mechanisms of the body are inadequate.

3. Chronic hypoxia. It occurs due to the exposure to low pO_2 (40–60 mm Hg) for long periods (e.g. as would occur after stay for extended period of time at altitudes of approximately 10,000–18,000 ft). Symptoms of chronic hypoxia are:

- Severe fatigue,
- Dyspnoea,
- Shortness of breath,
- Respiratory arrhythmias (e.g. Cheyne–Stokes breathing).

Signs of hypoxia

1. Cyanosis is the bluish discolouration of skin and mucous membrane caused by the presence of more than 5 g of deoxyhaemoglobin/100 mL of the capillary blood.

Cyanosis is not a reliable sign of hypoxia because:

- Anaemic patients may never develop cyanosis, even though they are extremely hypoxic because of an inadequate haemoglobin concentration.
- Cyanosis does not occur in histotoxic hypoxia either because the O_2 saturation of haemoglobin is normal.
- In contrast, patients with polycythaemia may be cyanotic as a result of high concentration of haemoglobin, even though their tissues are adequately oxygenated and further
- Methaemoglobin, with its slate-grey colour, can also impart a bluish colour to tissues.

IMPORTANT NOTE

There are two types of cyanosis:

Peripheral cyanosis is seen in the nail beds and is suggestive of *stagnant hypoxia*. This is because perfusion in these distally located areas is worst affected in hypotensive states. Large amount of O_2 is extracted from the haemoglobin and the concentration of deoxyhaemoglobin rises to produce cyanosis.

Central cyanosis is seen in the earlobes where skin is thin and in the mucous membrane of lips and tongue. These areas receive good blood supply and become cyanotic only if the O_2 saturation of blood is low, as occurs in the *hypoxic hypoxia*.

2. Tachycardia. It occurs as a peripheral chemoreceptor reflex response to the low arterial oxygen tension.

3. Tachypnoea presents in the hypoxic hypoxia where arterial pO_2 is low, but absent in both anaemic hypoxia and stagnant hypoxia in which the arterial pO_2 is normal.

Physiological compensatory responses to chronic hypoxia

Two types of physiologic compensatory responses known to occur in hypoxia are *accommodation and acclimatization.* For details see page 359.

Physiological basis of oxygen therapy in hypoxia

Oxygen therapy is of great value in certain types of hypoxia and at the same time of almost no value in other types.

In general, simple O_2 therapy is not of much help in treatment of hypoxia because diffusion across respiratory membrane depends upon the partial pressure of gases, therefore, alveolar pO_2 can be increased by:

- Inhalation of 100% pure oxygen or
- Inhalation of 100% pure oxygen at high barometric pressure called hyperbaric oxygen therapy.

Oxygen therapy with 100% pure oxygen at atmospheric pressure, i.e. at 760 mm Hg

1. Oxygen therapy is useful in most types of the hypoxic hypoxia. It is useful in different causes of hypoxic hypoxia include:

- In atmospheric hypoxia,
- In hypoventilation hypoxia and
- In hypoxia due to an impaired respiratory membrane diffusion.

2. Oxygen therapy is of limited value in an anaemic hypoxia, stagnant hypoxia and hypoxic hypoxia caused by the physiological or anatomical shunts; because in all these conditions oxygen is already available in the alveoli. However, in these conditions some extra oxygen can be transported in dissolved state in the blood when alveolar oxygen is increased to the maximum level and this extra oxygen may sometimes be the difference between life and death.

Therefore, the hyperbaric O_2 therapy is more useful in such conditions than O_2 therapy at atmospheric pressure.

3. Oxygen therapy is of no use in the histotoxic hypoxia because in this type of hypoxia, the tissue metabolic enzyme system is simply incapable of utilizing the oxygen that is delivered.

Hyperbaric oxygen therapy (inhalation of 100% pure oxygen at high barometric pressure)

Advantage of hyperbaric O_2 therapy over O_2 therapy at atmospheric pressure is that the former increases the amount of dissolved O_2 in plasma and is therefore unaffected by the haemoglobin concentration.

Amount of O_2 dissolved in plasma depends upon its partial pressures:

- Normally, plasma can have 0.3 mL of dissolved oxygen per 100 mL per 100 mm Hg pO_2, or 0.003 mL per 100 mL per mm Hg pO_2.
- *At 1 atmospheric pressure (760 mm Hg),* inhalation of 100% O_2 (in a patient with normal pCO_2 40 mm Hg and pH_2O 47 mm Hg) can raise the arterial pO_2 to a maximum of 760 – (40 + 47), or 673 mm Hg.

- Therefore, the maximum amount of O_2 that can be dissolved in plasma will be:
  - At 1 atmospheric pressure: 673 × 0.003, or 2 mL/100 mL,
  - At 2 atmospheric pressure: 673 × 0.003 × 2, or 4 mL/100 mL and
  - At 3 atmospheric pressure: 673 × 0.003 × 3, or 6 mL/100 mL.
- The normal demand of the body tissues (5 mL/100 mL/min), thus can be met only by dissolved O_2 in the plasma, if administered at atmospheric pressure of 2.5 or

Indications of hyperbaric O_2 therapy include:

- Carbon monoxide poisoning
- Anaemic hypoxia (due to severe anaemia)
- Decompression sickness and air embolism
- Wounds with poor blood supply
- Stagnant hypoxia (very limited value).

Caution. For therapeutic use, the hyperbaric 100% O_2 should not be used with pressures beyond two to three times atmospheric and should not be used more than 5 h because of high chances of developing O_2 toxicity.

Side effects of 100% O_2 (O_2 toxicity)

Mechanism of side effects. Inhalation of 100% O_2 produces side effects (harmful effects) due to the conversion of molecular oxygen into active oxygen, i.e. superoxide anion (O_2^-), which is free radical, and H_2O.

Side effects noted by inhalation of 100% O_2 include:

- *Irritation of airways* in the form of nasal congestion, sore throat, substernal discomfort, sneezing and coughing and bronchoconstriction may occur after about 8 h of inhalation.
- *Bronchopneumonia* may be initiated when O_2 therapy is continued for more than 24 h because of:
  - Inhibition of ability of lung macrophages to kill bacteria and
  - Decreased production of surfactant.
- *Complications in newborn infants* are very common, as they are very sensitive to get O_2 toxicity. Special dangers of O_2 therapy in the premature infants are occurrence of:
  - *Retinopathy of prematurity* (old name retrolental fibroplasia), which is characterized by retinal neovascularization and proliferation of fibrovascular tissue ultimately forming an opaque retrolental mass, leading to bilateral permanent blindness.
  - *Bronchopulmonary dysplasia* is characterized by the formation of lung cysts and opacities.

IMPORTANT NOTE

Special care is needed while treating newborns in incubators with O_2 therapy. It is cautioned that infants should never be given more than 40% O_2.

- *Nervous system complication,* i.e. derangement of cerebral activity is especially known to occur with administration of *hyperbaric O_2 therapy*. Nervous tissues are especially susceptible because of their high lipid content. Nervous symptoms include muscular twitching, tinnitus (ringing of bells in ears), convulsions, coma and even death.

HYPERCAPNIA

Hypercapnia refers to an increase in the arterial pCO_2 (normal value 40 mm Hg). When hypercapnia is the primary problem, it is associated with the respiratory acidosis (see page 346) since an increase in CO_2 promptly generates excess H^+ through following reaction:

$$H_2O + CO_2 \xrightarrow{\text{Carbonic anhydrase}} H_2CO_3 \rightarrow H^+ + HCO_3^-$$

Causes of hypercapnia

Hypercapnia rarely occurs due to an increased production of CO_2 because an increase in arterial pCO_2 causes a prompt increase in pulmonary ventilation through stimulation of central chemoreceptors, resulting in CO_2 washout and increase in pulmonary ventilation (see page 345).

Hypercapnia occurs due to:

1. ***Defective elimination of CO_2*** as occurs in:
   - Reduced pulmonary ventilation and
   - Reduced effective alveolar ventilation
2. ***Accidental inhalation*** of CO_2 in persons working in breweries and refrigeration plants.

Signs and symptoms of hypercapnia

1. ***Hyperpnoea*** occurs due to the stimulation of respiratory centre through central chemoreceptors.
2. ***Carbon dioxide narcosis*** develops when arterial pCO_2 increases above 50 mm Hg. For details see page 345.

HYPOCAPNIA

Hypocapnia, i.e. reduced pCO_2 is usually associated with *respiratory alkalosis,* since decrease in CO_2 promptly drives the following reaction in backward direction resulting in a decrease in H^+ concentration.

$$H_2O + CO_2 \xleftarrow{\text{Carbonic anhydrase}} H_2CO_3 \leftarrow H^+ + HCO_3^-$$

Causes. Hypocapnia occurs due to hyperventilation (see page 347).

ASPHYXIA

Asphyxia refers to a condition in which hypoxia (decreased pO_2) is associated with hypercapnia (increased pCO_2).

Causes

- Strangulation,
- Drowning,
- Acute tracheal obstruction (due to entry of food or due to choking) and
- Paralysis of diaphragm as in acute poliomyelitis.

Clinical stages of acute asphyxia

There are three stages of acute asphyxia:

Stage I: Stage of hyperpnoea. This stage lasts for 1 min and is characterized by:

- Increase in the rate and depth of respiration with more pronounced expiratory effort,
- Dyspnoea, cyanosis and sudden prominence of eyeballs.
- This stage occurs due to sudden and powerful stimulation of respiratory centres by acutely occurring rise in pCO_2. O_2 lack is not yet enough to stimulate ventilation.

Stage II: Stage of central excitation. This stage occurs due to excess CO_2 stimulating the centres directly and lack of O_2 stimulating the centres reflexly. It lasts for about 1 min and is characterized by all signs of central excitation, such as:

- Expiration becomes more *violent*,
- Heart rate is increased,
- Systemic blood pressure rises due to widespread vasoconstriction,
- Pupils are constricted,
- All the reflexes are exaggerated,
- Convulsions occur due to excess of pCO_2 and
- Consciousness is lost.

Stage III: Stage of central depression. This stage occurs due to direct effect of O_2 lack on vital centres causing their inhibition. It lasts for 2–3 min and its characteristic features are:

- Convulsions disappear,
- Respiration becomes slow and finally it becomes gasping (shallow and with low frequency),
- Heart rate is decreased,
- Blood pressure falls,
- Pupils are dilated,
- All the reflexes are abolished,
- The whole body lies still,
- Duration between the gasps is gradually increased and
- Finally, the death occurs.

Drowning

There are two main mechanisms by which effects of drowning, ultimately causing death:

1. Asphyxia is the cause of death in only 10% cases of drowning. Asphyxia occurs initially due to *breath-holding* and after the breaking in effect due to the severe *laryngospasm* induced by first gasp of water. The laryngospasm prevents entry of water into the lungs, but soon produces death due to asphyxia. Thus, the lungs remain dry in asphyxial deaths due to drowning.

2. Flooding of lungs with water occurs in 90% cases of drowning. The muscles of glottis relax and allow entry of water into the lungs. Further events depend upon the type of water:

- *Fresh water drowning* is associated with rapid absorption of water (since it is hypotonic) into the circulation, which causes plasma dilution and intravascular haemolysis.
- *Sea water drowning* is associated with hypovolaemia due to draining of water from the circulation into the lungs (since the sea water is hypertonic).

Note. When the patients with drowning are timely rescued and resuscitated with artificial respiration, the above described circulatory effects must be taken care of depending upon the type of water.

CARBON MONOXIDE POISONING

Carbon monoxide (CO) is a dangerous gas present in exhaust of gasoline engines, coal mines, gases from deep wells and underground drainage systems.

Toxic effects. Carbon monoxide produces anaemic hypoxia and derangement of cellular metabolic system.

Anaemic hypoxia. When CO inhaled accidentally from the above mentioned sources, carbon monoxide having 200 times more affinity than O_2 for haemoglobin combines with it to form *carboxyhaemoglobin.* The carboxyhaemoglobin produces severe anaemic hypoxia by following mechanisms:

- It does not allow the haemoglobin to take up oxygen from the alveolar air and
- The presence of carboxyhaemoglobin decreases the release of oxygen from haemoglobin, i.e. the oxygenhaemoglobin dissociation curve shifts to the left.

Derangement of cellular metabolic system. Carbon monoxide causes toxic effects on cytochrome system of the cells causing derangement of the cellular metabolic system.

Symptoms of CO poisoning. Depending on the concentration of CO in the inspired air symptoms are:

- Headache and nausea
- Loss of consciousness
- Death may occur when Haemoglobin is 50% saturated with CO.

Treatment of CO poisoning. When diagnosed timely, following measures should be taken promptly:

- Immediate termination of exposure to carbon monoxide,
- Immediate hyperbaric 100% O_2 therapy and
- Administration of air with few percent of CO_2 to stimulate the respiratory centres.

HIGH-ALTITUDE PHYSIOLOGY

Critical altitudes which are important from physiological point of view are:

- *At 10,000 ft altitude,* usually no symptoms of hypoxia are present because body can easily acclimatize to the oxygen lack. Therefore, *high altitude* is classically defined as an altitude in excess of 10,000 ft (3 km).
- *An altitude of 18,000 ft* is the highest altitude at which permanent inhabitation is possible.
- *Above 20,000 ft altitude,* hypoxia can endanger life unless O_2 is added to the inhaled air.
- *At an altitude of about 35,000 ft* commonly fly the modern aircrafts. The use of pressurized cabins in these aircrafts help to provide an environment similar to that at sea level.
- *Above 40,000 ft altitude* starts the ozone layer.

Composition of air and effect of altitude on it. Composition of air (Table 5.7-2) does not change with altitude, i.e. composition of atmosphere (% of gases) remains constant from sea level to about 30,000 ft.

Barometric pressure and partial pressure of gases. Barometric pressure at sea level is 760 mm Hg and it falls progressively with the increasing height (Table 5.7-3). With decrease in total pressure of air at increasing altitude partial pressure of gases will change.

HYPOXIA AT HIGH ALTITUDE

The effects of hypoxic hypoxia produced by decreasing pO_2 at high altitude depend upon:

- The level of altitude,
- The rate at which hypoxia develops, i.e. hypoxia occurs due to a rapid ascent (acute hypoxia) or slow ascent (subacute hypoxia) and
- Duration of exposure to hypoxia, i.e. whether short-term stay or long-term stay (chronic hypoxia).

Table 5.7-2 Concentration and partial pressure of gases in atmospheric air and alveolar air

Gas	Atmospheric air		Alveolar air	
	Concentration (%)	Partial pressure (mm Hg)	Concentration (%)	Partial pressure (mm Hg)
Nitrogen	78.62	597.0	4.9	569
Oxygen	20.84	159.0	13.6	104
Carbon dioxide	0.04	0.3	5.3	40
Water vapour	0.50	3.7	6.2	47
Total	**100**	**760**	**100**	**760**

BAROMETRIC PRESSURE AND pO_2 AT DIFFERENT ALTITUDES AND ITS EFFECT ON THE BODY

The barometric pressure, partial pressure of oxygen (pO_2) and common effects at different altitudes are given in Table 5.7-3.

Stages of hypoxic hypoxia. In a classical mould four stages of hypoxic hypoxia depending upon the value of pO_2 are described (Table 5.7-3):

1. *Stage of indifference* is usually characterized by no symptoms of hypoxia as pO_2 remains above 60 mm Hg. This occurs up to 10,000 ft altitude.
2. *Stage of reaction* starts above 10,000 ft altitude and is characterized by development of moderate hypoxia up to 15,000 ft altitude at pO_2 of 40–60 mm Hg. Hypoxic symptoms include:
   - *Cardiovascular involvement* in the form of tachycardia and hypertension,
   - *Respiratory symptoms* in the form of increased pulmonary ventilation and
   - *Early central nervous system (CNS) involvement* in the form of impaired judgement, feeling of overconfidence, talkativeness, reduction in visual acuity and emotional outburst of laughing or crying etc.
3. *Stage of disturbance* occurs when pO_2 values fall between 30 and 40 mm Hg, usually between 15,000 and 20,000 ft altitude. It is characterized by the development of severe hypoxia. In addition to the symptoms described above, the CNS involvement is aggravated.

CLINICAL TYPES OF HYPOXIC HYPOXIA AT HIGH ALTITUDE

Clinically, three types of hypoxia occurring at high altitude are described (see page 351):

CLINICAL SYNDROMES CAUSED BY HIGH ALTITUDE

The three specific entities (clinical syndromes) which need to be discussed in relation to the effects of low pO_2 at high altitude are:

- High-altitude pulmonary oedema (HAPO),
- Acute mountain sickness and
- Chronic mountain sickness.

High-altitude pulmonary oedema

High-altitude pulmonary oedema (HAPO) usually occurs as an effect of a rapid ascent at high altitude (above 10,000 ft). It is usually seen in individuals who engage in heavy physical work during first 3–4 days after a rapid ascent to high altitude due to sympathetic stimulation caused by hypoxia.

Table 5.7-3 Barometric pressure, pO_2 and common effects of different altitudes

I Level of altitude [feet (km)]	II Barometric pressure (mm Hg)	III Atmospheric air pO_2 (mm Hg)	IV Alveolar air pO_2 (mm Hg)	V Alveolar air pCO_2 (mm Hg)	VI % Oxygen saturation of haemoglobin	VII Common effects	VIII Stage of hypoxia
0 (sea level)	760	159	104	40	100		—
5000 (1.5)	630	130	80	40	95	NIL No effects	
10,000 (3)	520	110	60	40	90	Usually no symptoms except at night there may be some reduction in vision	*Stage of indifference* • The rapid ascent up to 10,000 ft *is* safe zone of ascent • Classically high altitude is in excess of 10,000 ft • No hypoxia up to pO_2 60 mm Hg
15,000 (4.5)	480	90	50	36	80	Effects of hypoxia in the form of CVS and respiratory system symptoms	*Stage of reaction* • At altitude 10,000–15,000 ft • Moderate hypoxic symptoms due to low pO_2 (40–60 mm Hg)
18,000 (5.5)	400	80	40	30	70	Above effects of hypoxia plus hypoxic symptoms due to involvement of CNS	*Stage of disturbance* • At altitude 15,000–20,000 • Severe hypoxia due to pO_2 30–40 mm Hg • Needs to be treated with O_2 therapy
20,000 (6)	350	70	<40	<30	<70	Hypoxic symptoms due to CNS involvement aggravate	*Stage of disturbance aggravate* Unconsciousness occurs when Hb saturation falls below 60%
30,000 (9)	226	47	21	24	20	Severe hypoxic symptoms	*Critical stage* • Survival is not possible without O_2 therapy above 20,000 ft altitude so-called *critical* survival altitude
40,000 (12)	140	30	12	24	15	Even with oxygen therapy	

Mechanism of development of HAPO. Since HAPO does not develop in individuals who ascent slowly at high altitude and avoid physical exertion for the first few days, so probably mechanism of development of HAPO may be following.

Sympathetic activity increased by the physical work is over and above the sympathetic stimulation caused by hypoxia (due to low pO_2) and cold (as the temperature falls by 2°C for every 1000 ft increase in altitude) products vasoconstriction leading to an increase in pulmonary capillary hydrostatic pressure. Normally, pulmonary capillary hydrostatic pressure is less than 10 mm Hg and osmotic pressure of 25 mm Hg keeps the alveoli dry. Thus, increased pulmonary capillary hydrostatic pressure drives the fluid out of the pulmonary capillaries producing pulmonary oedema. When the hypoxia is very severe, even generalized oedema may develop by the similar mechanism.

Characteristics of HAPO include:

- It responds to rest and O_2 therapy because it occurs due to aggravation of hypoxia and not due to cardiovascular or lung disease.

- It is associated with an increased pulmonary artery pressure, so it also responds to calcium channel blockers such as nifedipine, which lowers the pulmonary artery pressure.

Acute mountain sickness

Acute mountain sickness refers to the symptom complex which occurs in an individual residing at sea level, ascends to a high altitude over a period of 1–2 days for the first time. The symptoms develop 8–24 hours after arrival at high altitude and last for 4–8 days.

Symptoms are headache, nausea, vomiting, irritability, insomnia and breathlessness.

Cause of acute mountain sickness appears to be associated with cerebral oedema or alkalosis.

Mechanism of cerebral oedema. The low pO_2 at high altitude causes arteriolar dilation which is normally compensated by cerebral autoregulation. However, once the limit of cerebral circulation autoregulatory mechanism is reached, there occurs an increase in the capillary pressure that favours increased transudation of fluid into the brain tissue.

Treatment. The symptoms of acute mountain sickness can be reduced by:

- Decreasing cerebral oedema by the administration of large doses of glucocorticoids, and by
- Decreasing alkalosis by administration of acetazolamide. Acetazolamide decreases H^+ excretion through kidneys by inhibiting the enzyme carbonic anhydrase.

Chronic mountain sickness

Chronic mountain sickness (Monge's disease) occurs in some long-term residents of high altitude who develop extreme polycythaemia, cyanosis, malaise, fatigue and exercise intolerance. These individuals must be removed to a lower altitude to prevent rapid development of fatal pulmonary oedema.

PHYSIOLOGICAL COMPENSATORY RESPONSES TO HIGH ALTITUDE HYPOXIA

Two types of physiological compensatory responses known to occur in the individuals exposed to high-altitude hypoxia are accommodation and acclimatization.

I. Accommodation

Accommodation refers to the immediate reflex adjustments of the respiratory and cardiovascular systems to hypoxia. These include:

1. Hyperventilation. As mentioned above, hyperventilation occurs secondary to stimulation of peripheral chemoreceptors by low O_2 tension in the arterial blood. The increased ventilation is compensated by:

- Increasing pO_2 and reducing pCO_2,
- Reduced pCO_2 causes a respiratory alkalosis, which in turn, lowers the respiratory drive. The respiratory drive continues to increase during this time as the alkalosis is corrected by two ways:

(i) *By active regulation of cerebrospinal fluid (CSF) pH.* It is maintained by
- active transport of HCO_3^- from the CSF
- active transport of H^+ into the CSF
- During hypoxia, the anaerobic metabolic activity results in lactic acid formation, thus increases H^+ concentration in the surrounding area of central chemoreceptors. Their stimulation maintain pulmonary ventilation (action of H^+ and pCO_2 is additive).

(ii) *By active regulation of blood pH:* During hyperventilation alkalosis occurs which is corrected by kidney via more excretion of HCO_3^- in the urine.

2. Tachycardia, as mentioned above, also occurs as a peripheral chemoreceptor response to the low arterial oxygen tension. It increases O_2 delivery to the tissues by increasing cardiac output. In the individuals who go to high altitudes, the cardiac output returns to normal after several weeks.

3. Increase of 2,3-diphosphoglycerate (2,3-DPG) concentration in RBCs occurs in response to hypoxia and alkalosis. The increased 2,3-DPG concentration raises p_{50} of haemoglobin, which helps to maintain the tissue O_2 tension at slightly higher level than it would be otherwise.

II. Acclimatization

Acclimatization refers to the changes in body tissues in response to long-term exposure to hypoxia, such as when a person living at sea level goes and stays at high altitude for a long time. With longer stay, the person gradually gets acclimatized to low pO_2 by following changes in the body tissues:

1. Increase in red blood cell count or the polycythaemia, secondary to tissue hypoxia results from the release of *renal erythropoietic factor,* which acts on a plasma globulin to form erythropoietin. Erythropoietin stimulates the production of RBCs by the bone marrow. This leads to:

- Increase in haemoglobin concentration from 15 g/dL to about 20 g/dL,
- Increase in haematocrit from normal value of 40–45% to 60% after full acclimatization and
- Increase in blood volume by 20–30% leading to total increase in circulating haemoglobin by 50%.

These changes allow each unit of blood to carry additional O_2, which compensates for the decreased O_2 tension. Increase in haemoglobin and blood volume starts after

2 weeks, reaches half development in a month and is fully developed only after many months.

2. Increase in pulmonary ventilation. When an individual stays at high altitude for many days, there is a gradual increase in ventilation to an average of about five times the normal. This is because of loss of breaking effect of CO_2 due to renal correction of alkalosis, leading to decreased HCO_3^- ion concentration in CSF and brain tissues.

3. Cardiovascular changes in the form of increased heart rate, force of contraction and increased cardiac output which occur in the initial accommodation period, later on decrease back to normal once the O_2 supply to tissues becomes normal due to the changes in blood.

4. Pulmonary hypertension. It occurs secondary to the generalized hypoxic pulmonary vasoconstriction. The increased pulmonary artery pressure causes a more even distribution of pulmonary blood flow, which can improve gas exchange. However, the elevated pulmonary artery pressure can induce cor pulmonale if the hypoxia is sufficiently severe.

5. Increase in total lung capacity and diffusing capacity of the lung occur in high-altitude natives as compared to their sea-level counterparts. The increase in total lung capacity is evidenced by the enlarged chest that high-altitude natives develop.

Diffusing capacity of lungs increases due to an increase in the surface area of respiratory membrane. The greatly increased pulmonary capillary blood volume expands the capillaries thereby increasing surface area. Hypoxia increases pulmonary ventilation leading to an increase in lung volume which expands surface area of alveolar membrane. Pulmonary hypertension forces blood into greater number of alveolar capillaries than normally, especially in the upper parts of lungs which are poorly perfused.

6. Cellular and tissue acclimatization occurs after a long stay at high altitude. These include:

- Increase in oxidative enzyme concentrations within the mitochondria of many tissues, which allows more rapid generation of ATP via oxidative phosphorylation.
- Increase in mitochondrial density within the cells, which reduces the diffusion distance and provides more sites for O_2 utilization.
- Increase in capillary density in the skeletal and cardiac muscles, which reduces the diffusion distance from the blood into the cells.

7. Decreased respiratory drive is caused by lifelong exposure to hypoxia, i.e. for very prolonged periods. The reduced respiratory drive leads to higher CO_2 tension and lower O_2 tension, but it diminishes the work of respiration, which reserves more O_2 for the use by other skeletal muscles.

8. Work capacity. At high altitude (6000 m), the work capacity of unacclimatized person is 50% as that of at sea level. Acclimatization about 2 months improves the work capacity approximately to 70% of normal. The permanent residents of high altitude are so well acclimatized that their physical efficiency is almost similar to residents at sea level.

The differences observed in the high-altitude natives and an unacclimatized person at high altitude are summarized in Table 5.7-4.

Table 5.7-4 Differences between an acclimatized and an unacclimatized person at high altitude

Features	High altitude natives (acclimatized person)	New comer to high altitude (unacclimatized person)
Pulmonary ventilation (increase)	More, therefore • Chest is enlarged and barrel shaped • More alveolar ventilation • More functional residual capacity (FRC)	Less, therefore, • No change in size of chest • Less alveolar ventilation • Less FRC
Response to hypoxic stimulation	More therefore, • Hypocapnic alkalosis is less • Urine is alkaline	Less therefore, • Urine is acidic
RBC count	High (polycythaemia) therefore, • Hb contents increased • PCV increased	Comparatively low
Affinity of Hb for O_2	Less due to low arterial pO_2,Hb is not fully saturated, Hb–O_2 dissociation curve shift to right due to increased concentration of 2,3-DPG	More, therefore, body tissues are more affected by hypoxia
Vascularity of organs	More	Less
Tissue changes	At tissue level, there is increase in oxidative enzymes, myoglobin contents and number of mitochondria	Less

OTHER EFFECTS OF HIGH ALTITUDE

Factors other than hypoxia which produce changes in the body at high altitude are:

- Effects of expansion of gases,
- Effects of fall in atmospheric temperature and
- Effects of light rays.

EFFECTS OF EXPANSION OF GASES

According to the Boyle's law of gases, the pressure (P) of a given mass of gas in inversely proportional to its volume (V), i.e.,

$$P \propto \frac{1}{V}$$

Therefore at high altitude, barometric pressure and partial pressure of a gas is decreased and its volume is increased. For example, if at sea level, with atmospheric pressure of 760 mm Hg, the volume of a given gas is *1 L*.

- *At an altitude of 18,000 ft* where atmospheric pressure is about 400 mm Hg, the volume of gas increases to about *2 L*, and
- *At an altitude of 30,000 ft* where atmospheric pressure is about 225 mm Hg, the volume of the gas increases to about 3 L.

Effects of expansion of gases in the body are:

- *In gastrointestinal tract,* the expansion of gases may cause painful distension of stomach and intestines. This effect can be reduced by supporting the abdomen by a belt or by evacuation of gases while ascending rapidly.
- *In the lungs,* the expansion of gases may sometimes destroy the alveoli.
- *In the paranasal sinuses,* the expansion of gases may cause tissue damage.
- *Decompression sickness* may occur in an aviator if he is exposed to an ambient altitude in excess of about 22,000 ft (below this altitude decompression sickness is almost non-existent).

EFFECTS OF FALL IN ATMOSPHERIC TEMPERATURE

The atmospheric temperature falls by 2°C for every 1000 ft increase in altitude above sea level.

In general, the effects of low temperature on the body can be described (see page 959).

EFFECTS OF LIGHT RAYS

Ultraviolet (UV) rays at high altitude also cause many hazardous effects such as skin irritation.

PHYSIOLOGY OF HIGH ATMOSPHERIC PRESSURE

INTRODUCTION

Atmospheric pressure of 760 mm Hg at sea level is considered 1 atmospheric pressure.

Pressure increases by 1 atm for a depth of every 10 m (33 ft) as one descends beneath the sea. Thus, a person under sea at a depth of 10 m (33 ft) is exposed to a pressure of 2 atm, 1 atm due to the air above the sea level and another 1 atm due to 10 m column of water. As the depth under sea increases the pressure also increases proportionately (Table 5.7-5).

Decrease in volume of gases occurs due to compression as the pressure increases under sea. According to the Boyle's law the volume to which a given quantity of gas is compressed is inversely proportional to the pressure. With increase in the pressure the volume is decreased proportionately (Table 5.7-5).

High atmospheric pressure is met under following conditions:

- Deep sea diving,
- Going under the sea in submarines and
- Caisson's workers, i.e. the men who dig underwater tunnel, work in a chamber (Caisson's chamber) in which atmospheric pressure is high to prevent entry of water.

Physiological problems associated with life under high pressure may be divided into:

- Physiological problems at depth (due to compression effect of high atmospheric pressure) and
- Physiological problems of ascent (due to decompression phenomenon).

Table 5.7-5 Effect of depth on pressure and volume of gas

Depth [m (ft)]	Pressure (atm)	Volume (L)
Sea level	1	1
10 (33)	2	1/2 (0.5)
20 (66)	3	1/3 (0.33)
30 (100)	4	1/4 (0.25)
40 (133)	5	1/5 (0.2)
50 (166)	6	1/6 (0.167)
60 (200)	7	1/7 (0.143)
90 (300)	10	1/10 (0.1)
120 (400)	13	1/13 (0.077)
10 (500)	16	1/16 (0.062)

PHYSIOLOGICAL PROBLEMS UNDER DEPTH

At a depth of more than 30 m (100 ft) due to mechanical effects of increased atmospheric pressure there may occur:

- Caving in of the chest,
- Damage to the face and
- Squeezing of air in the paranasal sinuses and middle ear.

PHYSIOLOGICAL PROBLEMS DUE TO EFFECT OF HIGH PRESSURE ON RESPIRATORY GASES

Air under high atmospheric pressure is breathed under the sea. At high atmospheric pressure of air, the partial pressure of oxygen (pO_2), nitrogen (pN_2) and carbon dioxide (pCO_2) is also increased producing the following physiological problems:

1. Effects of increased pO_2 (oxygen toxicity)

Oxygen toxicity may be acute or chronic.

Acute oxygen toxicity occurs on exposures to 4 atm pressure of oxygen (pO_2 in lungs about 3000 mm Hg).
Acute oxygen poisoning is typically characterized by nervous system complications as the *brain tissue is especially susceptible to acute oxygen poisoning.* At high tissue pO_2 the molecular oxygen is converted into active oxygen, i.e. superoxide anion (O_2^-) which is free radical.

Nervous complications of acute oxygen poisoning include disorientation, dizziness, convulsions and even coma.

Chronic oxygen toxicity occurs due to prolonged exposure (8–24 h) to oxygen at 1 or 1.5 atmospheric pressure (see page 355).

2. Effects of increased pN_2 (Nitrogen narcosis)

Due to increased pN_2, the nitrogen dissolves *gradually* into the body fluids and more easily into fats. The cell membrane of neurons contains high lipid content, so more nitrogen is dissolved in the neurons of brain. The nitrogen dissolved in the cell membranes of neurons alters the ionic conductance through the membrane and finally decrease the neuronal excitability producing nitrogen toxicity known as nitrogen narcosis. Nitrogen narcosis is characterized by:

- *Euphoric symptoms.* The individual becomes jovial and carefree. These are followed by the impairment of mental functions and intelligence, individual becomes drowsy and has poor muscular co-ordination.

PHYSIOLOGICAL PROBLEMS OF ASCENT

The two physiological problems which occur when an individual ascends back to sea level after sufficient exposure to high atmospheric pressure in the deep sea are:

- Decompression sickness and
- Air embolism.

DECOMPRESSION SICKNESS

Decompression sickness is also known as Caisson's disease, dysbarism, compressed air sickness, the bends and diver's palsy. When the individual ascends rapidly to the sea level after sufficient exposure to high atmospheric pressure deep in the sea, nitrogen is decompressed and escapes from the tissues at a faster rate. Being gas it forms bubbles while escaping rapidly from the tissues. The gas bubbles block the blood vessels producing tissue ischaemia and sometimes the tissue death. The symptoms produced by escaping gas bubbles constitute the *decompression sickness.*

Symptoms of decompression sickness are:

- *Pain in joints and muscles* of legs or arms. The joint pain accounts for the term *'bends'* that is often used to describe the decompression sickness.
- *Sensation of numbness.*
- *The chokes.* The chokes refer to the serious shortness of breath which is often followed by severe pulmonary oedema, and occasionally death.
- *Paralysis* of muscles may occur temporarily due to escape of nitrogen bubbles from the myelin sheath of motor nerves. This is called *diver's palsy* (one of the names of this disease).
- *Coronary ischaemia* or myocardial infarction may occur due to the blockage of coronary capillaries by the nitrogen bubbles.
- *Neurological symptoms* like dizziness, paralysis of muscles, or collapse and unconsciousness may occur due to the blockage of blood vessels of brain and spinal cord.

Treatment of decompression disease. Tank decompression is used for treatment of the decompression disease.

AIR EMBOLISM

Air embolism is another physiological problem which may occur during the rapid ascent from a depth below the sea level.

Manifestations of air embolism include chest pain, tachypnoea, systemic hypotension and hypoxaemia. In severe cases, air emboli may travel to the systemic circulation, block the blood flow to some vital organs and may even result in death.

PREVENTION OF PHYSIOLOGICAL PROBLEMS OCCURRING AT DEPTH AND ON ASCENT

Deeper and longer dives can be made safe by following preventive measures.

1. Use of breathing apparatus

Use of breathing apparatus which delivers gas to breath and either absorbs carbon dioxide (closed circuit apparatus) or release carbon dioxide as bubbles into the surrounding

(open circuit apparatus). An example of such an apparatus is SCUBA diving.

SCUBA diving. SCUBA (self-contained underwater breathing apparatus) is a compact arrangement for breathing which the diver can carry with him under the water. In this apparatus the air is compressed so that more air is carried in less volume and also the gas amounts to a substantial quantity even when the ambient pressure is high.

2. Use of breathing mixture containing helium and low oxygen concentration

Use of breathing mixture containing helium and low oxygen concentration is less harmful than natural air because:

Low oxygen concentration prevents occurrence of oxygen toxicity.

Helium when replaced with nitrogen provides following advantages:

- Because of its smaller molecule and lower density than nitrogen it is easier to breathe, it diffuses faster and it is easier to eliminate its bubbles from the body.
- The amount of helium trapped in the body under high atmospheric pressure is much less than that of nitrogen because its solubility in the body fluids is less than half that of nitrogen.
- Being less toxic than nitrogen, its narcotic effect is only one-fifth that of nitrogen.

IMPORTANT NOTE

At high pressure oxygen–helium mixture can produce high pressure nervous syndrome. This condition is characterized by tremors, drowsiness and decreases alpha wave activity of electroencephalogram. The cause of this condition is not clearly known but may be because of other gases (like xenon, krypton, argon and neon). These gases at atmospheric pressure are physiologically inert and have anaesthetic effect at high pressure. The anaesthetic activity of these gases depends on fat solubility.

3. Slow ascent or use of decompression tank

Slow ascent with short stay at regular intervals, i.e. slow and stepwise ascent ensures that only a small amount of bubbles are formed at a time. They are eliminated before further ascent. In this way, decompression sickness can be prevented effectively.

Decompression tank is based on the principle of slow ascent. After a rapid ascent, the individual is put into a pressurized tank whose pressure is lowered gradually up to a normal atmospheric pressure. Usually no decompression is required after a 30 m dive for less than 30 min. Decompression for a period of 3 h is required for a 60 min dive at a depth of 60 m and 20 min dive at a depth of 100 m.

ARTIFICIAL RESPIRATION AND CARDIOPULMONARY RESUSCITATION

ARTIFICIAL RESPIRATION

Artificial respiration (AR) alone is required as an emergency life-saving procedure:

I. *When there is sudden stoppage of breathing* as seen in:
- Drowning,
- Electrocution,
- Anaesthetic accidents,
- Carbon monoxide poisoning,
- Strangulation and
- Accidents.

II. *Artificial respiration may also be needed when breathing is expected to stop gradually* as in paralysis of muscles in:
- Poliomyelitis,
- Diphtheria and
- Ascending paralysis.

It is important to note that the tissues of brain, particularly cerebral cortex, develop irreversible damage if oxygen supply is stopped for 5 min. So, the resuscitation must be started quickly without any delay before the development of cardiac failure.

METHODS OF ARTIFICIAL RESPIRATION

Mouth-to-mouth breathing method

Various manual methods of artificial respiration have been described in past and discarded. Presently, the only manual method employed is mouth-to-mouth breathing (exhaled air ventilation) (Fig. 5.7-4) because:

- It can be applied quickly without waiting for the availability of any aid.
- It is simple and effective measure of resuscitation.
- It can be applied in all age groups.
- It is the only technique capable of producing adequate ventilation.
- It also works by expanding the lungs.

Procedure

- The procedure should be performed swiftly and alertly.
- The procedure is performed after placing the patient in a supine position.
- It is essential to provide and maintain a clear airway for the procedure to be effective. Therefore, any foreign material present in the mouth cavity must be removed with fingers, e.g. grass, straw, etc. (in case of drowning patients), artificial denture if any; mucus, saliva and blood clot, etc. The tongue must be drawn forward and it must be prevented from falling posteriorly causing airway obstruction. The clothes around the neck and

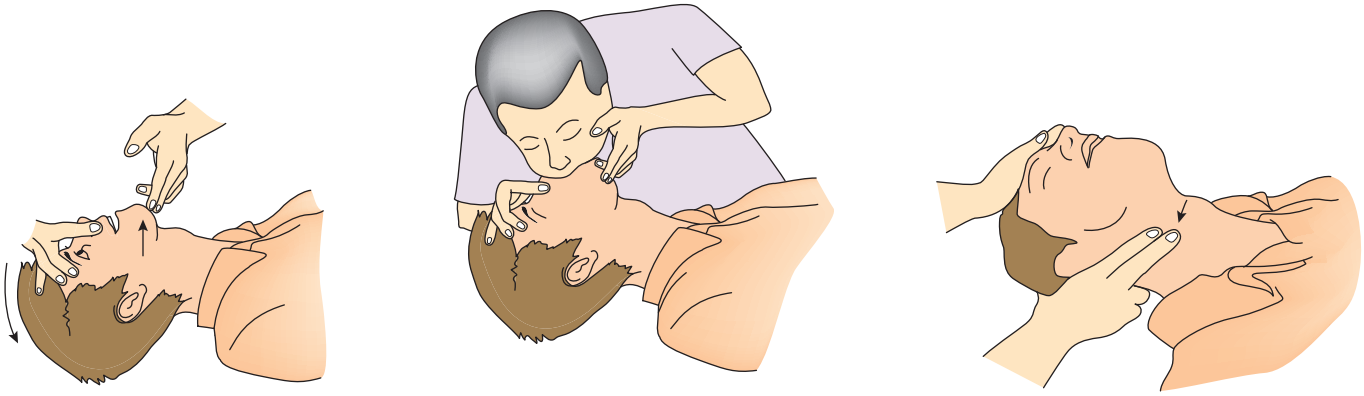


Fig. 5.7-4 Mouth-to-mouth breathing: A, the neck is extended by placing one hand under the chin and pressing the forehead with other hand; B, nostrils are closed with thumb and index finger and resuscitator exhales into the patient's airway by tightly placing his mouth over the patient's mouth and C, allows the patient to exhale passively by unsealing nose and mouth.

chest region must be loosen. If the mouth is full of blood, mouth-to-nose respiration should be given.

- To begin with patient's neck is extended by placing one hand under the chin and lifting it and pressing the forehead with the other hand (Fig. 5.7-4A). This prevents the flaccid tongue from falling back into the pharynx.
- Then the patient's nostrils are closed by the thumb and index finger of the hand (Fig. 5.7-4B).
- The resuscitator then takes a deep breath and exhales air into the patient's airway after tightly placing his mouth over patient's mouth and noting the expansion of the chest at the same time. The volume of the air exhaled must be twice the normal tidal volume. This expands the patient's lungs.
- Then, the resuscitator removes his mouth from that of the patient, allowing expiration to occur passively due to the elastic recoil of the lungs and chest (Fig. 5.7-4C).
- Some of the air is likely to enter the stomach through the oesophagus. It can be easily expelled by pressure on the epigastrium.
- The above procedure is repeated 12–16 times/min till spontaneous breathing returns, or till the patient is shifted to a hospital.

It is important to remember that:

- Mouth-to-mouth method is most effective manual method because the CO_2 present in the expired air by the resuscitator can also directly stimulate the respiratory centres and facilitate the onset of respiration.

Mechanical methods of artificial respiration

The mechanical respirators are employed when artificial respiration has to be continued for long periods. The mechanical respirators are of two types:

- Tank respirators and
- Ventilators.

1. Tank respirators or the so-called iron lung chambers as the name indicates consist of an airtight chamber made of iron or steel. There are various types of mechanical respirators. A commonly used is Drinker respirator. In this respirator, the patient is kept inside the tank by placing the head outside the chamber. In Drinker method, alternate positive and negative pressure breathing machines produce periodic inflation and deflation of the lungs.

2. Ventilators are the artificial respiration machines by which air or oxygen is pumped into the lungs with pressure intermittently through a rubber tube introduced into the patient's trachea. Inflation occurs when air is pumped and expiration occurs by elastic recoil of chest and lungs, when it is stopped. Presently, two types of ventilators are available:

- Volume ventilator pumps a constant volume of air into the patient's lungs intermittently with minimum pressure.
- Pressure ventilator pumps the air with a constant high pressure into the patient's lungs.

CARDIOPULMONARY RESUSCITATION

Cardiopulmonary resuscitation (CPR) is required in some patients when heart and respiration both stop. Breathing usually stops before the heart stops, so artificial respiration should be started immediately.

Emergency plan of cardiopulmonary resuscitation

The following plan called ABC of CPR has proved useful in reviving such patients:

A. Airway care is required in the unconscious patients. Immediately, tilt head back with a hand under the neck to maintain an open airway.

B. Breathing by artificial respiration (AR) method is required when the patient is not breathing. Mouth-to-mouth respiration

should be immediately started. Feel carotid pulse, if present continue AR only (Fig. 5.7-4).

C. Cardiac massage is required when carotid pulse cannot be felt. During external cardiac massage sternum should be depressed by 4–5 cm at a rate of 80–90 times/min. The cardiac compression should be alternated with mouth-to-mouth respiration at a rate of one ventilation to five chest compression (Fig. 5.7-5).

PULMONARY FUNCTION TESTS

ROLE OF PULMONARY FUNCTION TESTS IN CLINICAL PRACTICE

The roles of pulmonary function tests in clinical practice are:

1. In diagnosis of pulmonary diseases, i.e. for confirmation of clinical diagnosis,
2. To follow the progress of disease and its response to treatment,
3. To objectively assess the severity of disease,
4. To assess respiratory status before anaesthesia,
5. To assess physical fitness for certain jobs, such as those involving strenuous physical exercise, flying at high altitude, etc.
6. To obtain medicolegal information in certain situation.

CLASSIFICATION

Pulmonary function tests can be classified into following groups:

A. Ventilatory function tests,
B. Tests of diffusion and
C. Tests of ultimate purpose of respiration.

VENTILATORY FUNCTION TESTS

Ventilatory function tests are meant for the assessment of the expansion of lungs and chest wall; and for the assessment of restrictive and obstructive ventilatory defects. The assessment of ventilatory functions can be accomplished by:

I. Measurement of various lung volume and capacities,
II. Measurement of dead space,
III. Measurement of compliance and
IV. Measurement of airway resistance.

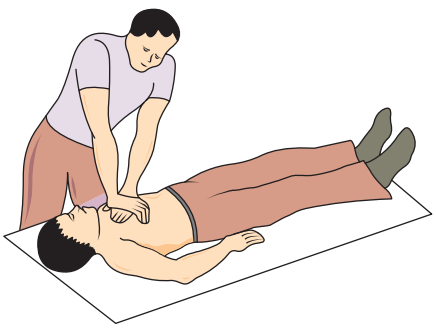


Fig. 5.7-5 Procedure showing external cardiac massage.

I. MEASUREMENT OF VARIOUS LUNG VOLUMES AND CAPACITIES

Various lung volumes and capacities have been described on page 301. Most of the lung volumes and capacities except residual volume, functional residual capacity and total lung capacity can be measured by spirometry. Functional residual capacity is determined by nitrogen wash-out method or helium dilution method, and then residual volume and total lung capacity are calculated.

Spirometry

Spirometry refers to the recording of volume changes during various clearly defined breathing manoeuvres. It can be performed using a simple spirometer, a modified spirometer called respirometer or computerized spirometer.

Simple spirometer (Fig. 5.7-6) is made of metal. It consists of following parts:

- *Outer chamber* or container which is filled with water.
- *Floating drum,* or a gas bell with 6 L capacity, floats in the water in an inverted manner. It is attached to a chain which passes over a pulley bearing a balancing weight and a

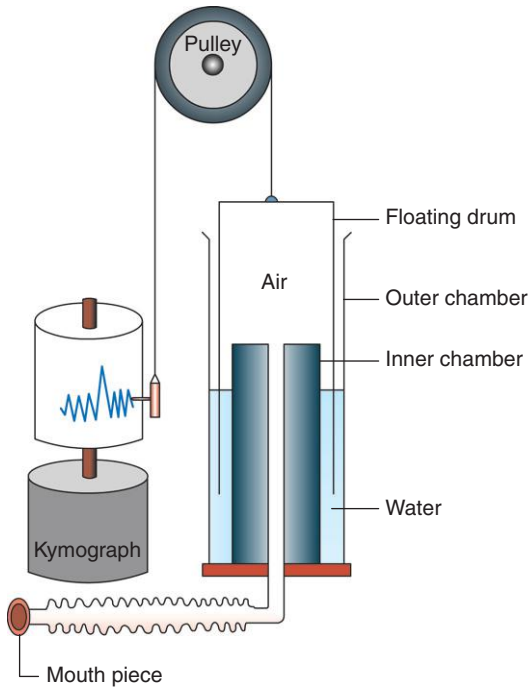


Fig. 5.7-6 Simple spirometer.

writing needle (pen). The needle (pen) moves with the movement of the floating drum. The floating drum is thus counterpoised and has very little inertia and friction.

- *Inner chamber* is open at the top end which lies above the water level in the outer chamber and is connected to a tube at the bottom end. At the end of tube, a mouth piece is attached through which the subject is made to respire.
- *Kymograph* is a recording drum on which the movements of the needle are recorded.

II. MEASUREMENT OF DEAD SPACE

Dead space air is the portion of minute ventilation that does not take part in the exchange of gases. Normally, it is constituted by the air present in the conducting zone of respiratory passages *(anatomical dead space),* but in some diseases may additionally include also poorly perfused alveoli (physiological dead space). For details see page 319.

III. MEASUREMENT OF COMPLIANCE

Compliance (C) expresses the distensibility (expansibility) of the lungs and chest wall. Reduced compliance produces a condition called restrictive lung disease. In clinical testing, the restrictive lung diseases are evaluated indirectly by measurement of various lung volumes and capacities as described above.

IV. MEASUREMENT OF AIRWAY RESISTANCE

Airway resistance is the resistance caused by the friction of gas molecules between themselves and the walls of the airways. Airway resistance is increased in many obstructive lung diseases. Like compliance, airway resistance is also seldom measured directly for clinical use. However, it may be required for the research purposes.

TESTS OF DIFFUSION

Pulmonary diffusion refers to the transfer of gases from the alveoli to the capillary blood across the respiratory membrane. The exchange of gases in the lungs was earlier believed to be dependent merely on the ability of the gases to diffuse across the respiratory membrane. This term led to the use of term *diffusion capacity.* However, later it was realised that many other factors like ventilation–perfusion balance, pulmonary capillary blood volume, Hb concentration of the blood and rate of reaction of gases with Hb are also involved in the exchange of gases. Therefore, nowadays, the term *transfer factor,* rather than diffusion capacity is used.

TESTS OF ULTIMATE PURPOSE OF RESPIRATION

Since the ultimate purpose of respiration is to supply O_2 from atmosphere to the tissues and removal of CO_2 from the tissues into the atmosphere; so, the estimation of the arterial blood pO_2, pCO_2 and pH (blood gas analysis) are most fundamental of all the pulmonary function tests.

Estimation of arterial pO_2, pCO_2 and pH

For blood gas analysis, arterial blood sample is usually taken from the radial artery or femoral artery. The estimation of pO_2, pCO_2 and pH can be done within a minute or so using a very small sample of blood with the help of miniaturised glass electrodes.

Arterial pO_2 levels in young healthy adult vary from 85 to 105 mm Hg with a mean of 95 mm Hg. The value may drop by up to 15% in healthy elderly subjects due to an increase in ventilation–perfusion inequality.

Causes of decreased arterial pO_2 are:

- Alveolar hypoventilation, i.e. inadequate intake of air.
- Diffusion defect, i.e. inadequate transport of O_2 across the respiratory membrane.
- Arteriovenous admixture, i.e. right to left vascular shunt and
- Decreased ventilation–perfusion ratio, i.e. physiological shunt as seen in the patients with emphysema.

Arterial pCO_2 and pH level in normal adult are about 40 and 7.4 mm Hg, respectively and are basically determined by the volume of alveolar ventilation:

- Hypoventilation causes increased pCO_2 and reduction in arterial pH (respiratory acidosis),
- Hyperventilation produces decreased pCO_2 and increase in arterial pH (respiratory alkalosis).

Chapter 5.8

Physiology of Exercise

INTRODUCTION

Physiology of exercise has generated significant interest and has gained importance because of:

- The current concern with physical conditioning and improvement in performance of athletes, sports persons, military and paramilitary personnel all over the world,
- Role of exercise in prevention of cardiovascular diseases and physical fitness of population groups,
- Role of exercise (stress tests) in evaluation of the cardiovascular and respiratory systems and
- Role of exercise in rehabilitation of the cardiac invalids.

EXERCISE: TYPES AND GRADING

Types of exercise

Exercise may be dynamic or isotonic and static or isometric.

Dynamic exercise involves isotonic muscle contractions. External work is involved in this type of exercise.

Static exercise involves isometric muscle contractions.

Grading of exercise

WHO (1978) has classified muscular exercise into four grades depending upon the heart rate and oxygen consumption. Oxygen consumption can be expressed as litres per minute or as relative load index, i.e. percentage of maximum O_2 utilization. The oxygen utilization can also be expressed as metabolic energy expenditure (MET) test. One MET is equivalent to resting O_2 uptake of 250 mL/min for an average adult man and 200 mL/min for an average woman. The four grades of exercise are shown in Table 5.8-1.

ADJUSTMENTS TO EXERCISE

Adjustments to physical (muscular) exercise depend upon the type of exercise, grade of exercise, cardiac reserve (i.e. efficiency of the heart), muscle power, training, motivation and the state of nutrition.

RESPONSES TO EXERCISE

Exercise, basically, is a period of enhanced energy expenditure. The energy for muscular exercise is provided by the increased fuel consumption, which is reflected as *greater* O_2 consumption and CO_2 production. The increased O_2 delivery to the tissues and removal of CO_2 from the tissues is achieved by:

- Cardiovascular responses to exercise,
- Respiratory responses to exercise and
- Changes at tissue levels during exercise.

Table 5.8-1 Grades of exercise

Grade	Level	Heart rate (beats per min)	O_2 consumption (L/min)	Relative load index (RLI) (% of max. O_2 consumption)	METs
I	Light (mild)	<100	0.4–0.8	<25	<3
II	Moderate	100–125	0.8–1.6	25–50	3.1–4.5
III	Heavy	125–150	1.6–2.4	51–75	4.6–7
IV	Severe	>150	>2.4	>75	>7

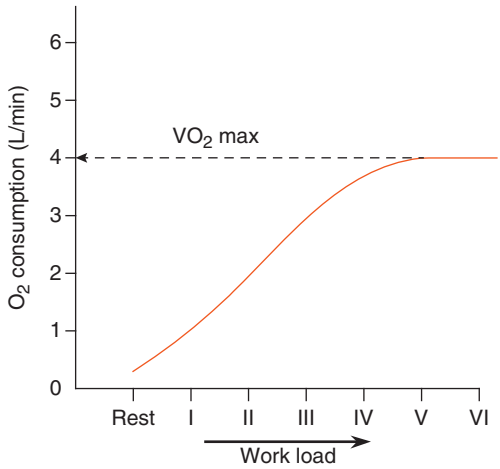


Fig. 5.8-1 Oxygen consumption during exercise.

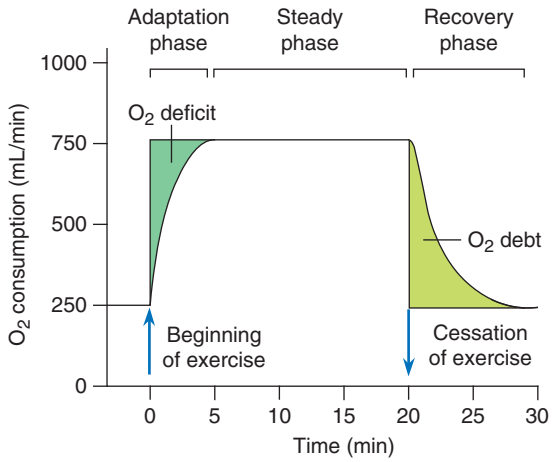


Fig. 5.8-2 Oxygen deficit and oxygen debt.

In addition, endocrine responses to exercise occur and play a regulatory role by regulating the water loss and availability of fuel during exercise.

OXYGEN CONSUMPTION DURING EXERCISE

Oxygen consumption during exercise. The energy for muscular work during exercise is provided by the increased fuel consumption, which is reflected in greater O_2 consumption and CO_2 production.

Oxygen consumption (VO_2) during rest is about 250 mL/min, which increases linearly with severity of exercise up to a certain limit, beyond which a plateau is reached (Fig. 5.8-1).

Maximal oxygen consumption (VO_2 max) is the term used to define the level of oxygen consumption beyond which no further increase in O_2 consumption occurs with further increase in the severity of exercise.

- Average VO_2 max in an adult is 3 L/min and in a trained athlete it may be as high as 5 L/min.
- VO_2 max represents the highest attainable rate of aerobic metabolism during the performance of rhythmic muscular work that exhausts the subject within 5–10 min.
- VO_2 max increases during childhood and reaches a peak during early adulthood, after that a gradual and steady decline takes place with the increasing age.

Criteria for establishing VO_2 max are:

(i) Oxygen consumption (VO_2 reaches a plateau)
(ii) Achievement of maximum heart rate which can be calculated by a simple formula depending on age is 220–age (in years).
(iii) Respiratory quotient (RQ) reaches more than 1.15
(iv) Blood lactic acid level increases more than 70–80 mg/dL (normal range is 20–40 mg/dL).

OXYGEN DEFICIT AND O_2 DEBT

The period of muscular exercise can be divided into three phases (Fig. 5.8-2):

1. Adaptation phase refers to the beginning of muscular exercise (first 2–4 min) during which oxygen consumption increases linearly and reaches the maximal O_2 consumption (VO_2 max). The VO_2 max at this stage is much less than the oxygen demand; thus an *oxygen deficit* is established. So, the energy requirement over and above the limits of O_2 consumption is met with by the anaerobic pathway.

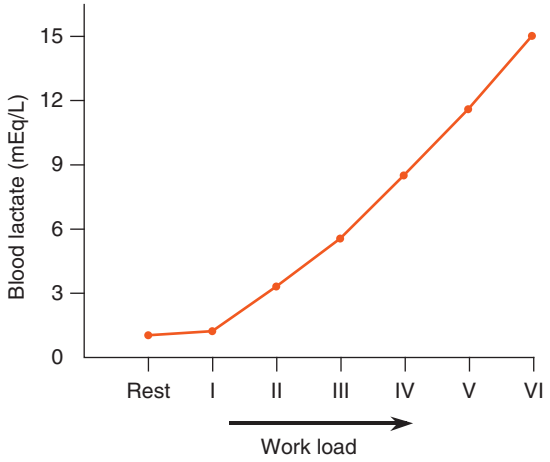


Fig. 5.8-3 The relationship between blood lactate and severity of exercise.

2. Steady phase of exercise is characterized by a maximum O_2 consumption (VO_2 max) throughout, i.e. a plateau phase of O_2 consumption and work done relationship. During this phase also, as mentioned above, the excess energy requirement is met with by the anaerobic pathway, i.e. by breakdown of creatine phosphate and muscle glycogen. As a result of anaerobic release of energy in the muscles, the blood levels of lactic acid begin to rise steeply when the oxygen consumption exceeds 2 L/min (Fig. 5.8-3). In the blood, lactic acid is buffered by the bicarbonate buffer as:

$$H^+ + HCO_3^- \Leftrightarrow H_2CO_3 \Leftrightarrow CO_2 + H_2O$$

The extra CO_2 so evolved is removed by hyperventilation. Since CO_2 evolved is more than O_2 consumed, the respiratory quotient (CO_2 evolved/O_2 consumed) may reach 1.5–2 during severe exercise.

3. Recovery phase refers to the period after the cessation of exercise during which an extra amount of O_2 is consumed. The amount of extra O_2 consumed during recovery phase is called O_2 debt and is proportionate to the extent to which *oxygen deficit* occurred during exercise. In other words, the O_2 deficit which occurs during exercise is repaid during the recovery phase in the form of O_2 debt. The extra amount of O_2 consumed during the recovery phase (O_2 debt) is used:

- To remove the excess lactate collected due to anaerobic glucose breakdown,
- To replenish the ATP and phosphoryl creatine store,
- To replace the small amounts of O_2 that have come from the myoglobin and
- To resupply dissolved O_2 in the tissue fluids and blood.

During recovery phase, respiratory quotient falls to low values since CO_2 is retained to form HCO_3^- and lactate is mobilized by the Cori's cycle.

Table 5.8-2 Redistribution of cardiac output in standing posture during exercise

	At rest	During heavy exercise	Change
Cardiac output	5 L/min	24 L/min	Increased by five to six times
Blood flow to:			
• Skeletal muscles	750–800 mL/min	20 L/min	25 times
• Heart	250 mL/min	1 L/min	four times
• Brain	750 mL/min	750 mL/min	No change
• Visceral	2600 mL/min	500 mL/min	Decreased by 80%
• Cutaneous	500 mL/min	400 mL/min	Initially decreased
		1000 mL/min	Later increased

CARDIOVASCULAR RESPONSES TO EXERCISE

To meet the increased energy demand of muscles during exercise the primary cardiovascular response is in the form of:

- Increase in the skeletal muscle blood flow,
- Redistribution of blood flow in the body,
- Increase in the cardiac output,
- Blood pressure changes and
- Changes in the blood volume.

SKELETAL MUSCLE BLOOD FLOW

At rest the blood flow to the skeletal muscle is about 2–4 mL/100 g/min of muscle tissue. During strenuous exercise muscle blood flow can increase up to 20 times, i.e. about 50–80 mL/100 g/min muscle tissue. This is called *exercise hyperaemia.* This tremendous increase in the muscle blood flow during exercise is made possible by:

- Arteriolar dilatation and
- Opening up of the closed capillaries which greatly increase the surface area and the rate of blood flow (For detail see page 275).

REDISTRIBUTION OF BLOOD FLOW

As mentioned earlier, the tremendous increase in the skeletal muscle blood flow is possible due to increased cardiac output (discussed later in detail) and redistribution of cardiac output in following manner (Table 5.8-2):

Coronary blood flow. During exercise, coronary blood flow is increased by four to five times with 100% O_2 utilization (see page 266).

Visceral blood flow is temporarily curtailed in co-ordination with increase in muscle blood flow. It is brought about by the increased sympathoadrenal discharge.

Splanchnic blood flow is decreased by 80% in severe exercise.

Renal blood flow is also decreased by 50–80% in severe exercise.

Cutaneous blood flow at rest is about 500 mL/min.

- *Decrease* in cutaneous blood flow occurs initially in the beginning of exercise due to reflex vasoconstriction.
- *Increase* in cutaneous blood flow is noted in sustained exercise when body temperature rises, to dissipate the heat generated during exercise, as the blood flow through the skin is controlled predominantly by the requirements of temperature regulation.

Cerebral blood flow at rest is about 750 mL/min and remains unchanged during any grade of muscular exercise.

Adipose tissue blood flow is increased by four times during exercise. This helps to deliver fatty acids mobilized from triglyceride stores to the working muscles.

INCREASE IN CARDIAC OUTPUT

Normal cardiac output is about 5–6 L/min. During exercise, the cardiac output is increased depending upon the severity of exercise. In maximum exercise it may increase by five to six times. Since, cardiac output is the product of heart rate and stroke volume, an increase in both contributes to the increase in a cardiac output during exercise (Fig. 5.8-4).

Increase in heart rate

Heart rate increases linearly with the severity of exercise. The increase in heart rate occurs as soon as the exercise begins or may be seen even before the exercise begins (anticipatory tachycardia).

Factors contributing to tachycardia during exercise are:

- *Increased sympathetic discharge.*
- *Peripheral reflexes* originating from the exercising muscles (muscle spindles, muscle-tendon receptors and organ of Corti) and joints.
- *Local metabolic factors.* Muscle tissue has free nerve endings which are stimulated by the lactic acid potassium ions and other metabolites which collect in exercising muscles possibly contribute to the sustained increase in heart rate during prolonged exercise.
- *Humoral factors,* such as release of adrenaline and noradrenaline and possibly thyroid hormones during exercise.
- *Intrinsic factors.* Stimulation of sinoatrial node in the right atrium due to increased venous return, which increase the heart rate during exercise. This is known as *Bainbridge reflex.*
- *Increased temperature* in the myocardium due to increased activity of the heart during exercise may directly increase the rhythmicity of the pacemaker.

Increase in stroke volume

Under normal conditions, the average stroke volume is about 80 mL/beat and may increase up to twice the normal value during exercise (Fig. 5.8-4).

Mechanisms responsible for increase in stroke volume

It has been stated that an increase in the stroke volume during exercise occurs due to gearing up of both the control mechanisms, i.e.

- Intrinsic autoregulation or Frank–Starling mechanism (for details see page 218).
- Extrinsic regulation or autonomic and neural mechanism (for details see page 220) as explained.

BLOOD PRESSURE CHANGES DURING EXERCISE (FIG. 5.8-4)

In systemic circulation

Systolic blood pressure is always raised by exercise since it depends upon the cardiac output which is increased in exercise.

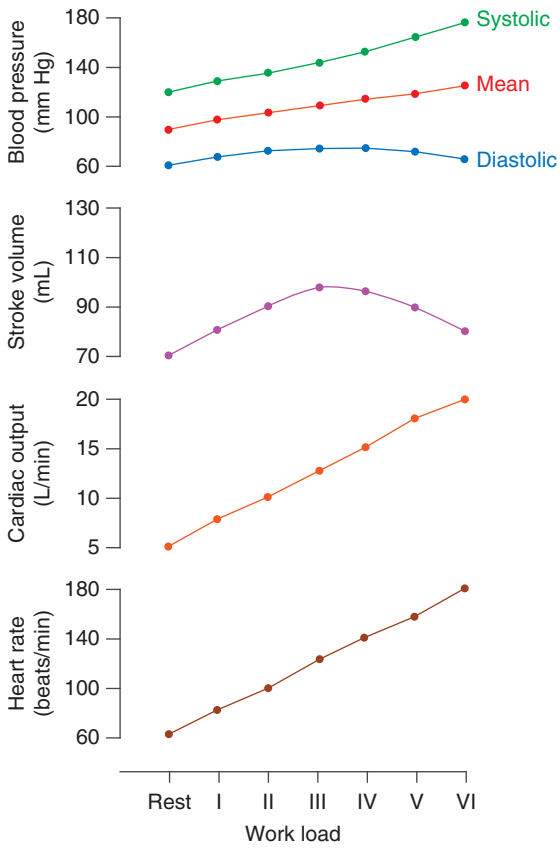


Fig. 5.8-4 Effect of severity of muscular exercise on cardiovascular functions.

The blood pressure remains elevated during exercise and is not reflexly corrected by baroreceptor reflex. This has been explained by the fact that the neurons descending from the hypothalamic defence centre inhibit the baroreceptor afferents.

Diastolic blood pressure which primarily depends upon the peripheral resistance may mildly increase or decrease or remain unchanged depending upon the change in total peripheral resistance. Mostly, the vasodilatation in the skeletal muscles balances the vasoconstriction in other tissues, so diastolic blood pressure is usually not changed much.

Mean blood pressure is usually increased. It helps to increase the skeletal muscle blood flow by providing greater pressure head in the face of dilated resistance vessels.

In pulmonary circulation

- *Systolic blood pressure* in the pulmonary artery may rise during heavy exercise to 25–30 mm Hg from 15–20 mm Hg at rest,
- *Diastolic blood pressure* may rise from 5–8 mm Hg at rest to 8–10 mm Hg and
- *Mean blood pressure* may reach to 15 mm Hg from 8–12 mm Hg at rest.

CHANGES IN BLOOD VOLUME DURING EXERCISE

Blood volume during exercise is decreased by 15% resulting in haemoconcentration. Blood volume is decreased due to more plasma loss at the capillary level due to following reasons:

- Increased hydrostatic pressure in capillaries and
- Increased tissue fluid osmotic pressure due to accumulation of osmotically active metabolites in tissue spaces such as potassium, phosphate and lactic acid.

SUMMARY OF CARDIOVASCULAR RESPONSES TO EXERCISE

The cardiovascular responses to exercise are summarized in Fig. 5.8-5.

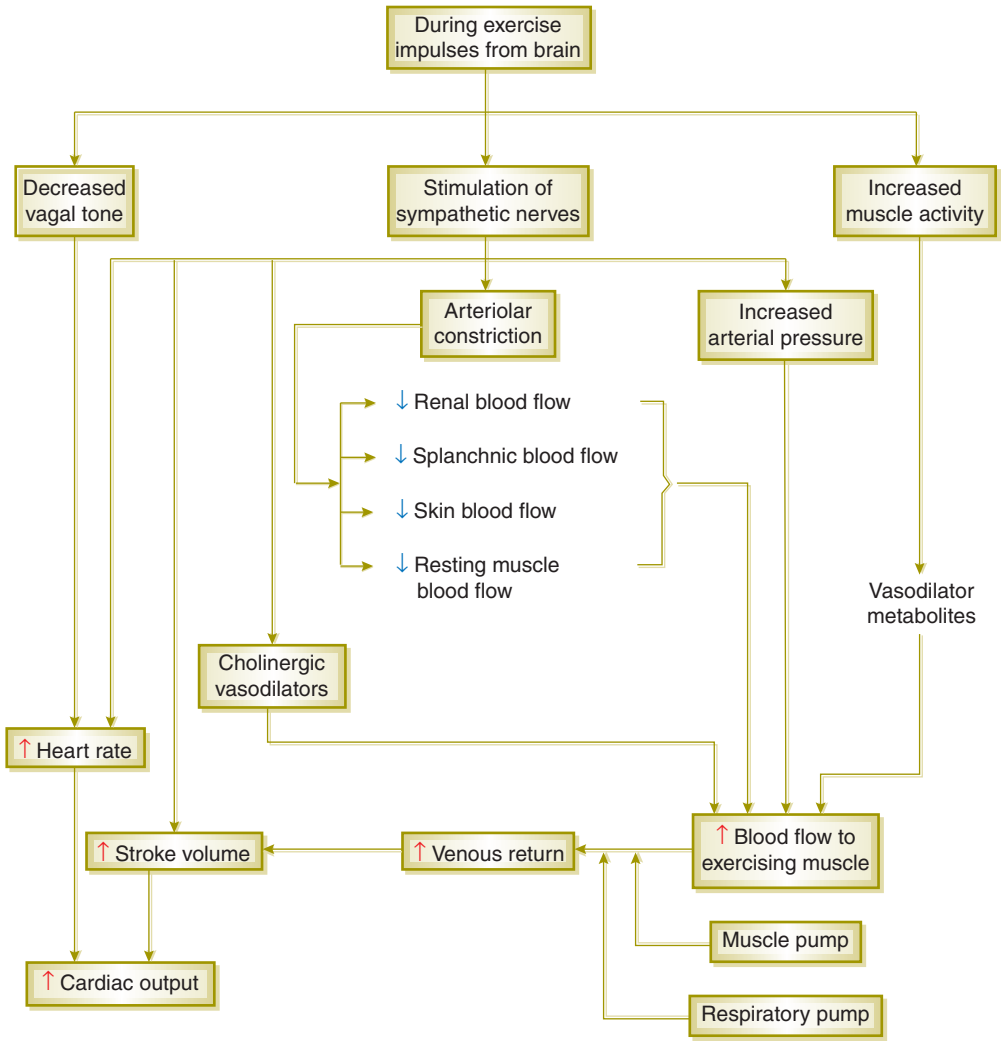


Fig. 5.8-5 Summary of cardiovascular responses to exercise.

RESPIRATORY RESPONSES TO EXERCISE

I. INCREASE IN PULMONARY VENTILATION

The pulmonary ventilation increases linearly with the increase in intensity of exercise (O_2 consumption) until the anaerobic threshold is reached (Fig. 5.8-6). The anaerobic threshold occurs at approximately 60% of the maximal exercise level, regardless of the level of physical fitness. Above the anaerobic threshold, the pulmonary ventilation increases out of proportion to the increase in O_2 consumption because lactic acid that is generated imposes an additional respiratory drive.

Mechanism of increased pulmonary ventilation

There are several factors which can account for the marked increase in the pulmonary ventilation occurring in severe exercise. The probable factors are:

1. Neural control mechanisms have been suggested to play a main role than chemical mechanisms in increasing pulmonary ventilation during exercise.

Neural control mechanisms which have been suggested to make contribution to exercise hyperpnoea are:

- *Cerebral cortex,* the seat of conscious thought and voluntary activity may be responsible for the anticipatory hyperpnoea which may occur due to psychic stimuli just before the beginning of exercise.
- *Afferent impulses from proprioceptors* in the muscles and joints are at least partly responsible for exercise hyperpnoea.
- *Increase in body temperature* during sustained exercise may also make some contribution to the exercise hyperpnoea through neural mechanism.

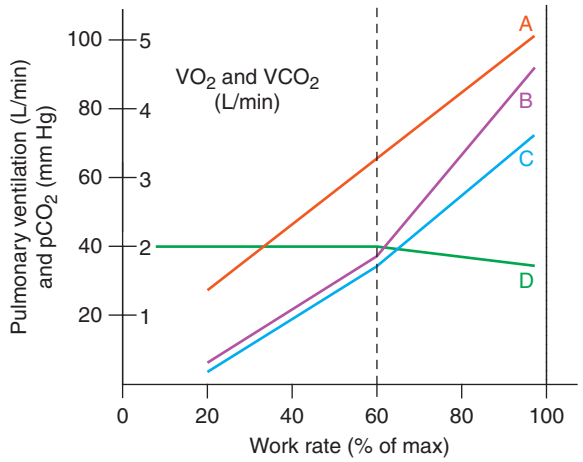


Fig. 5.8-6 The effect of exercise on respiratory parameters of gas exchange: A, pulmonary ventilation; B, CO_2 excretion in expired air (VCO_2); C, oxygen consumption (VO_2) and D, arterial pCO_2. The dashed line indicates onset of anaerobic threshold.

2. Chemical mechanism does not play the main role in exercise hyperpnoea as the alveolar and arterial pO_2 and pCO_2 are well maintained during exercise. Following roles have been suggested, however:

- *Accentuations of the normal oscillations* in pO_2 and pCO_2 synchronous with respiration might stimulate the carotid body chemoreceptors and explains part of exercise hyperpnoea.
- *Acidosis produced due to accumulation of lactic acid* during severe exercise (above the aerobic threshold level) is responsible for the increase in pulmonary ventilation.

3. Humoral mechanisms are not reported to play any role in exercise hyperpnoea.

II. INCREASE IN OXYGEN UPTAKE IN THE LUNGS

The oxygen uptake by the blood in the lungs increases from 250 mL/min at rest to about 4 L/min during heavy exercise. This is made possible by the following changes:

1. Increased pulmonary perfusion. During exercise, about six times more blood passes through the lungs per minute and so more O_2 per minute is carried by the blood from the lungs.

2. Increased alveolar capillary pO_2 gradient. During exercise, because of greater extraction of O_2 by the muscles, the O_2 content of the mixed venous blood reaching the lungs may be as low as 3 mL/100 mL of blood (as compared to 14–15 mL% at rest). Thus the alveolar capillary pO_2 gradient is increased due to a marked desaturation of the venous blood and so more O_2 is taken up by the blood in the lungs.

3. Increased pulmonary diffusion capacity. During exercise, about six fold increase in the pulmonary blood flow due to the opening of several pulmonary capillaries which are closed at rest. As a result, the alveoli are better perfused with blood. The larger number of open up capillaries increases the surface area available for diffusion. In this way, there occurs threefold increase in the diffusion capacity of the lungs.

CHANGES AT THE TISSUE LEVEL

The changes at the tissue level which facilitate transfer of a large amount of O_2 from the blood to the exercising muscles and that of CO_2 from the tissues to the blood are:

Blood flow to the skeletal muscles is increased during strenuous exercise which brings more O_2 to the tissue per minute. *Capillary bed* of the contracting muscles is dilated and many previously closed capillaries are open. Because of these changes the mean distance from the blood to the tissues cells is greatly decreased; this facilitates movement of O_2 from the blood to the cells.

Gradient of pO_2 between capillary blood and tissue fluid is increased. Due to the increased gradient of pO_2 between

capillaries and tissue fluid more O_2 is removed from the capillary blood into the tissue fluid.

O_2–Hb dissociation curve shifts to the right due to accumulation of CO_2, rise in temperature and rise in red blood cell 2,3-DPG. This results in a three fold increase in O_2 extraction from each unit of blood.

ENDOCRINAL RESPONSES TO EXERCISE

Endocrines play an important role in adjustment to exercise. The hormones which are increased during exercise along with the role played by them are given.

1. Antidiuretic hormone (ADH) secretion is markedly increased during exercise. It helps to maintain fluid balance by reducing urine flow.

2. Adrenocorticotrophic hormone is released during the endurance events and probably helps by mobilizing fats for providing energy directly as well as by stimulating glucocorticoid secretion.

3. Endorphin secretion is significantly increased during exercise. They improve tolerance to discomfort associated with exercise by relieving pain. They also relieve mental stress and induce a feeling of well-being.

4. Cortisol secreted during exercise is helpful in reducing exercising stress. It also mobilizes protein and fats. Fatty acids are particularly useful as fuel during exercise. Consequently, carbohydrates are spared to be used by brain.

5. Aldosterone secreted during exercise reduces urinary loss of water and sodium like ADH. This helps to maintain fluid balance in the presence of excessive sweating during exercise.

6. Adrenaline and noradrenaline secretion is increased significantly in intense exercise. These hormones mobilize fatty acids and glucose and thus improve the availability of fuel.

7. Insulin secretion is decreased during exercise. However due to training tissue sensitivity to insulin improves at rest. The improved sensitivity at rest account for the improvement in glucose tolerance test seen as a result of regular exercise. Because of this effect exercise is considered one of the most useful components of treatment of diabetes.

8. Glucagon. Prolonged exercise stimulates secretion of glucagon. It mobilises glucose from glycogen and fatty acids from the adipose tissue, and thereby improves fuel availability during prolonged exercise.

EFFECTS OF TRAINING

Training to the body tissues is provided by different sets of exercise regimes. Endurance training (aerobic) produces different effects in the skeletal muscles. Only aerobic exercises produce cardiovascular conditioning.

Usefulness of training to the body systems is highlighted: Training is most essential for the performance by athletes and sports persons and forms the main aspect of sports physiology.

- Training by regular physical exercise is likely to slow the ageing process, which helps to prevent several degenerative and metabolic diseases and thereby make life healthier and longer.
- Patients who exercise regularly 'feel better'. Such effects may also be attributed to the release of endorphins during exercise which are reported to relieve mental stress and induce a sense of well-being.
- Regular exercise is one of the most useful components of treatment of diabetes because it reduces the insulin requirement by virtue of improvement in the glucose tolerance.
- There is some evidence that regular exercise decreases the incidence and severity of myocardial infarctions.

Effects of training on the body tissues can be described as:

- Effects of training on cardiovascular system,
- Effects of training on respiratory system,
- Effects of training on skeletal muscles,
- Psychological effects of training and
- Metabolic effects of training.

EFFECTS OF TRAINING ON CARDIOVASCULAR SYSTEM

Only aerobic exercises produce cardiovascular conditioning for a heart rate of 60–70% of the maximum for 20–30 min, three to four times a week for at least 3 months. Athletic training produces following effects on cardiovascular system:

1. Low resting heart rate. Therefore, increases the vagal tone. Consequently, trained athletes have low-resting heart (50–60/min). This is useful during exercise because it increases the range through which the heart rate can increase.

2. Higher resting stroke volume. The aerobic athletic training leads to a cardiac hypertrophy and an increase in end-diastolic volume which increases the resting stroke volume to about 105 mL as compared to about 75 mL in an untrained individual.

3. Much larger cardiac output during exercise. Because of the low resting heart rate and higher resting stroke volume, a trained athlete can achieve much larger cardiac output during exercise than an untrained individual as illustrated in Table 5.8-3 by arbitrary but plausible figures.

EFFECTS OF TRAINING ON RESPIRATORY SYSTEM

1. Increase in maximal oxygen consumption

Maximal O_2 consumption (VO_2 max) increases by 5–20% by conditioning (athletic training).

Table 5.8-3 Effects of aerobic training on cardiovascular functions

	At rest		During maximal exercise	
	Untrained individual	Trained athlete	Untrained individual	Trained athlete
Heart rate (beats/min)	72	50	180	180
Stroke volume (mL/beat)	75	100	100	160
Cardiac output (L/min)	5	5	18	29

2. Increase in maximal minute ventilation

- Endurance (aerobic) training increases maximal minute ventilation that is achieved during exercise.
- Specific respiratory muscle training allows one to increase the duration and intensity of exercise.

3. Increase in pulmonary oxygen diffusing capacity

Athletic training allows more increase in diffusion capacity of lungs for oxygen because by training the pulmonary capillary density increases.

EFFECTS OF TRAINING ON SKELETAL MUSCLES

Regular muscular exercise may lead on to following changes in the muscles:

Increased muscle strength due to training results from:

- *Increase in muscle mass* which is entirely the result of an increase in the size of muscle fibres *(hypertrophy* and not due to increase in the number of muscle fibres (hyperplasia).
- *More effective and efficient deployment of motor units.*
- *Increase in the production of contractile proteins,* such as actin and myosin, which is mediated by somatomedins which are required for generating force during the muscle contraction.

Changes in the muscle fibres, which enhance the capacity of the muscles to extract more O_2 and improve the ability of the muscle fibres to provide energy during prolonged exercise, are:

- Increase in the capillary network.
- Increase in the number of mitochondria in the muscle fibres.
- Increase in the mitochondrial enzymes involved during oxidative metabolism.
- Increase in the muscle glycogen stores.
- Increase in the stored triglycerides.

PSYCHOLOGICAL EFFECTS OF TRAINING

Regular training improves the psychology of the individual and thus the psychic stimuli to vasomotor centre and respiratory centres are reduced. Consequently, during exercise there occurs:

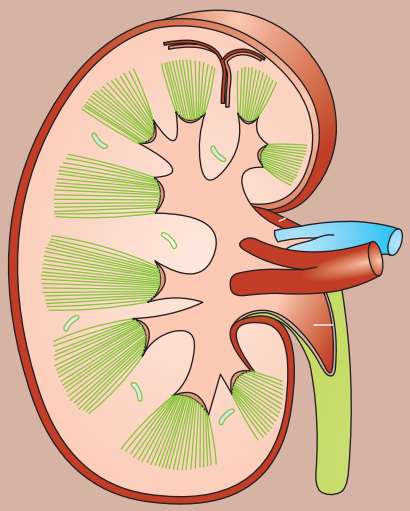
- Less increase in the sympathetic activity and
- Less decrease in the parasympathetic activity.

METABOLIC EFFECTS OF TRAINING

Metabolic adjustments during exercise are the result of increase in energy stores and mitochondrial changes in the muscles described above. Due to these changes the ability of the muscle to extract oxygen improves and there is shift towards aerobic metabolism which is more efficient than the anaerobic metabolism. Consequently, there is less accumulation of lactic acid and smaller fall in the pH of the body fluids. These changes facilitate mobilization of fatty acids from the tissue stores into the blood. Shift in metabolism towards more utilization of fats is a very useful adaptation because fat stores are virtually unlimited as compared to the extremely meager glycogen store. So fat utilization spares glycogen. As physical performance is a direct function of glycogen stores therefore endurance of the individual increases.

SECTION 6

Excretory System



Concept of excretory system: Excretion. Literally, the word excretion means elimination of any matter from the body of an organism. The organs which are involved in the process of excretion include:

- *Kidneys,* which excrete water and water soluble waste products,
- *Lungs,* which excrete carbon dioxide, water vapours and other volatile substances such as acetone,
- *Skin,* which excretes water and salts mainly in the form of sweat and
- *Gastrointestinal tract,* which excretes faeces (excreta).

However, sensu stricto, the term excretion refers to elimination of principal products of metabolism except carbon dioxide. The principal products of metabolism, other than carbon dioxide, are ammonia, urea, uric acid, creatinine, various pigments and inorganic salts.

Excretory organs. Thus, in strictest sense, kidneys are the excretory organs. Together with a pair of ureters and a urinary bladder, kidneys constitute the excretory system.

FUNCTIONS OF EXCRETORY KIDNEY

The kidneys serve several major functions:

1. *Excretory function*. As mentioned above, the kidneys excrete a number of end products of metabolism in urine. Thus, formation of urine is the major function of kidneys. In addition to the metabolic wastes, the kidneys also excrete foreign substances from the body.
2. *Regulation of water and inorganic ion balance*. Control of volume of body fluids and their inorganic ion balance is an important homeostatic role of kidneys.
3. *Regulation of acid-base balance*. Kidneys, in co-ordination with lungs, liver and buffers in the body play a role in regulation of acid-base balance.
4. *Hormonal function*. As an endocrine gland kidneys produce and secrete renin, calcitriol, and erythropoietin.

Chapter 6.1

Kidneys: Functional Anatomy and Blood Flow

FUNCTIONAL ANATOMY OF KIDNEYS

GROSS ANATOMY OF KIDNEY

External features

Gross anatomical features of a human kidney, illustrated in Fig. 6.1-1A, are:

Location. The kidneys are bean-shaped organs that lie retroperitoneally on the posterior abdominal wall, one on each side of the vertebral column at the level of T_{12}–L_1 vertebrae. The right kidney lies slightly inferior to the left kidney.

Size and shape. During life, the kidneys are reddish-brown in colour. Each kidney in an adult human weighs about 150 g and measures approximately 10 cm in length, 5 cm in width and 2.5 cm in thickness.

Renal hilum and sinus. The renal hilum is a vertical cleft present on the concave medial margin. It is the entrance to space within the kidney—the renal sinus. Through renal hilum the renal artery enters, and the renal vein and renal pelvis leave the renal sinus. The *renal sinus* is thus occupied by the renal pelvis, calyces, vessels, nerves and a variable amount of fat.

Renal pelvis and calyces. The renal pelvis is the flattened, funnel-shaped expansion of the superior end of ureter. Within the renal sinus, the pelvis divides into two (or three) parts called *major calyces.* Each major calyx divides into a number of *minor calyces.* The end of each minor calyx is shaped like a cup into which fits a projection of kidney tissue called renal papilla (the apex of renal pyramid).

Gross internal structure

Gross internal structure of the kidney, as seen in the coronal section through the organ, exhibits that kidney tissue consists of an outer region called the cortex and an inner region called the medulla (Fig. 6.1-1B).

Medulla. It is made up of triangular areas of renal tissue that are called the renal pyramids. Pyramids are 4–14 in number and separated from each other by cortical columns of Bertin. Each pyramid has a base directed towards the cortex and an apex (or renal papilla) which is directed towards the renal pelvis and fits into the minor calyx. Pyramids show striations that pass radially towards the apex. These striations are due to a straight portion of the nephron and extend some distance upwards into the cortex where they are called medullary rays. The medulla can be subdivided into two parts:

- *Outer medulla* that is further subdivided into the outer stripe and the inner stripe.
- *Inner medulla* is also called papillary zone.

Cortex. The renal cortex can be divided into two parts which are continuous with each other:

- *Cortical arches* or cortical lobules refer to the tissue lying between the bases of pyramids and surface of the kidney.

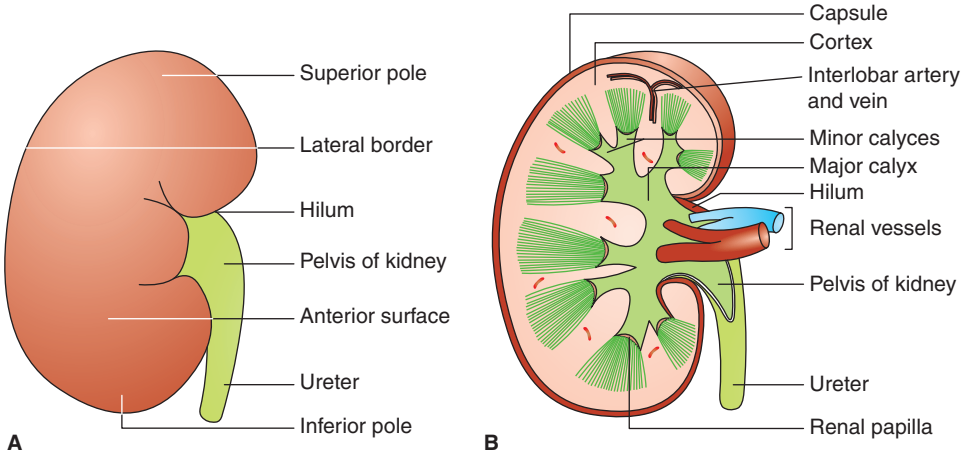


Fig. 6.1-1 A, gross anatomical features and B, coronal section through human kidney.

- *Renal columns* refer to the cortical tissue that lies in between the pyramids.
- *Lobe of kidney.* Each pyramid, surrounded by a shell of cortex constitutes a lobe of the kidney.

MICROSCOPIC STRUCTURE OF KIDNEY

Microscopically, the cortex and medulla of the kidney are composed of nephrons, blood vessels, lymphatics and nerves.

Nephron is a structural and functional unit of the kidney. Each kidney contains approximately 1.2 million nephrons. Each nephron is capable of forming urine.

Structure of the nephron

A nephron consists of two major parts (Fig. 6.1-2):

- Renal corpuscle and
- Renal tubule.

Renal corpuscle

Renal corpuscle or Malpighian corpuscle is a rounded structure comprising glomerulus surrounded by a glomerular capsule (Fig. 6.1-3A).

Glomerulus. Glomerulus refers to a rounded tuft of anastomosing capillaries. Blood enters the glomerulus through an *afferent arteriole* and leaves it through an *efferent arteriole* (note that the efferent vessel is an arteriole and not a venule).

Glomerular capsule. Glomerular capsule, also known as Bowman's capsule, encloses the glomerulus and is formed of two layers: the inner layer covering the glomerular capillaries is called *visceral layer* and the outer layer is called *parietal layer.* In fact, the Bowman's capsule represents the cup-shaped blind end in the beginning of the renal tubule. Space between the visceral and parietal layer of the capsule (called Bowman's space or urinary space) is continuous with the lumen of the renal tubule.

Ultrastructure of glomerular membrane. *Glomerular membrane refers to the membrane that separates blood of the glomerular capillaries from the* fluid present in the Bowman's space. It is also called filtration barrier and consists of three major layers (Fig. 6.1-3B):

1. Capillary endothelium. It is fenestrated (i.e. it contains pores with diameter 70–90 nm) and is freely permeable to water to small solutes and even to small proteins.

2. Basement membrane. It consists of a matrix of glycoproteins and mucopolysaccharides. As compared to the typical membranes, the glomerular basement membrane is very thick.

No pores have been demonstrated in the basement membrane; however, its permeability corresponds to pore size (of about 8 nm).

3. Bowman's visceral epithelium or the inner layer of Bowman's capsule, which forms the third layer of glomerular membrane, is formed by special cells called podocytes. The podocytes have finger-like processes that encircle the outer surface of capillaries. The processes of podocytes interdigitate to cover the basement membrane and are separated by gaps called the filtration slits (of approximately 25 nm diameter). Each filtration slit is covered by a layer of fine filaments that constitute the diaphragm.

Mesangium is an important component of renal corpuscle, it consists of mesangium cells that are present between the capillary endothelial cells and the basement membrane. These cells provide a structural support for the glomerular capillaries, secrete the extracellular matrix and exhibit phagocytic activity.

Renal tubule

Renal tubule is a long complicated tubule that is divisible into the following main parts (Fig. 6.1-2):

1. Proximal tubule. The proximal tubule initially forms several coils, proximal convoluted tubule, followed by a

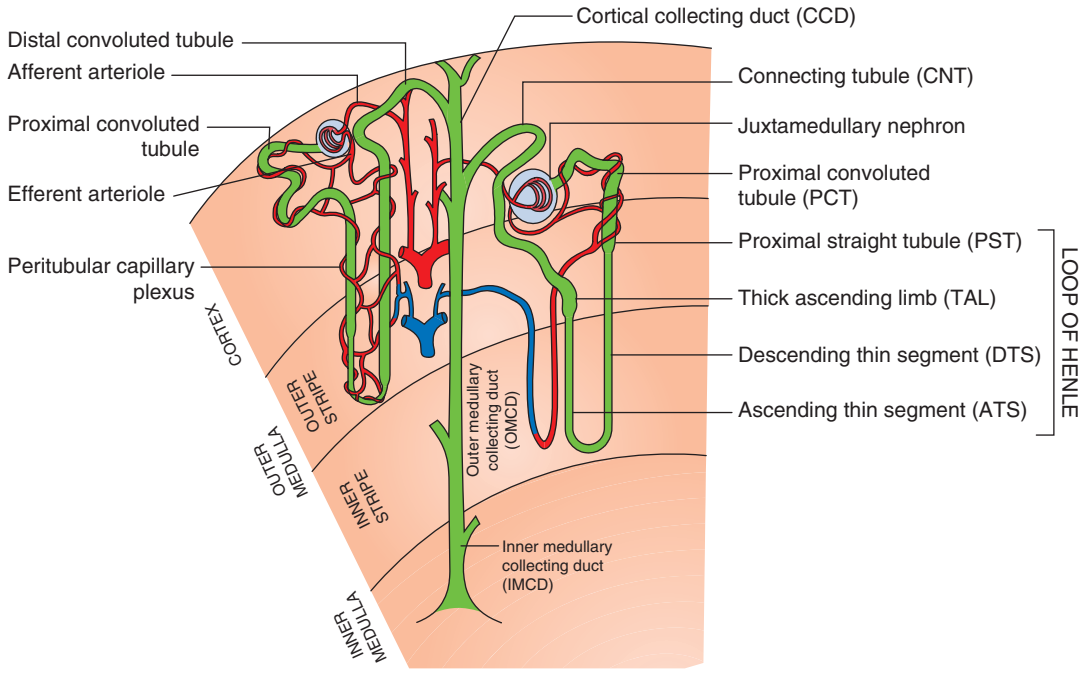


Fig. 6.1-2 Parts of a typical nephron. Organization of cortical and juxtamedullary nephrons showing different parts. Note differences between two types of nephrons.

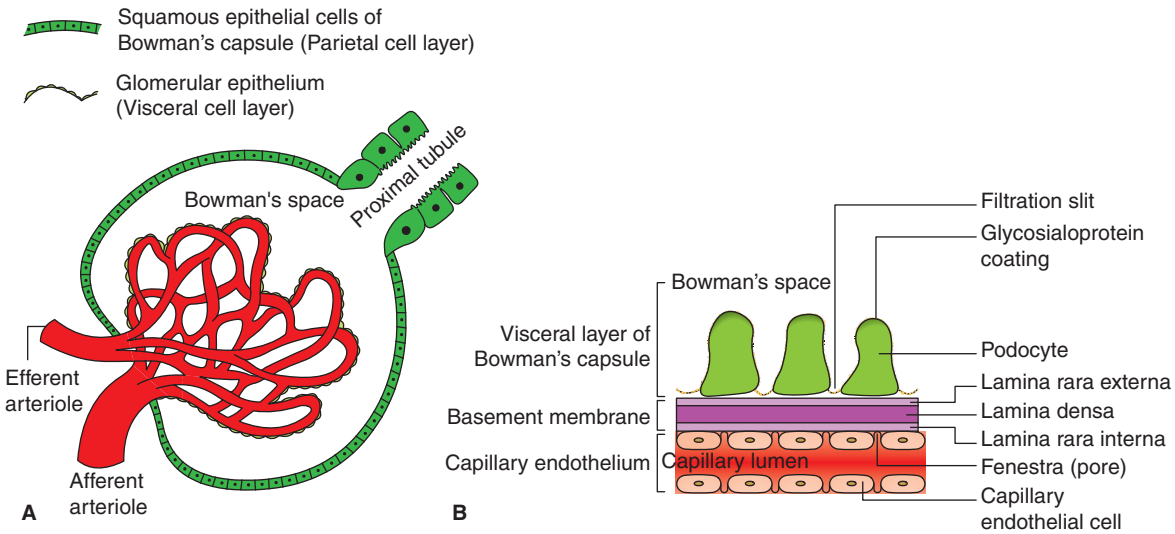


Fig. 6.1-3 A, structure of a glomerulus and B, glomerular membrane.

straight segment, proximal straight tubule or pars recta that descend towards the medulla.

2. Intermediate tubule or loop of Henle that consists of:

- Descending thin segment (DTS),
- Ascending thin segment (ATS), and
- Thick ascending limb (TAL).

In juxtamedullary nephrons, the DTS joins ATS to form the hair pin band (loop). The ATS reaches up to the junction of outer and inner medulla. In *cortical nephrons,* there is no ATS, the DTS is continuous at the bend of loop with the TAL. Near the end of the TAL, the nephron passes between its afferent and efferent arteriole. This short segment of the TAL is called the *macula densa.*

3. Distal convoluted tubule. It begins a short distance beyond the macula densa and extends to a point in the cortex when the connecting tubules of two or more nephrons join to form the cortical collecting ducts.

4. Collecting duct. The collecting duct is divisible into three parts:

- *Cortical collecting duct*, i.e. the portion present in the cortex,
- *Outer medullary collecting duct*, i.e. the portion present in the outer medulla and
- *Inner medullary collecting duct* (IMCD), i.e. the portion present in the inner medulla. Several IMCDs coalesce together before finally opening at the tip of the renal papilla.

Characteristics of epithelium lining the renal tubule. The epithelium lining the different segments of renal tubule has some special characteristic features which are suited to perform specific transport functions (Figs 6.1-2 and 6.1-4).

Type of cells. The cells lining the renal tubule are mostly cuboidal except in the thin segment where these are flat or squamous type.

Apical surface of cuboidal cells bear a few microvilli in general, which are numerous, dense and amplified in proximal tubule cells to form the so-called brush border.

Basolateral membrane of the proximal convoluted tubule cells, thick ascending segment cells and distal convoluted tubule cells is highly invaginated and contain many mitochondria. These infoldings create basal spaces. In contrast, the cells of the descending thin limb and ascending thin limb of loop of Henle have poorly developed basolateral surfaces and contain a few mitochondria.

Lateral surfaces of the cells of renal tubules bear the lateral cell process which interdigitate with lateral processes of the adjacent cells. Lateral intercellular spaces are present in between the interdigitations. Lateral intercellular spaces do not communicate with the basal spaces. The lateral surfaces of cells form two types of tight junctions:

- *Leaky tight junctions* that permit water and solutes to diffuse across them. These are present in a *proximal tubule.*

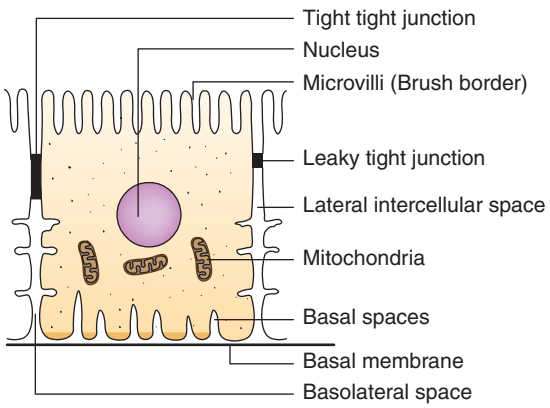


Fig. 6.1-4 Ultrastructure of an epithelial cell lining the proximal convoluted tubule.

- *Tight tight junctions* that do not permit water and solutes to diffuse across them easily. They are present in the *distal tubule.*

Cortical collecting duct is composed of two cell types:

- *Principal cells* have a moderately invaginated basolateral membrane and contain few mitochondria.
- *Intercalated cells* have a high density of mitochondria.

Inner medullary collecting duct is composed of a single layer of cells that have poorly developed apical and basolateral surfaces and a few mitochondria.

Types of nephrons

There are two types of nephrons: cortical (superficial) and juxtamedullary. Differences between the cortical and juxtamedullary nephrons are depicted in Fig. 6.1-2 and Table 6.1-1.

Juxtaglomerular apparatus

Juxtaglomerular (JG) apparatus as the name indicates (juxtanear) refers to the collection of specialised cells located very near to the glomerulus. It forms the major component of renin–angiotensin–aldosterone system. The JG apparatus comprises three types of cells (Fig. 6.1-5):

- Juxtaglomerular cells,
- Macula densa cells and
- Mesangial cells.

1. Juxtaglomerular cells. JG cells are specialised *myoepithelial* (modified vascular smooth muscle) cells located in the media of the *afferent arteriole* in the region of JG apparatus.

Characteristic features of JG cells are:

- They have well-developed Golgi apparatus and endoplasmic reticulum, abundant mitochondria and ribosomes.
- They synthesize, store and release an enzyme called *renin.* Renin is stored in the secretory granules of JG cells and, therefore, these are also called *granular cells.*
- They act as *baroreceptors* (tension receptors) and respond to changes in the transmural pressure gradient between the afferent arterioles and the interstitium.
- They are densely innervated by the *sympathetic nerve* fibres and release their renin content in response to the sympathetic discharge.
- As these cells act as vascular volume receptors, they monitor renal perfusion pressure and are stimulated by hypovolaemia or decreased renal perfusion pressure.

2. Macula densa cells. Macula densa cells refer to the specialised renal tubular epithelial cells of a short segment of the thick ascending limb of loop of Henle which passes between the afferent and efferent arterioles supplying its glomerulus of origin.

Table 6.1-1 Differences between cortical and juxtamedullary nephron

Feature	Cortical nephron	Juxtamedullary nephron
Location of glomerulus	Upper region of cortex	Near junction of cortex and medulla
Percentage of total nephron	85%	15%
Size of glomeruli	Small	Larger
Size of loop of Henle	Small, extend up to outer layer of medulla	Large, extend deep into the medulla
Descending limb of loop of Henle comprises	Thin segment	Thin segment
Ascending limb of loop of Henle comprises	Thick segment	Thin segment
Efferent arterioles	Have large diameter and break-up into peritubular capillaries	Have small diameter and continue as vasa recta
Rate of filtration	Slow	High
Major function	Excretion of waste products in urine	Concentration of urine by the counter current system

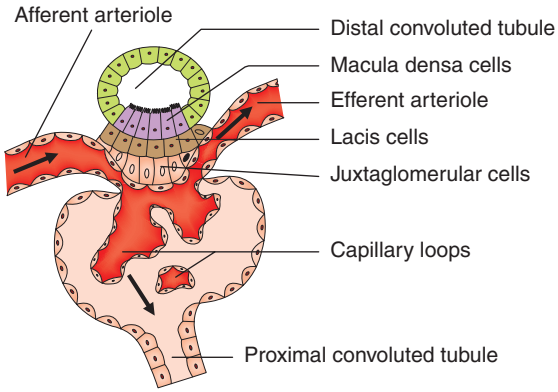


Fig. 6.1-5 Juxtaglomerular apparatus.

Characteristic features of macula densa cells are:

- They are not well adapted for reabsorption.
- They are not innervated.
- These cells are in direct contact with the mesangial cells and in close contact with the JG cells.
- They act as *chemoreceptors* and are stimulated by decreased NaCl concentration and thereby causing increased renin release.

3. Mesangial cells Mesangial cells or *lacis* cells are the interstitial cells of the JG apparatus.

Characteristic features of these cells are:

- They are in contact with both the macula densa cells (on one side) and JG cells (on the other side).
- Functionally, these cells possibly relay the signals from macula densa to the granular cells after modulating the signals. In this way, a decreased intraluminal Na^+ load, Cl^- load, or both in the region of macula densa stimulates the JG cells to secrete renin.
- They also show granulation to secrete renin in conditions of extreme hyperactivity.
- They also secrete various substances and take up immune complexes.

RENAL BLOOD FLOW

RENAL BLOOD VESSELS

Arrangement of arterial vessels (renal artery and its branches) in the kidney (Fig. 6.1-6)

Renal artery (one for each kidney), a major branch from the aorta, divides into a number of lobular arteries at the hilum of kidney.

Lobular artery (one for each pyramid) divides into two or more interlobar arteries.

Interlobar arteries enter the tissue of the renal columns and run towards the surface of kidney. Reaching the level of the bases of the pyramids, the interlobar arteries divide into arcuate arteries.

Arcuate arteries run at right angles to the parent interlobar arteries. They lie parallel to the renal surface at the junction of pyramid and cortex. They give a series of interlobular arteries.

Interlobular arteries run through the cortex at right angles to the renal surface to end in a subcapsular plexus. It has been held that interlobular arteries divide the renal cortex into small lobules. Each interlobular artery gives off a series of afferent arterioles.

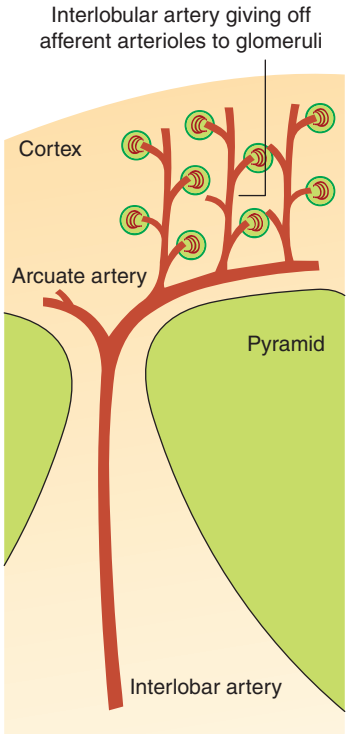


Fig. 6.1-6 Scheme to show arrangement of arteries within the kidney.

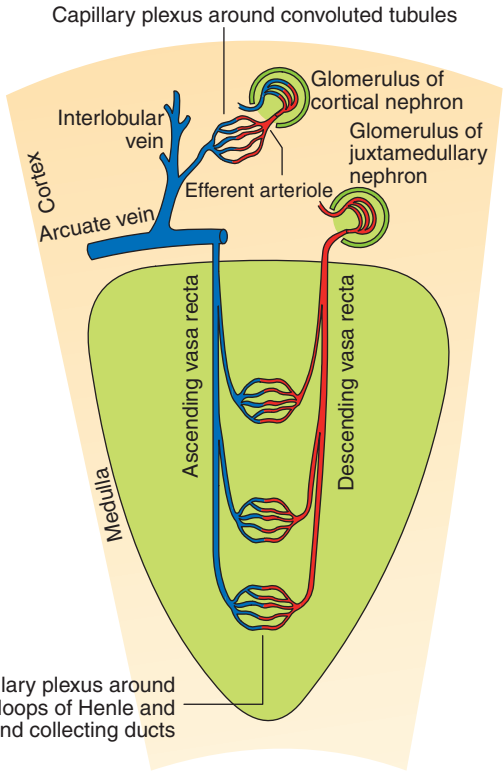


Fig. 6.1-7 Scheme to show behaviour of efferent arterioles arising from the glomeruli of cortical and juxtamedullary nephrons.

Afferent arterioles. Each afferent arteriole enters the Bowman's capsule and divides into a rounded tuft of anastomosing capillaries called *glomerulus.* As mentioned earlier, this capillary network has special features owing to which it works as a sieve allowing plasma filtration with retention of plasma proteins and blood cells. The glomerular capillaries join to form the efferent arteriole.

Efferent arterioles leaving the glomeruli of two types of nephrons exhibit different behaviour:

- *Efferent arterioles* arising from the cortical nephrons divide into the *peritubular capillaries* that surround the proximal and distal convoluted tubule forming a rich meshwork of microvessels. These capillaries drain into the interlobular veins.
- *Efferent arterioles arising from the juxtamedullary nephrons* give rise to *vasa recta.* The *vasa recta* descend with the long loops of Henle into renal medulla and return to the area of the glomerulus and drain into interlobular or arcuate vein.

Side branches arising from the vasa recti form a capillary network at different levels along the loop of Henle (Fig. 6.1-7).

Arrangement of venous vessels (renal veins)

The pattern of renal venous system is similar to that found in the end arterial system, except for the presence of multiple anastomoses between veins at all levels of the venous circulation. The corresponding veins which run parallel to the arterial vessels are the interlobular veins, the arcuate veins, the interlobar veins and the renal veins which exit the kidney at hilus.

Innervation of kidney

Renal vessels are innervated by sympathetic and parasympathetic fibres.

Parasympathetic innervation is by vagus nerve, but its function is uncertain.

Sympathetic innervation. Pre-ganglionic sympathetic fibres arise from the neurons of lower thoracic and upper lumbar (T_{10}–L_2) intermediolateral segments of spinal cord. The cell bodies of the post-ganglionic neurons are located in the ganglia of sympathetic chain and superior mesenteric ganglion. The fibres from these neurons are carried by the renal nerves, which travel along the renal blood vessels as they enter the kidney. The efferent fibres are mainly distributed to afferent and efferent arterioles, cells of renal tubule and also to JG cells.

Afferents from the kidney (afferents of renorenal reflex and pain fibres) run along with the efferent fibres and enter in the spinal cord through the thoracic and upper lumbar dorsal roots.

Note. *Renorenal reflex.* An increase in ureteral pressure of one kidney reflexly reduces efferent nerve activity of

contralateral kidney and causes increase excretion of sodium and water is known as renorenal reflex.

CHARACTERISTICS OF RENAL BLOOD FLOW

Amount and rate of blood flow

- *Rate of renal blood flow* under basal conditions, approximately 1200 mL/min (400 mL/100 g tissue/min) is very high compared to other tissues.
- *Total renal blood flow* is approximately 20% of resting cardiac output, while the two kidneys make <0.5% of total body weight.
- *Higher blood flow* to kidneys is related to its excretory function rather than its metabolic requirement.
- *In face of blood pressure changes* the renal blood flow shows remarkable constancy due to autoregulation.
- *During exercise,* sympathetic tone to renal vessels increases and shunts renal blood flow to the skeletal muscles.

Renal blood flow and oxygen consumption

- *Renal* O_2 consumption (approximately 6 mL/100 g tissue/min) is very high being only second to heart (i.e. 8 mL/100 g tissue/min) in the body.
- Arteriovenous O_2 difference (approximately 1.5 mL/dL) of blood is smallest of the major organ system.
- Oxygen consumption (VO_2) in kidneys is directly proportional to the renal blood flow. Thus, unlike other organs, where the blood flow is related to O_2 requirements of the organ, in the kidneys, the O_2 consumption is a function of blood flow.

Hydrostatic pressure in renal vessels and their physiological significance

- ***Glomerular capillaries*** have relatively high hydrostatic pressure (45 mm Hg), which is an important factor in the formation of glomerular filtrate.
- ***Peritubular capillaries*** have very low hydrostatic pressure (8 mm Hg only) due to drop in the pressure in efferent arterioles. This low hydrostatic pressure in the peritubular capillaries facilitates the reabsorptive function of the proximal and distal *convoluted* tubules.
- ***Renal veins*** have hydrostatic pressure of only 4 mm Hg.

Renal portal circulation

It has been mentioned above that the afferent arterioles arise from the interlobular arteries and each breaks up into a bunch of capillaries called the glomerulus. The glomerular capillaries drain into the efferent arterioles which again break up into a peritubular capillary network which ultimately drains into an interlobular vein.

In this way, two sets of capillaries are formed and thus renal circulation becomes a sort of portal circulation.

REGULATION OF RENAL BLOOD FLOW

The regulatory mechanisms affect the renal blood flow (RBF) and glomerular filtration rate (GFR) by changing the arteriolar resistance. Therefore, before discussing the various regulatory mechanisms it will be worthwhile to understand the *relationship between selective changes in the resistance of afferent and/or efferent arterioles and RBF and GFR* (Fig. 6.1-8):

- *Constriction of afferent arteriole* decreases both RBF and GFR without change in the filtration fraction (FF) (Fig. 6.1-8A).
- *Dilatation of the afferent arteriole* increases both RBF and GFR without change in the FF (Fig. 6.1-8B).
- *Constriction of the efferent arteriole* decreases the RBF and increases GFR and FF (Fig. 6.1-8C).
- *Dilatation of the efferent arteriole* increases the RBF and decreases the GFR and FF (Fig. 6.1-8D).

Regulatory mechanisms of renal blood flow include:

- Autoregulation,
- Hormonal regulation and
- Nervous regulation.

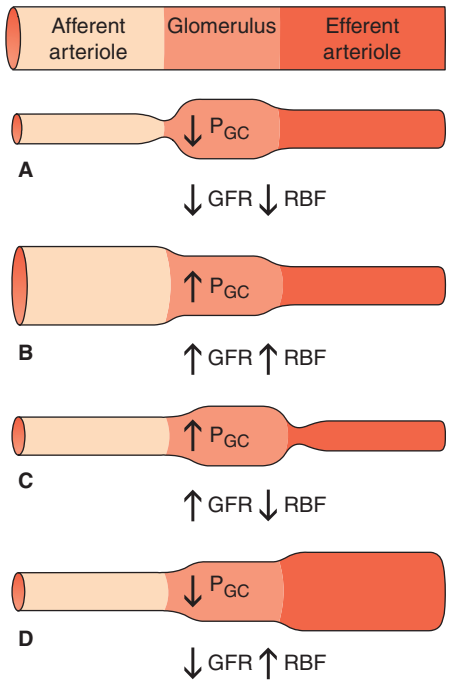


Fig. 6.1-8 Relationship between selective changes in arteriolar resistance and of either afferent or efferent arteriole and renal blood flow (RBF) and glomerular filtration rate (GFR). (P_{GC} = Hydrostatic pressure in glomerular capillaries.)

Autoregulation of renal blood flow

The RBF and thus the GFR remain constant over a wide range of renal arterial pressures (80–200 mm Hg) (Fig. 6.1-9). This occurs due to an intrarenal mechanism known as autoregulation.

Mechanisms of autoregulation

Two mechanisms are considered responsible for the autoregulation of RBF and GFR: one mechanism that responds to changes in the arterial pressure and another that responds to changes in NaCl concentration of tubular fluid.

1. Myogenic mechanism. It is related to an intrinsic property of vascular smooth muscle: the tendency to contract when it is stretched. Thus, when renal arterial pressure is raised, the afferent arterioles are stretched, which contract and increase the vascular resistance. The increased vascular resistance offsets the effect of increased arterial pressure and thereby maintains a constant RBF and GFR (Fig. 6.1-9).

2. Tubuloglomerular feedback mechanism. Tubuloglomerular feedback (TGF) mechanism is based on the NaCl concentration of tubular fluid. It involves a feedback loop which operates as (Fig. 6.1-10):

- Changes in the GFR cause changes in the NaCl concentration of fluid in the loop of Henle.
- Changes in the NaCl concentration are sensed by the macula densa cells and converted into a signal.
- The signal from the macula densa cells changes the vascular resistance in the afferent arterioles.
- Signals obtained due to an increased concentration of NaCl produce vasoconstriction; conversely, signals obtained due to decreased NaCl cause vasodilatation of afferent arterioles.
- The effector mechanism responsible for vasoconstriction and vasodilatation is not exactly known. Perhaps, adenosine triphosphate (ATP), which selectively constricts the afferent arterioles and metabolites of arachidonic acid, may contribute to TGF mechanism.

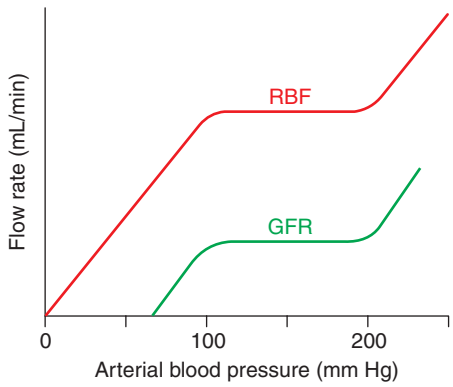


Fig. 6.1-9 Autoregulation maintains renal blood flow (RBF) and glomerular filtration rate (GFR) constant over a wide range of arterial pressure (80–200 mm Hg).

Physiological significance and certain important facts about autoregulation

Physiological significance. A small change in GFR has great effect on urinary output and therefore on loss of solutes and water. If RBF and GFR were to change suddenly in proportion to change in blood pressure, urinary excretion of fluid and solute would also change suddenly. Such changes in water and solute excretion, without comparable alterations in intake, would prove disastrous due to alterations in fluid and electrolyte balance. Thus, autoregulation of RBF and GFR is an effective mechanism for uncoupling renal function from fluctuations in the arterial pressure and maintain fluid and electrolyte balance.

Certain important facts about autoregulations to be noted are:

- Autoregulation of RBF and GFR is virtually absent at the mean arterial blood pressure below 80 mm Hg,
- Autoregulation is not a perfect mechanism, i.e. RBF and GFR do change slightly with variation in the arterial blood pressure and
- Several hormones and other factors can change RBF and GFR, despite autoregulation mechanisms.

Hormonal regulation

As mentioned above, despite autoregulation, several hormones and other factors have a major effect on RBF and

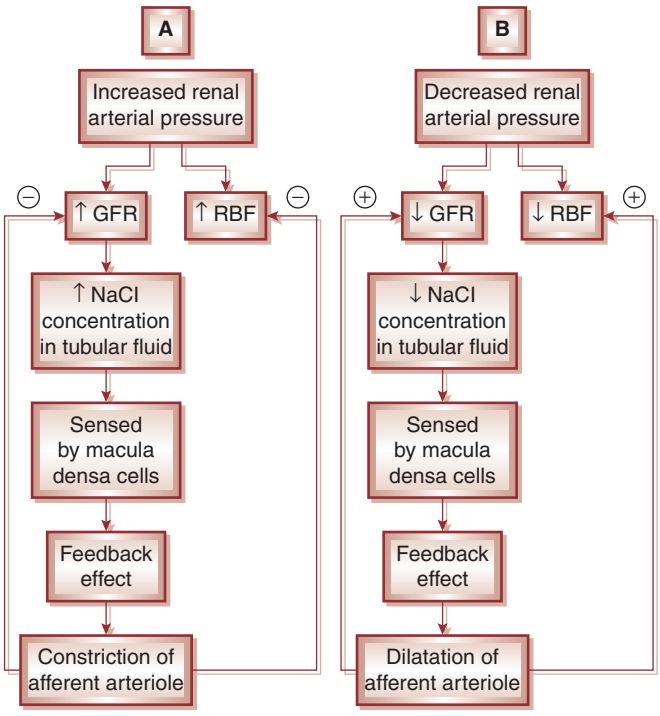


Fig. 6.1-10 A, tubuloglomerular feedback mechanism which maintains a constant RBF and GFR when renal arterial pressure increases and B, decreases.

Table 6.1-2 Hormones influencing the RBF and GFR

Hormone	Effect on GFR	Effect on RBF
Vasoconstrictors		
• Norepinephrine	↓	↓
• Angiotensin II	↓	↓
• Endothelin	↓	↓
Vasodilators		
• Prostaglandins (PGI_2 and PGE_2)	NC	↑
• Nitric oxide (NO)	↑	↑
• Bradykinin	↑	↑
• ANP	↑	↑
• Glucocorticoids	↑	↑
• Dopamine	NC	↑
• Histamine	NC	↑

GFR by affecting afferent and/or efferent arteriolar resistance (Table 6.1-2).

1. Hormones that cause vasoconstriction, and thereby decrease RBF and GFR include:

- *Norepinephrine* causes an intense vasoconstriction of both afferent and efferent arterioles.
- *Angiotensin II,* in low concentrations, causes a predominant constriction of the efferent arterioles. However, at higher concentrations, it causes constriction of both afferent as well as efferent arterioles.
- *Endothelin* causes profound vasoconstriction of the afferent and efferent arterioles. It is secreted by the endothelial cells of renal vessels, mesangial cells and distal tubular cells.

2. Hormones that cause vasodilatation and thereby increase RBF and GFR include:

- Prostaglandins,
- Nitric oxide (NO),
- Bradykinin,
- Atrial natriuretic peptide (ANP),
- Glucocorticoids,
- Dopamine and
- Histamine.

Nervous regulation

Under normal circulatory conditions, sympathetic tone is minimum.

Mild-to-moderate stimulation of sympathetic nerves usually has mild effects on RBF because of autoregulation mechanism.

Strong acute stimulation of sympathetic nerves may produce marked fall in RBF (even to 10–30% of normal) temporarily due to constriction of both afferent and efferent arterioles. This effect is mediated mainly by α_1-adrenergic receptors and to a lesser extent by post-synaptic α_2-adrenergic receptors.

Note. This system works to preserve arterial pressure at the expense of maintaining normal RBF in conditions of acute hypotension due to severe haemorrhage. Further, an increase in sympathetic activity also increases the release of epinephrine and angiotensin-II, enhancing vasoconstriction (vide infra).

MEASUREMENT OF RENAL BLOOD FLOW/RENAL PLASMA FLOW

Renal blood flow can be measured by the following methods:

1. With the help of electromagnetic flow meter (see page 234).
2. *Para-amino hippuric acid (PAH) clearance method.* RBF as well as renal plasma flow can be measured by this method, see page 437.
3. Renal blood flow can also be measured indirectly from the filtration fractions (see page 388).

Chapter 6.2

Mechanism of Urine Formation: Glomerular Filtration and Tubular Transport

INTRODUCTION
- Processes concerned with urine formation

GLOMERULAR FILTRATION
- Characteristics of filtration membrane
- Composition of glomerular filtrate
- Dynamics of glomerular filtration
- Normal glomerular filtration rate
- Filtration fraction
- Factors affecting glomerular filtration rate
- Measurement of glomerular filtration rate

TUBULAR REABSORPTION AND SECRETION
- **General principles of renal tubular transport**
  - Transport mechanisms across cell membrane
  - Transepithelial transport pathways
  - Tubular mechanisms, patterns of renal handling of substances and concept of renal clearance
  - Quantification of renal tubular transport
  - Tubular fluid concentration/plasma concentration ratio
- **Transport across different segments of renal tubule**
  - Transport across proximal tubule
  - Transport across loop of Henle
  - Transport across distal tubules and collecting duct
- **Renal handling of common solutes and water**
  - Renal handling of sodium and water
  - Renal handling of potassium
  - Renal handling of glucose
  - Renal handling of proteins, peptides and amino acids
  - Renal handling of urea
  - Renal handling of uric acid
  - Renal handling of para-amino hippuric acid

INTRODUCTION

The main function of the kidneys is to clear waste products from the blood and excrete them in the urine. The kidneys accomplish their excretory function by the formation of urine.

Processes concerned with urine formation. Three processes are involved in the urine formation (Fig. 6.2-1):

- Glomerular filtration,
- Tubular reabsorption and
- Tubular secretion.

Thus, the excretion of each substance in the urine involves a specific combination of filtration, reabsorption and secretion as expressed by the following relationship: Urinary excretion = Filtration − Reabsorption + Secretion (Fig. 6.2-1).

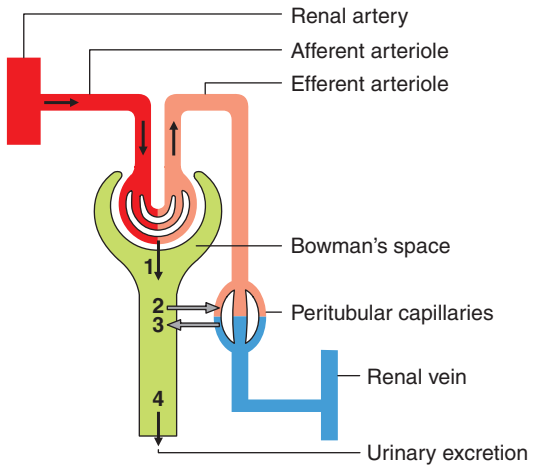


Fig. 6.2-1 Steps involved in the formation of urine: 1, filtration; 2, reabsorption; 3, secretion and 4, excretion.

GLOMERULAR FILTRATION

Glomerular filtration refers to the process of ultrafiltration of plasma from the glomerular capillaries into the Bowman's capsule. The understanding of process of glomerular filtration involves a review of:

- Characteristics of filtration membrane,
- Composition of glomerular filtrate,
- Dynamics of glomerular filtration,
- Glomerular filtration rate,

- Filtration fraction,
- Factors affecting glomerular filtration and
- Measurement of glomerular filtration.

CHARACTERISTICS OF FILTRATION MEMBRANE

As described on page 378 (Fig. 6.1-3B), the filtration membrane consists of three layers: capillary endothelium, glomerular basement membrane (GBM) and Bowman's visceral epithelium (podocytes). Characteristic features of the filtration membrane are:

High permeability. The glomerular membrane is highly permeable to water and 100% dissolved substances because of its porous nature.

Permeability selectivity. The filtration membrane exhibits a high degree of permeability selectivity based on two factors, i.e. pore size and electrical charge of the molecule.

- *Pore size.* The capillary endothelial cells have pores that are 70–90 nm in diameter, the GBM has no pores but its permeability corresponds to pore size of 8 nm.
  - Molecules less than 4 nm in size are freely filtered.
  - Molecules with diameter more than 8 nm are not filtered at all (i.e. zero permeability).
  - Filtration of molecules having diameter between 4 and 8 nm is inversely proportional to their diameter.
- *Electrical charge.* The pores in the filtration membrane are negatively charged due to the presence of glycoproteins rich in sialic acid (sialo proteins). Thus, with the same molecular size, compared to anionic particles, there is, in order, increasing permeability for neutral and cationic particles (Fig. 6.2-2). This explains why albumin (with a molecular diameter of 7 nm but a negative charge) is not filtered.

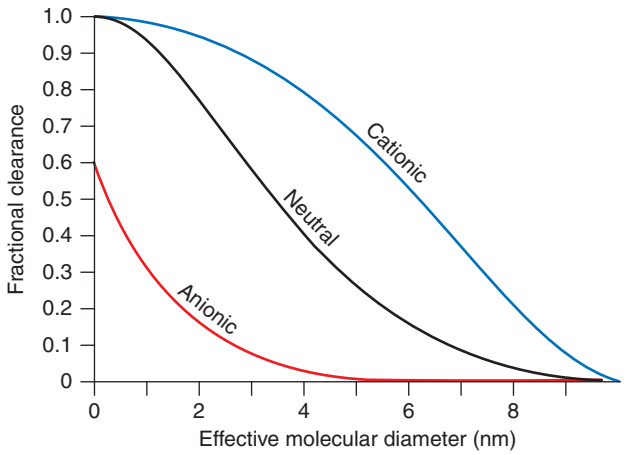


Fig. 6.2-2 Effect of electrical charge and effective molecular diameter on filterability of dextran molecule through glomerular filtration membrane. A value of one indicates that it is filtered freely, whereas a value of zero indicates that it is not filtered.

COMPOSITION OF GLOMERULAR FILTRATE

The unique characteristic features of the glomerular filtration membrane determine the composition of the glomerular filtrate, in that it is like that of the plasma except for absence of proteins (colloids) and cells. It is important to note that normally the amount of proteins in the urine is less than 100 mg/dL, and most of this is not filtered but comes from the shedded tubular cells. Filtration membrane permeability alteration in diseases, however, may alter diffusibility of colloids and cells. As a result, filtration of proteins is increased and albumin appears in the urine in significant amount *(albuminuria* or *proteinuria).*

DYNAMICS OF GLOMERULAR FILTRATION

The forces which determine the bulk flow or ultrafiltration of protein-free plasma across the glomerular membrane are the same which determine the formation of tissue fluid.

The glomerular filtration rate (GFR) will depend upon the balance of Starling forces (Figs 6.2-3 and 6.2-4). According to the Starling hypothesis, the GFR can be expressed as

$$GFR = Kf\,[(P_{GC} - P_{BS}) - (\pi_{GS} - \pi_{BS})],$$

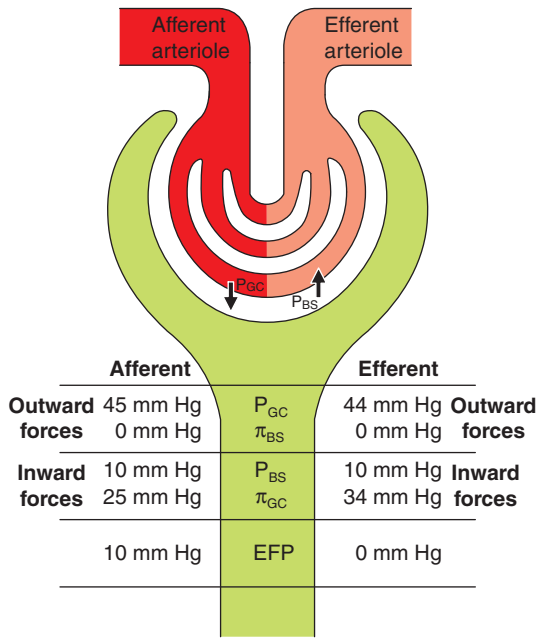


Fig. 6.2-3 Depiction of the Starling forces across the glomerular filtration membrane. (P_{GC} = glomerular capillary hydrostatic pressure; P_{BS} = Bowman's space hydrostatic pressure; π_{GC} = glomerular capillary oncotic pressure; π_{BS} = Bowman's space oncotic pressure.)

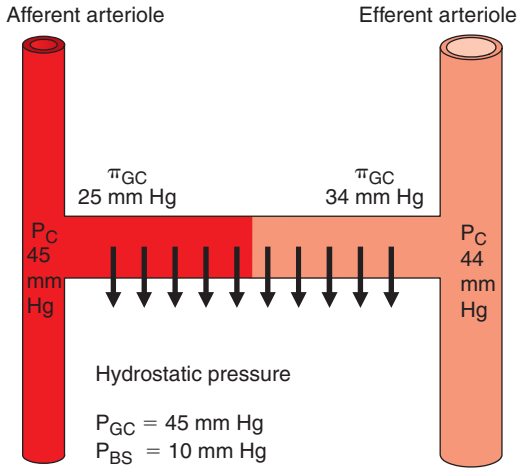


Fig. 6.2-4 Balance of the Starling forces in glomerular capillary.

where

GFR is the filtration across the glomerular membrane.
Kf or the filtration coefficient of the glomerular membrane. The filtration coefficient (Kf) normally equals 12.5 m^2/min/mm Hg.
P_{GC} is glomerular capillary hydrostatic pressure. Its normal value is about 45 mm Hg.
P_{BS} or Bowman's space hydrostatic pressure. Its normal value is about 10 mm Hg.
π_{GC} or glomerular capillary oncotic pressure. Its normal value is 25 mm Hg.
π_{BS} or Bowman's space oncotic pressure. Its normal value is zero because glomerular filtrate contains no proteins.

Effective filtration pressure is the net outward force and is calculated as the difference between the outwardly (i.e. P_{GC} and π_{BS}) and inwardly (P_{BS} and π_{GC}) directed forces (Fig. 6.2-4). Thus under normal circumstances,

$$GFR = 12.5\ (45 - 10) - (25 - 0) = 125\ mL/min$$

NORMAL GLOMERULAR FILTRATION RATE

As calculated above, the normal GFR in an average sized man is about 125 mL/min (range 90–140 mL/min). Its values in women are 10% lower than those in men. Thus, in a 24 h period, as much as 180 L/day of plasma is filtered at the glomerulus. Of the 180 L/day of glomerular filtrate which passes through the remaining part of the nephron, 99% or more is reabsorbed and only 1% or less is excreted as urine. After age of 30 years, GFR declines with age.

FILTRATION FRACTION

The filtration fraction (FF) is the ratio of GFR to the renal plasma flow (RPF).

At normal values of GFR 125 mL/min and RPF 650 mL/min; the FF is approximately 0.2 (125/650). In other words, normally only about 20% the renal plasma flow is actually filtered per minute.

Table 6.2-1 Agents causing contraction and relaxation of mesangial cells

Contraction of mesangial cells (decrease GFR)	Relaxation of mesangial cells (increase GFR)
Angiotensin-I	Atrial natriuretic peptide (ANP)
Antidiuretic hormone (ADH)	Cyclic AMP
Endothelins	Dopamine
Histamine	Prostaglandins (PGE_2)
Leukotrienes–C_4 and D_4	Nitric oxide (NO)
Prostaglandins (PGF_2)	
Platelet derived growth factor (PDGF)	
Platelet activating factor	
Thromboxane-A_2	

FACTORS AFFECTING GLOMERULAR FILTRATION RATE

1. Filtration coefficient (Kf). Increased Kf raises GFR and decreased Kf reduces GFR. As mentioned earlier, Kf is the product of permeability and filtration area of the glomerular capillary membrane. Therefore,

(i) Permeability of the glomerular capillaries is increased in abnormal conditions like hypoxia and presence of toxic agents. In such conditions GFR is increased because plasma proteins are also filtered to a variable degree. Decreased capillary permeability occurs due to thickening of capillary membrane in some diseases leading to decreased GFR.

(ii) Alteration in GFR filtration area of glomerular capillaries can alter the Kf. Thus:
- *Contraction of mesangial cells* leading to decreased Kf.
- *Relaxation of mesangial cells* leading to increased Kf.

Substances causing contraction and relaxation of mesangial cells are listed in Table 6.2-1.

2. Hydrostatic pressure in Bowman's space fluid (P_{BS}) opposes filtration and therefore GFR is inversely related to it. It is decreasing in an acute obstruction of urinary tract (e.g. a ureteric obstruction by stone).

3. Glomerular capillary hydrostatic pressure (P_{GC}). GFR is directly related to P_{GC}. P_{GC} is mainly dependent on:

(i) *Arterial pressure.* GFR is autoregulated between arterial pressure of 80–200 mm Hg. Increased arterial pressure above 200 mm Hg may raise GFR and decreased arterial pressure below 70 mm Hg may lower GFR.

(ii) *Renal blood flow.* GFR is directly proportional to the renal blood flow (as described above page 383). However, renal blood flow is controlled by the auto-regulatory mechanisms.

(iii) *Afferent and efferent arteriolar resistance.* Relation of afferent and efferent arteriolar resistance with GFR is described on page 383.

Note. In *acute renal failure,* GFR declines because of fall in P_{GC}.

4. Glomerular capillary oncotic pressure (π_{GC}). GFR is inversely proportional to π_{GC}. In hyperproteinaemia and in haemoconcentration, the π_{GC} is raised leading to a decrease in GFR. Conversely in hypoproteinaemia and haemodilution, the π_{GC} is reduced leading to increased GFR.

5. Sympathetic stimulation (see page 385).

6. State of glomerular membrane (see page 388).

MEASUREMENT OF GLOMERULAR FILTRATION RATE

Glomerular filtration rate can be measured by the renal clearance of inulin, urea and creatinine.

The renal clearance can be defined as volume of plasma, i.e. cleared of substance in 1 min by excretion of the substance in urine and is calculated by the following formula:

$$C(mL/min) = UV/P,$$

where

C = Renal clearance of the substance,
U = Urine concentration of substance,
V = Rate of urine flow and
P = Plasma concentration of the substance

For details see page 436.

TUBULAR REABSORPTION AND SECRETION

Of the 180 L glomerular filtrate formed per day, about 1.5 L (i.e. less than 1%) per day is excreted as urine. The different segments of the renal tubule viz. proximal tubule, loop of Henle, distal tubule and collecting duct determine the composition and volume of the urine by process of *selective reabsorption* of solutes and water and *selective secretion* of solutes. For conceptual understanding renal tubular reabsorption and secretion can be considered in the following subsections:

- General principles of renal tubular transport,
- Transport across different segments of renal tubule,
- Tubular transport of common solutes and water.

GENERAL PRINCIPLES OF RENAL TUBULAR TRANSPORT

Transport mechanisms across cell membrane

The water moves across the cell membrane of renal tubular cells passively, while the solute movement occurs by both passive and active mechanisms.

Passive transport does not need energy and occurs spontaneously, down an electrochemical gradient by following mechanisms:

- Diffusion,
- Facilitated diffusion (channels, uniport, coupled transport, uniport or symport) and
- Solvent drag. For details see page 19.

Active transport requires direct input of energy and is abolished if cell metabolism is inhibited. Active transport can occur against an electrochemical gradient. Most of the active transports are carrier mediated.

Transepithelial transport pathways

In the renal tubule, a substance can be reabsorbed or secreted by two pathways (Fig. 6.2-5):

Transcellular pathway refers to the transport through the cells. Its example includes transcellular Na^+ reabsorption by the proximal tubule, which is a two-step process:

- *Movement of* Na^+ into the cell across the apical membrane occurs down an electrochemical gradient established by Na^+–K^+–ATPase.
- *Movement of* Na^+ into the extracellular fluid across the basolateral membrane occurs against an electrochemical gradient via Na^+–K^+–ATPase.

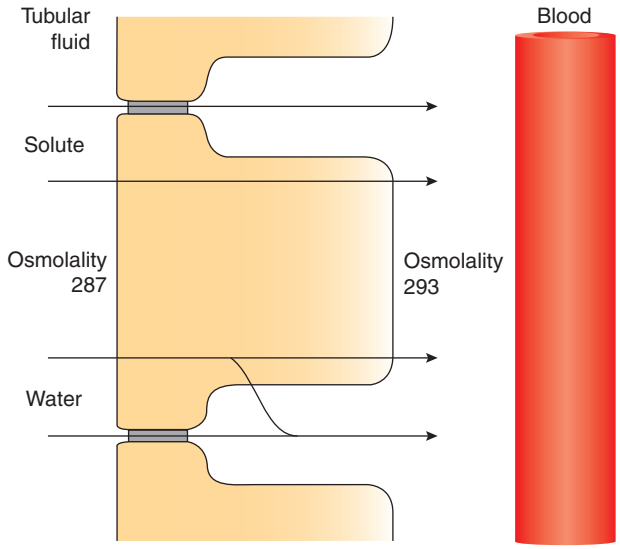


Fig. 6.2-5 Routes of water and solute reabsorption across the proximal tubule (for understanding see text).

Paracellular pathway refers to the transport between the cells. Examples of paracellular pathway include:

- Reabsorption of Ca^{2+} and K^+ across the proximal tubule,
- Some of the water reabsorbed across the proximal tubule crosses the paracellular pathway and
- Some solutes dissolved in this water (in particular Ca^{2+}, and K^+) are carried along with the reabsorbed fluid across the paracellular pathway by the process of solvent drag.

Tubular mechanisms, patterns of renal handling of substances and concept of renal clearance

Tubular mechanisms

As mentioned earlier, the two main tubular mechanisms involved in renal handling of a substance are tubular reabsorption and tubular secretion.

Tubular reabsorption denotes the active transport of solutes and passive movement of water from the tubular lumen into the peritubular capillaries. In other words, reabsorption is the removal of substances of nutritive value, such as glucose, amino acids, electrolytes (Na^+, K^+, Cl^-, HCO_3^-) and vitamins from the glomerular filtrate. Small proteins and peptide hormones are reabsorbed in the proximal tubules by endocytosis.

Tubular secretion refers to the transport of solutes from the peritubular capillaries into the tubular lumen, i.e. it is the addition of a substance to the glomerular filtrate.

Active secretion of substances occurs into the tubular fluid with the help of certain non-selective carriers. The carrier which secretes para-aminohippuric (PAH) acid can also secrete uric acid, bile acids, oxalic acid, penicillin, probenecid, cephalothin and furosemide.

Patterns of renal handling of a substance and concept of renal clearance

Different patterns of renal handling of a substance include:

1. Glomerular filtration only, i.e. the substances are freely filtered, but neither reabsorbed nor secreted (e.g. inulin) (Fig. 6.2-6A). Such substances are called glomerular markers and said to have renal clearance equal to GFR.

2. Glomerular filtration followed by partial reabsorption (Fig. 6.2-6B). Such substances have renal clearance less than GFR.

3. Glomerular filtration followed by complete tubular reabsorption (Fig. 6.2-6C). Such substances have lowest renal clearance, e.g. Na^+, glucose, amino acids, HCO_3^- and Cl^-. The substances that are not filtered at all (e.g. protein) also have lowest renal clearance.

4. Glomerular filtration followed by tubular secretion (Fig. 6.2-6D). Such substances that are both filtered across the glomerular capillaries and secreted from the peritubular capillaries into urine have the highest renal clearances (e.g. PAH).

5. Glomerular filtration followed by partial reabsorption and secretion (Fig. 6.2-6E). In such circumstances, depending upon which of the two processes are dominant, there may be net reabsorption or net secretion of the substance. Net absorption is said to occur if the amount of substance excreted in urine is less than GFR in the same time. Similarly, net secretion is said to occur when the amount excreted is more than GFR, in the same time.

6. No glomerular filtration, no absorption, only secretion (Fig. 6.2-6F). Many organic compounds are bound to plasma proteins and are therefore unavailable for ultrafiltration. Secretion is thus their major route of excretion in urine.

Renal clearance. Concept of renal clearance can be defined as the volume of plasma that is cleared of a substance in 1 min by excretion of the substance in the urine. Further, from the above description and examples, the relative clearance of the common substances is:

PAH > K^+ (high K^+ diet) > Inulin> Urea > Na^+ > Glucose, amino acids and HCO_3^-.

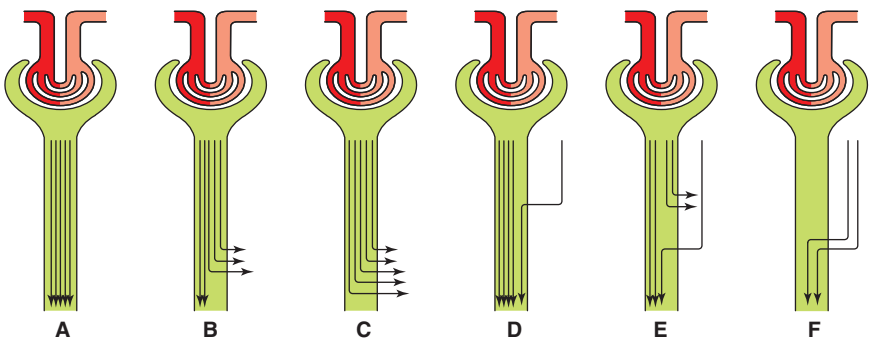


Fig. 6.2-6 Different patterns of tubular handling of substances (for details see text).

Quantification of renal tubular transport

The parameters used for the quantitative analysis of renal tubular transport are denoted by the capital letters with dots above them and include:

Filtered load. It is the amount of a substance entering the renal tubule by glomerular filtration per unit time.

- Filtered load (F°) is calculated by multiplying the GFR with plasma concentration of the substance (Px), i.e. F° = GFR × Px mg/min.

Excretion rate. It is the amount of substance that appears in the urine per unit time. The excretion rate (E°) can be calculated by multiplying urine flow rate (V) and the urinary concentration of the substance (Ux).

That is, $E^\circ = V \times Ux$ mg/min.

Reabsorption rate (R°). Reabsorption of a substance is said to occur when the filtered load exceeds the excretion rate. The reabsorption rate of a substance is calculated by subtracting excretion rate from the filtered load, that is:

$$R^\circ = F^\circ - E^\circ.$$

Secretion rate (S°). The net secretion of a substance is said to occur when the excretion rate is more than the filtered load. Under such circumstances, the secretion rate is calculated by subtracting filtered load from the excretion rate, that is:

$$S^\circ = E^\circ - F^\circ$$

Renal tubular transport maximum (T_m). It refers to the maximal amount of a solute that can be actively transported (reabsorbed or secreted) per minute by the renal tubules. In other words, the point at which carriers are saturated is the T_m. Therefore, it is important to note that T_m pertains to solutes that are actively transported only and the substances that are passively transported (e.g. urea) do not exhibit T_m.

- Substances that have T_m are phosphate ion, sulphate, glucose, many amino acids, uric acid, albumin, acetoacetate, β-hydroxybutyrate and β-ketoglutarate.
- Substances that do not have T_m include reabsorption of Na^+ along the nephron and HCO_3^-.

Threshold concentration is defined as the plasma concentration at which a substance first appears in the urine.

Tubular fluid concentration (TF)/plasma concentration (Px) ratio

The TF/Px ratio compares the concentration of a substance in tubular fluid at any point along the nephron with its concentration in plasma. The tubular fluid, for such studies, is collected by a micropuncture technique. A micropipette is inserted into the Bowman's space and different portions of the tubules of the living kidney in experimental animals and the composition of aspirated tubular fluid is determined by the use of microchemical techniques.

Significance of TF/Px ratio

The TF/Px ratio may be 1, <1 or >1.

TF/Px ratio of 1.0 signifies that either there has been no reabsorption or reabsorption of the substance has been exactly proportional to the reabsorption of water.

TF/Px ratio of <1.0 signifies that reabsorption of a substance has been greater than the reabsorption of water and its concentration in tubular fluid is less than that of the plasma.

TF/Px ratio of >1.0 signifies that either the reabsorption of a substance has been less than the reabsorption of water or there has been secretion of the substance.

TRANSPORT ACROSS DIFFERENT SEGMENTS OF RENAL TUBULE

The substances transported across the different segments of renal tubules are described below and enlisted in Table 6.2-2.

Table 6.2-2 Transport of substances across different segments of renal tubule

Reabsorption		Non-reabsorption	Secretion
Active	**Passive**		
Proximal tubule			
Na^+	Cl^-	Inulin	H^+
K^+	HCO_3^-	Creatinine	Water
Ca^{2+}	HPO_4^-	Sucrose	Penicillin
Mg^{2+}	Water	Mannitol	Sulphonamide
HPO_4^{2-}	Urea		Creatinine
SO_4^{2-}			Urate
NO_3^-			Water
Glucose			
Amino acids			
Protein			
Urate			
Vitamins			
Acetoacetate			
β-hydroxybutyrate			
Henle's loop			
Na^+	Cl^-		
K^+	HCO_3^-		
Ca^{2+}	Water		
Distal tubule and collecting duct			
Na^+	Cl^-		K^+
Ca^{2+}	HCO_3^-		H^+
Mg^{2+}	Water		
Water			

Transport across proximal tubule

The proximal tubule reabsorbs:

- Approximately 67% of the filtered water, Na^+, Cl^-, K^+ and other solutes.
- Almost all the glucose and amino acids filtered by the glomerulus.

The proximal tubule does not reabsorb inulin, creatinine, sucrose and mannitol.

The proximal tubule secretes H^+, PAH, urate, penicillin, sulphonamides and creatinine.

Transport across loop of Henle

About 20% of filtered Na^+ and Cl^-, 15% of filtered water and cations, such as K^+, Ca^{2+} and Mg^{2+} are reabsorbed in the loop of Henle.

Transport across distal tubules and collecting duct

Approximately 7% of the filtered NaCl and about 8–17% of water is reabsorbed and K^+ and H^+ are secreted in these segments.

RENAL HANDLING OF COMMON SOLUTES AND WATER

Renal handling of common solutes and water which include:

- Renal handling of sodium and water,
- Renal handling of potassium,
- Renal handling of glucose,
- Renal handling of proteins, peptides and amino acids,
- Renal handling of urea,
- Renal handling of uric acid and
- Renal handling of PAH.

RENAL HANDLING OF SODIUM AND WATER

Percentage reabsorption of the filtered sodium in different segments of the renal tubule is:

- Proximal tubule : 67%
- Loop of Henle (mainly thick ascending limb) : 20%
- Distal tubule : 7%
- Cortical collecting duct : 5%

Reabsorption in proximal tubule

The process of sodium reabsorption in the proximal tubule is *isosmotic,* i.e. the reabsorption of sodium and water is exactly proportional.

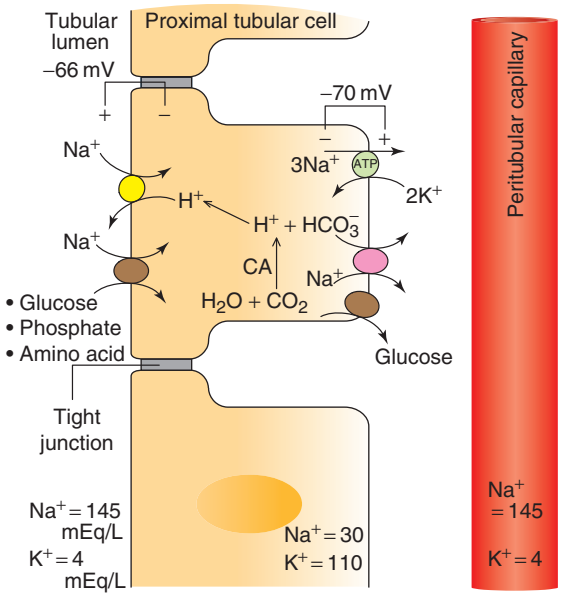


Fig. 6.2-7 Mechanism of reabsorption of sodium and other solutes across early proximal tubule.

Mechanisms of Na^+ reabsorption

Mechanism of Na^+ reabsorption in the early proximal tubule and late proximal tubule is different.

In early proximal tubule, Na^+ is reabsorbed by *cotransport with H^+* or organic solutes (glucose, amino acids, phosphate and lactate). The Na^+ absorption is a two-step process (Fig. 6.2-7):

Across the basolateral membrane, Na^+ moves against an electrochemical gradient via Na^+–K^+–ATPase pump, which pumps Na^+ into the paracellular spaces and lowers the intracellular Na^+ concentration.

Across the apical membrane, the sodium moves down an electrochemical gradient as above. The entry of Na^+ is mediated by the specific antiporter and symporter proteins. These include:

- *Na^+–H^+* antiporter is the main determinant of Na^+ and H_2O reabsorption in the proximal tubule. Na^+–H^+ exchange is linked directly to the reabsorption of HCO_3^-.

  Note. Carbonic anhydrase inhibitors (e.g. acetazolamide) are diuretics that act in the early proximal tubule by inhibiting the reabsorption of filtered HCO_3^-.

- *Na^+*–glucose (and other organic solutes) symporter mechanisms are also involved in the entry of Na^+ in the proximal cells. The glucose, amino acids, phosphate and lactate are almost completely absorbed along with Na^+ by the symporter (carrier) proteins which are different for different molecules.

  Note. The reabsorption of Na^+–HCO_3^- and Na^+-organic solutes across the proximal tubule establishes a

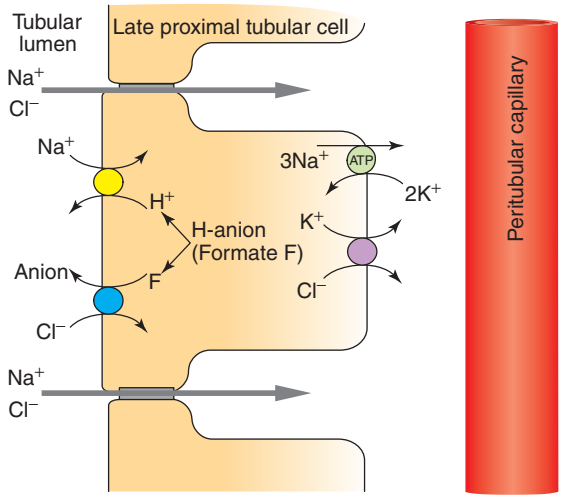


Fig. 6.2-8 Mechanism of reabsorption of sodium and other solutes across late proximal tubule.

transtubular osmotic gradient that provides the driving force for the passive absorption of water by osmosis. Because more water than Cl^- is reabsorbed in the early segment of proximal tubule, the Cl^- concentration in tubular fluid rises along the length of the early proximal tubule.

In the late proximal tubule, the Na^+ is reabsorbed primarily by *chloride-driven sodium transport mechanism* across both the transcellular and paracellular pathways.

- *Reabsorption via paracellular pathway.* The filtered glucose, amino acids and HCO_3^- have already been almost completely removed from the tubular fluid by reabsorption in the early proximal tubule. So the fluid entering the late proximal tubule contains very little of these substances but contains a high concentration of Cl^- (140 mEq/L) compared with that in the early proximal tubule (105 mEq/L). This high Cl^- (140 mEq/L) concentration in the lumen of late proximal tubule and comparatively low concentration (105 mEq/L) in the interstitium creates a concentration gradient which favours the diffusion of Cl^- from the tubular lumen across the tight junctions into lateral intercellular space. Movement of the negatively charged Cl^- causes the tubular fluid to become positively charged relative to the blood. This causes the diffusion of Na^+ across the tight junctions into the blood.
- *Transcellular Na^+* reabsorption across the luminal membrane of late proximal tubule cells occurs due to parallel operation of Na^+–H^+ and one or more Cl^- anion (formate) antiporters (Fig. 6.2-8).
- *Across the basolateral membrane,* the Na^+ leaves the cell by the action of Na^+–K^+–ATPase pump and Cl^- leaves by K^+–Cl^- co-transporter (Fig. 6.2-8).

Reabsorption in loop of Henle

Reabsorption occurring in different parts of the loop of Henle is:

Thin descending limb of loop of Henle

Water absorption occurs passively (because of hypertonic interstitial fluid) in this part of the loop of Henle. It is accompanied by diffusion of sodium ions from the interstitial fluid into the tubular lumen.

Thin ascending limb of loop of Henle

Limited passive reabsorption of Na^+ and Cl^- occurs in this water-impermeable limb. Because of impermeability to water, the fluid leaving this limb is hypotonic relative to plasma.

Thick ascending limb of loop of Henle

This limb is impermeable to water but is involved in the reabsorption of 20% of the filtered Na^+, Cl^- and other cations. About half of the Na^+ is reabsorbed actively and transcellularly while other half of the Na^+ is reabsorbed passively by the paracellular pathway along with other cations as described below:

Na^+, K^+–2Cl^- symporter-mediated active transport of sodium. Salient points are (Fig. 6.2-9):

- The key element involved is Na^+–K^+–ATPase located in the basolateral membrane which extrudes Na^+ and leading to low intracellular Na^+ concentration.
- Due to low intracellular Na^+, a chemical gradient is created which favours movement of Na^+ from the lumen into the cell.
- The movement of Na^+ across of apical membrane is mediated by Na^+–K^+–2Cl^- symporter.
- This symporter, with downhill movement of Na^+ and Cl^-, drives the uphill movement of K^+ influx.
- Na^+ leaves the cell across the basolateral membrane by the action of Na^+–K^+–ATPase.
- Due to the presence of 'tight' tight junctions, Na^+ is unable to leak back into the tubule to produce a luminal potential; however, some of the K^+ which enters the cell leaks back across the apical membrane into the tubular lumen, generating a lumen-positive transepithelial potential difference of +6 to +10 mV.

  Note. Thick ascending limb is the site of action of the loop diuretics (e.g. furosemide, ethacrynic acid), which inhibit Na^+–K^+–2Cl^- transporter.

Na^+–H^+ antiporter-mediated active reabsorption of sodium also occurs transcellularly leading to H^+ secretion (HCO_3^- reabsorption) (Fig. 6.2-9).

Paracellular passive reabsorption of Na^+, K^+, Ca^{2+} and Mg^{2+} is the function of voltage across the thick

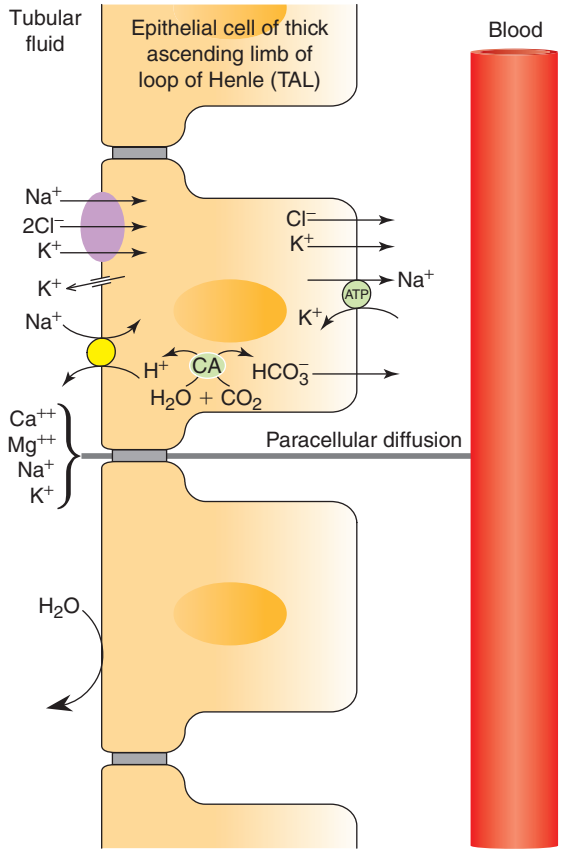


Fig. 6.2-9 The active (transcellular) and passive (paracellular) transport mechanism operating across the tubular cells in thick ascending limb (TAL) of loop of Henle.

ascending limb. Because of unique location of transport proteins in the apical and basolateral membranes, the tubular fluid is positively charged relative to the blood. The increased salt reabsorption by the thick ascending limb increases the magnitude of positive charge in the lumen, which plays a major role in driving passive paracellular reabsorption of cations (Fig. 6.2-9).

Note. Thick ascending limb is *impermeable* to water. Thus, NaCl and other solutes are reabsorbed without water. As a result, tubular fluid Na^+ and tubular fluid osmolarity decreases to less than their concentration in plasma (i.e. $TF/P_{osm} < 1.0$). This segment is therefore called the *diluting segment.* Further, Na^+ reabsorbed from this segment is the main driving force behind the counter-current multiplier system which concentrates Na^+ and urea in medullary interstitium.

Reabsorption across distal tubule and collecting duct

Early distal tubule (initial segment of distal tubule) reabsorbs Na^+, Cl^- and is impermeable to water.

- *Na^+–Cl^- symporter* mechanism is involved in the transport of Na^+–Cl^- across the apical membrane. Across the basolateral membrane, Na^+ leaves the cell via the action of Na^+–K^+–ATPase, and Cl^- leaves the cell by diffusion via channels (Fig. 6.2-10).
- Because of *impermeability to water*, the reabsorption of NaCl in this segment occurs without water leading to dilution of tubular fluid. This is why it is also called cortical diluting segment.
- *Thiazide diuretics* reduce NaCl reabsorption by inhibiting Na^+–Cl^- co-transport.

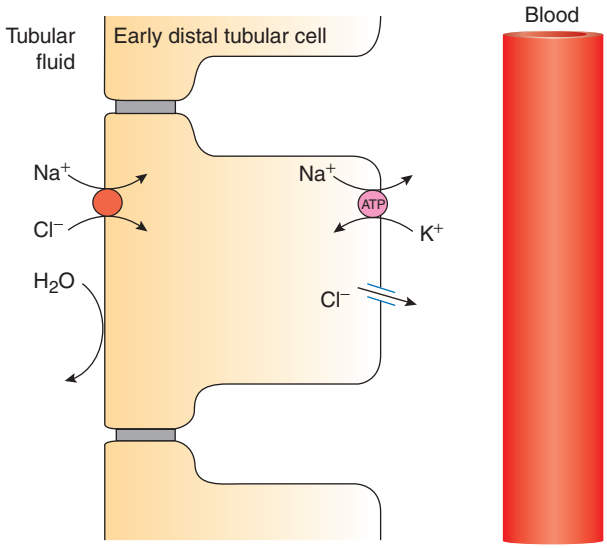


Fig. 6.2-10 Mechanism of reabsorption of Na^+ and Cl^- in the early distal tubule. This segment is impermeable to water (see text for details).

Late distal tubule and collecting duct have two cell types (principal cells and intercalated cells), which perform both reabsorption and secretory functions:

Principal cells reabsorb Na^+, Cl^- and H_2O and secrete K^+ (Fig. 6.2-11).

- *Na^+ reabsorption.* Na^+ is actively transported using Na^+–K^+–ATPase across the basolateral membrane. Across the apical membrane Na^+ diffuses passively due to the chemical gradient.
- *Cl^- reabsorption* occurs passively through the paracellular pathway. Cl^- is driven by the lumen-negative charge generated by the diffusional influx of sodium.
- *H_2O absorption* occurs in response to the effect of antidiuretic hormone (ADH) on the principal cells. ADH increases H_2O permeability by directing the insertion of H_2O channels in the luminal membrane. In the absence of ADH, the principal cells are virtually impermeable to water.
- *K^+ secretion.* K^+ uptake across the basolateral membrane occurs via the action of Na^+–K^+–ATPase followed by the diffusion down its electrochemical gradient across the apical cell membrane into the tubular fluid.

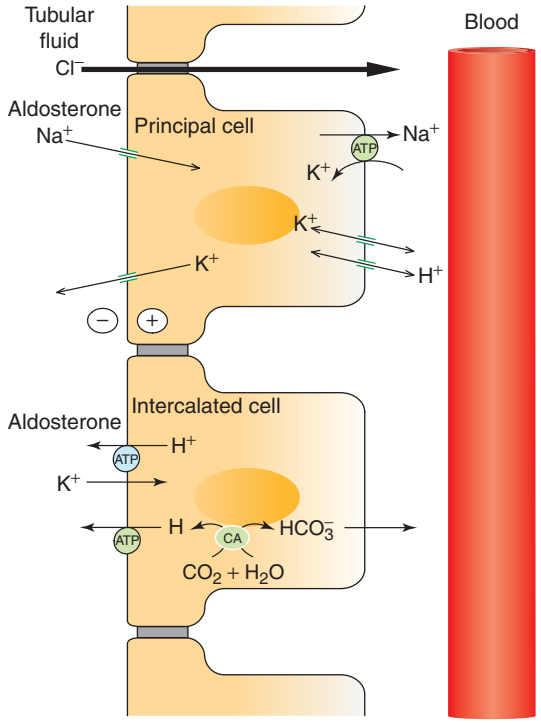


Fig. 6.2-11 Mechanism of transport in principal cells and intercalated cells of the late distal tubule and collecting duct. CA = Carbonic anhydrase.

Role of aldosterone on principal cell functions. Aldosterone increases Na^+ reabsorption and increases K^+ secretion. Like other steroid hormones, the action of aldosterone takes several hours to develop because new protein synthesis is required. About 2% of overall Na^+ absorption is affected by aldosterone.

Intercalated cells reabsorb K^+ and secrete H^+. Aldosterone also increases H^+ secretion by the intercalated cells by stimulating the H^+–ATPase.

Note

- This function is important for acid-base balance (see page 397).
- *Aldosterone* increases H^+ secretion by intercalated cells by stimulating the H^+-ATPase (in addition to its actions on the principal cells).

Mechanism of absorption of Na^+ in different segments of renal tubules is summarized in Table 6.2-3.

Table 6.2-3 Summary of mechanism of Na^+ absorption across different segments of renal tubule

Segment of the tubule	Absorption active/passive/ impermeable	Mediated by
Proximal tubule		
• Early proximal tubule	Active	• Na^+, K^+ antiport • Na^+–glucose (and other organic solutes) symport
• Late proximal tubule	Active	• Cl^- driven Na^+ transport
Loop of Henle		
• Descending thin segment (DTS)	Passively secreted in interstitium	
• Ascending thin segment (ATS)	Passive	
• Thick ascending limb (TAL)	Active (Transcellular)	• Na^+–K^+–$2Cl^-$ symporter • Na^+, H^+– antiporter
Distal tubule and collecting duct		
• Early distal tubule	Active	• Na^+, Cl^- symport
• Late distal tubule and collecting duct (Principal cell)	Active	• Regulated by aldosterone

Water reabsorption

Site and mechanism of reabsorption

Water is absorbed passively by osmosis in response to a transtubular osmotic gradient. Rapid diffusion of water across the cell membrane occurs through water channels made up of proteins called *Aquaporins*. Different types of aquaporins are aquaporin-1, 2, 5 and 9. Mostly, these aquaporins are present in the kidney. In the collecting ducts, reabsorption of water is controlled by ADH. Renal handling of water by different segments of renal tubule is as:

Proximal tubule: Passively reabsorbed (67%).

Loop of Henle

- Descending thin segment (DTS) : Passively reabsorbed (15%)
- Ascending thin segment (ATS) : Impermeable
- Thick ascending limb : Impermeable (TAL)

Distal tubule and collecting duct (8–17%)

- Distal convoluted tubule (DCT) : Impermeable
- Connecting tubule (CNT) : Impermeable
- Cortical collecting duct (CCD) : Reabsorbed (ADH)
- Outer medullary collecting duct (OMCD) : Reabsorbed (ADH)
- Inner medullary collecting duct (IMCD) : Reabsorbed (ADH)

Obligatory and facultative reabsorption of water

Obligatory reabsorption. About 85% of the filtered water is always reabsorbed, irrespective of the body water balance. This reabsorption occurs by osmosis in response to a transtubular osmotic gradient and is called obligatory (must occur) reabsorption.

- About 67% of obligatory reabsorption occurs in the proximal tubules and
- About 15–18% of obligatory reabsorption occurs in the descending thin segment of loop of Henle.

Facultative reabsorption. The remaining 15–18% of the water may or may not be absorbed depending upon the body water balance. It is called facultative (optional) reabsorption.

Facultative reabsorption of water occurs from the collecting tubule and is under the control of ADH.

RENAL HANDLING OF POTASSIUM

Functions of K^+

Potassium (K^+) is one of the most abundant cations in the body. It is the *principal intracellular cation*, and is equally important in the extracellular fluid for specific functions. Potassium is required for the following biochemical functions:

- Maintenance of intracellular osmotic pressure,
- Optimal activity of enzyme pyruvate kinase (of glycolysis),
- Proper synthesis of DNA and proteins by ribosomes,
- Optimal cell growth,
- Transmission of nerve impulse,
- Generation of cell membrane potential and muscle contraction,
- Extracellular K^+ influences cardiac muscle activity and
- Regulation of acid-base balance and water balance in the cells.

Transport of potassium across major nephron

Glomerular filtration

Filtration occurs freely across the glomerular capillaries, potassium is not bound to plasma proteins.

Tubular reabsorption and secretion

As shown in Fig. 6.2-12, 67% of the filtered K^+ is reabsorbed in the proximal tubule, 20% in the loop of Henle and approximately 10% is delivered to the early distal tubule. In contrast to proximal tubule and loop of Henle, which are capable of only reabsorbing K^+, the distal tubule (DT) and collecting duct (CD) are able to either reabsorb or secrete K^+. The role of reabsorption or secretion by DT and CD depends on a variety of hormones and factors.

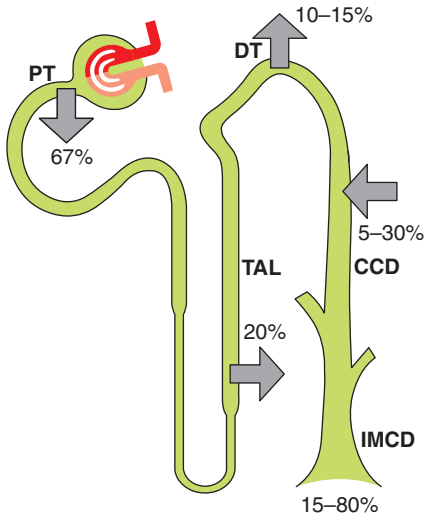


Fig. 6.2-12 Potassium transport along the nephron.

Reabsorption of K^+ by proximal tubule

In proximal tubule, approximately 7% of the filtered K^+ is reabsorbed passively in proportion to the H_2O reabsorption (solvent drug) and about 60% of the filtered K^+ is reabsorbed actively by the paracellular transport mechanism, which has following steps:

- *Concentration gradient* is created between the paracellular space and tubules fluid by active K^+ uptake via Na^+–K^+–ATPase located in the lateral cell membrane facing the lateral intercellular spaces.
- *Diffusion of K^+ along the concentration gradient* occurs from the tubular lumen into the lateral intercellular spaces. In this way, luminal fluid equilibrates with low K^+ concentration in the intercellular space.
- *Exit from the basolateral membrane* of most of the K^+ that enters the cell actively from lateral surface (as described above) occurs by three pathways:
  - The conductive K^+ channel,
  - The K^+–Cl^- co-transporter and
  - The Na^+–K^+–ATPase pump.

The K^+ that exits from the basolateral membrane is immediately absorbed in the peritubular capillaries.

Reabsorption of K^+ by loop of Henle

Twenty percent of the filtered K^+ is reabsorbed in the *thick ascending limb* (TAL) of loop of Henle along with Na^+ reabsorption by two mechanisms:

1. *Na^+–K^+–$2Cl^-$* active transport mechanism.
2. *Paracellular passive reabsorption* occurs as a function of voltage gradient across the thick ascending limb. For details see page 394 and Fig. 6.2-9.

Reabsorption and secretion of K^+ by distal tubule and collecting duct

Early distal tubule. Normally, in the distal convoluted tubule, K^+ is secreted. However, when there is need to conserve body K^+ (e.g. during K^+ depletion), K^+ is reabsorbed. Both secretion as well as reabsorption occurs in the same cell, i.e. distal tubular cell, depending upon the status of K^+ balance in the body.

Late distal tubule and collecting duct either reabsorb or secrete K^+, depending upon the dietary intake.

Reabsorption of K^+ occurs only when the dietary intake is very low (i.e. during K^+ depletion). Under these circumstances, K^+ excretion can be as low as 1% of the filtered load because the kidneys conserve as much K^+ as possible.

Secretion of K^+ is variable and accounts for the wide range of urinary K^+ excretion, depending upon the dietary K^+ intake, aldosterone levels, acid–base status and urine flow rate. *Principal cells* are involved in the K^+ secretion.

Mechanism of K^+ secretion is as follows (Fig. 6.2-11):

- At the basolateral membrane, K^+ is actively transported into the cell by Na^+–K^+–ATPase. As in all cells, this mechanism maintains a high intracellular K^+ concentration.
- At the apical membrane, K^+ passively secreted into the lumen through K^+ channels, down its electrical and chemical gradient.

Hormones and factors that regulate urinary K^+ excretion

Hormones and other factors involved in the regulation of K^+ tubular secretion and thus of urinary K^+ excretion include:

1. Plasma K^+ level, which mainly depends on
Dietary intake of K^+ is an important determinant of K^+ secretion by principal cells (Fig. 6.2-13). Hyperkalaemia, resulting from a high K^+ diet or any other factor (e.g. rhabdomyolysis) stimulates K^+ secretion within minutes.

Hypokalaemia, resulting from a low K^+ diet or other factors (e.g. diarrhoea) decreases K^+ secretion by mechanisms opposite to those described for hyperkalaemia.

2. Aldosterone. Salient points about role of aldosterone in regulating K^+ secretion by principal cells are:

Aldosterone secretion is increased by hyperkalaemia and angiotensin II (after activation of renin–angiotensin system).

Aldosterone secretion is decreased by hypokalaemia and atrial natriuretic peptide (ANP).

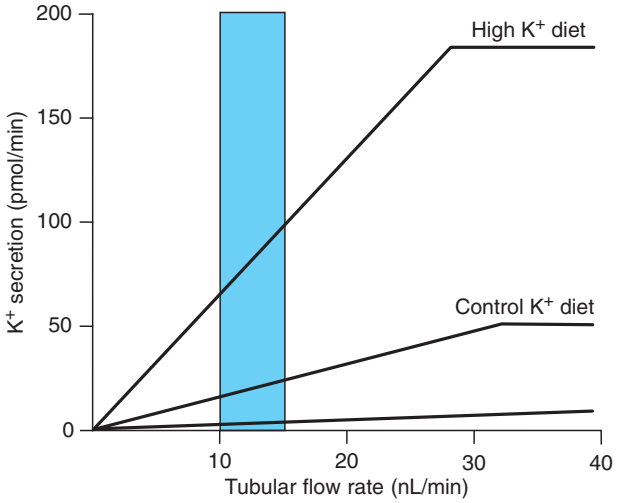


Fig. 6.2-13 Effect of dietary intake of K^+ on the relationship between tubular flow rate and K^+ secretion by the distal tubule and collecting duct.

Chronic rise in aldosterone level increases K^+ secretion by the principal cells by following mechanisms:

(i) *By increasing Na^+–K^+–ATPase activity.* Aldosterone increases the amount of Na^+–K^+–ATPase in the principal cells. This leads to increased pumping of Na^+ out of the cell at basolateral membrane and increased Na^+ entry into the cells across the luminal membrane.

(ii) *By making the transepithelial potential difference (TEPD) more lumen negative.* By increasing the Na^+ reabsorption from lumen, the aldosterone makes the TEPD more lumen negative which in turn favours K^+ secretion.

(iii) *By increasing the permeability* of apical membrane to K^+, aldosterone increases K^+ secretion.

3. Glucocorticoids indirectly increase K^+ excretion by increasing GFR which increases tubular flow. Increased tubular flow increases K^+ secretion.

4. Antidiuretic hormone. ADH increases Na^+ and water reabsorption and decreases the tubular flow which in turn decreases K^+ secretion (as discussed below). However, the ADH-induced increased Na^+ uptake across the luminal membrane of the principal cells creates an electrochemical gradient which increases K^+ secretion into the lumen. In this way, the inhibitory effect (through decreasing tubular flow) and stimulatory effect (through increasing electrochemical driving force for K^+) of ADH on K^+ secretion enable urinary K^+ excretion to be maintained constant despite wide fluctuations in water excretion.

5. Flow of tubular fluid. Increase in the tubular fluid flow *increases K^+ secretion* rapidly, while decrease in tubular

fluid flow decreases the secretion of K^+ by distal tubule and collecting ducts.

6. Acid–base balance effects the K^+ secretion by distal tubule and collecting ducts in the following manners:

Acute acidosis reduces K^+ secretion by two mechanisms:

- *By decreasing Na^+–K^+–ATPase activity* across basolateral membrane, it reduces the intracellular K^+ concentration and thus reduces the electrochemical driving force for K^+ exit across apical membrane.
- *By reducing the permeability of apical membrane*, it decreases K^+ secretion and also tends to increase intracellular K^+ concentration.

So, as a net result of the above two mechanisms, K^+ secretion by the principal cells decreased while their K^+ content remains unchanged.

Acute alkalosis has exactly the opposite effects to acute acidosis and thus as a net result increases K^+ secretion by the principal cells.

RENAL HANDLING OF GLUCOSE

Glucose reabsorption

Mechanism of tubular reabsorption

All the filtered glucose is completely reabsorbed into the proximal tubule by an active transport mechanism (Fig. 6.2-14):

Carrier mediated Na^+–glucose co-transport. Carrier protein located at the apical membrane in the proximal tubule reabsorbs glucose from tubular fluid into the blood.

- The carrier protein for glucose in early and late proximal tubule is called SGLT-2 and SGLT-1, respectively (SGLT = sodium-dependent glucose transporter).
- The carrier is driven by the Na^+ concentration gradient which exists between the high tubular (Na^+) concentration and the low intracellular (Na^+) gradient produced by the pumping out of Na^+ through the basolateral surface.

Facilitated diffusion moves the glucose out of the cell through the basolateral membrane. The carrier for facilitated diffusion across the basolateral membrane in early and late proximal tubule is called GLUT-2 and GLUT-1, respectively (GLUT = glucose transporter).

Characteristics of glucose transport and glucose excretion

Glucose is reabsorbed by a transport maximum process, i.e. there are limited number of Na^+–glucose carriers. The characteristics of glucose transport and glucose excretion can be elicited from the glucose titration curve,

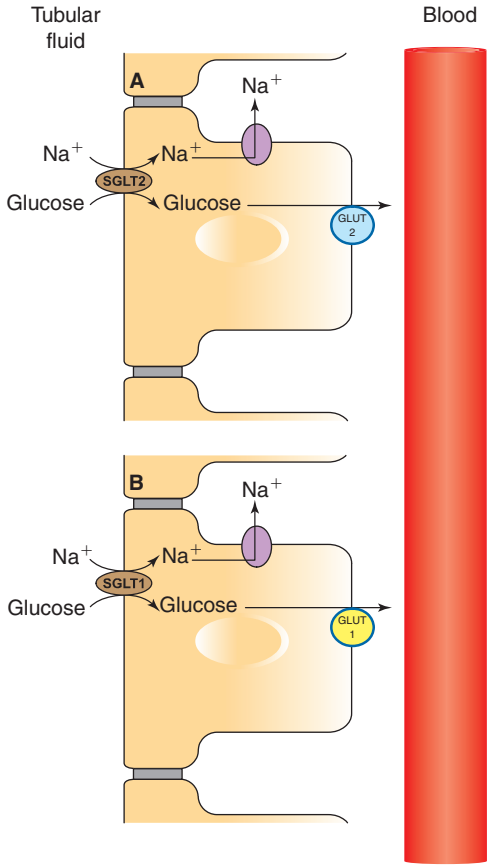


Fig. 6.2-14 Mechanism of glucose reabsorption in: A, early proximal tubule and B, late proximal tubule.

which is constructed by plotting the following pairs of variables:

- The filtered load against plasma glucose concentration,
- The excretion rate against plasma glucose concentration and
- The difference between the filtered load and excretion rate (i.e. maximum tubular reabsorption capacity, Tr) against plasma concentration.

Glucose titration curve depicts that (Fig. 6.2-15)

Filtered load increases with the plasma glucose concentration (P_G).

Renal threshold, i.e. the plasma glucose concentration at which glucose first appears in the urine (glycosuria) is about 180–200 mg/dL. At plasma levels below renal threshold, the reabsorption of glucose is complete (100%), i.e. all of the filtered glucose can be reabsorbed because plenty of carriers are available and hence no glucose is excreted in urine. In this region, the line of reabsorption is the same as that of filtration.

Transport maximum (T_m) refers to the plasma concentration at which carriers are fully saturated. As shown in Fig. 6.2-15, beyond plasma glucose concentration of

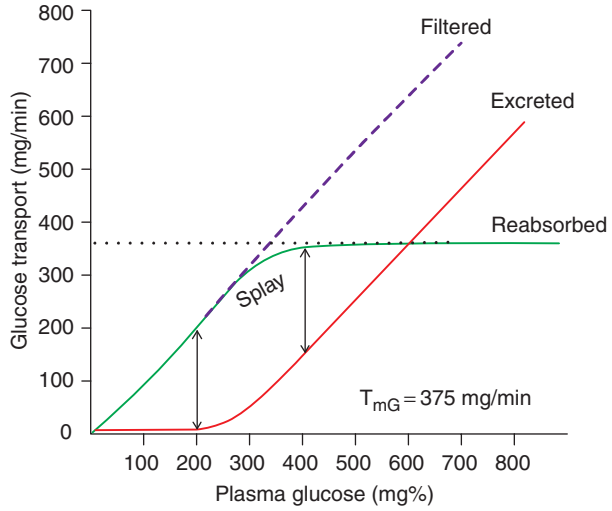


Fig. 6.2-15 Glucose titration curve (for details see text).

350 mg/dL (T_{mG}) the reabsorption rate does not increase, i.e. becomes constant and is independent of P_G. Therefore, as the T_{mG} is reached, the urinary excretion rate increases linearly with increase in plasma glucose concentration (Fig. 6.2-15).

Splay refers to the region of the glucose curve between threshold and T_{mG}, i.e. between P_G 180 and 350 mg/dL. It represents the excretion of glucose in urine before the T_{mG} is fully achieved. Note in the region of splay, the reabsorption curve is rounded indicating that though the reabsorption rate is increasing with increase in P_G, but reabsorption is less than filtration. Similarly, the excretion curve is also rounded in the region of splay, indicating that though the urinary excretion is increasing with increase in P_G, but there is no linear relation. *Causes of splay* are:

- *Heterogenicity in glomerular size,* proximal tubular length and number of carrier proteins for glucose reabsorption.
- *Variability in T_{mG}* of the nephron.

For example, there is variability in the number of glucose carrier, the transport rate of the carriers and the binding affinity of the Na^+ glucose carriers.

RENAL HANDLING OF PROTEINS, PEPTIDES AND AMINO ACIDS

Normally, up to 150 mg of proteins are excreted daily in urine, of this only 15 mg is albumin and the rest are low molecular weight proteins (LMWP). About 25 mg of LMWP are the *Tamm–Horsfall* proteins derived from the cells of TAL. The rest are derived from plasma proteins and include microproteins, lysozymes and light chains of immunoglobulins.

Proteinuria

Proteinuria is labelled when excretion of proteins in urine is more than 150 mg/day. It may be of following types:

1. Glomerular proteinuria occurs when the glomerular permeability increases and allows albumin and other large protein to be filtered.

2. Tubular proteinuria. Normally, low molecular weight proteins enter the glomerular filtrate in fairly large amounts. When tubular reabsorption for these proteins is impaired, e.g. in *tubulointerstitial disorders and Fanconi's syndrome*, then large amounts of low molecular weight proteins are excreted in urine.

3. Overflow proteinuria occurs when the amount of LMWP filtered exceeds the reabsorptive capacity of the tubules. Such a situation may arise when plasma levels of LMWP is increased, e.g. in multiple myeloma, in which large amounts of an abnormal protein called *Bence–Jones* protein appear in the plasma.

4. Nephrogenic proteinuria occurs when tubular enzymes such as N-acetyl β-glucosaminidase (NAG) and γ-glutamyl transferase (γ-GT) are released following damage to proximal tubular cells.

RENAL HANDLING OF UREA

Glomerular filtration

Urea is freely filtered into the glomerular filtrate. The amount of urea filtered by the glomerular capillaries varies with the protein intake.

Tubular transport

Proximal tubules reabsorb 5% of the filtered urea passively.

Proximal straight tubule (PST) *descending thin segment* (DTS) and *ascending thin segment* (ATS) of nephron receive urea by diffusion (*tubular secretion*) from the interstitium of renal medulla in which urea is present in high concentration.

Thick ascending limb, DCT, CNT, CCD and OMCD are all impermeable to urea.

Inner medullary collecting duct (IMCD) reabsorbs large amount of urea employing a specialized *urea transport protein* (UT-A). There are at least four isoforms of transport protein UT-A in the kidneys (UT-A1–UT-A4) (UT-B is found in erythrocytes). This protein is stimulated by ADH, which consequently increases urea permeability of the IMCD.

Urea recycling

Urea recycling involves the following steps (Fig. 6.2-16):

1. Concentration of urea in collecting duct (CCD and OMCD), as mentioned above, the nephron segment distal to ATS (i.e. TAL, DCT, CNT, CCD and OMCD) are all impermeable to urea. Therefore, as water is reabsorbed from the CCD, OMCD and the initial part of IMCD, urea gets more and more concentrated within the collecting duct.

2. Rapid and massive reabsorption of urea by IMCD, as described above, increases the concentration of urea in the medullary interstitium.

3. Carriage of urea by vasa recta to renal cortex interstitium. From the medullary interstitium, most of the urea enters the vasa recta and is carried upwards towards the renal cortex by the ascending vasa recta.

4. Tubular secretion of urea from the renal cortical interstitium occurs into the PST of cortical nephrons. Some of the urea also enters the thin segment of the long loops of the juxtamedullary nephrons. In this way, the urea is again carried back to the IMCD from where it diffuses out again resulting in a constant recycling.

Urea recycling plays an important role in the countercurrent system (see page 403).

RENAL HANDLING OF URIC ACID

Glomerular filtration

Urate is freely filtered by the glomerular capillaries.

Tubular transport

- Early proximal tubule (S1 segment) reabsorbs 95% of the filtered uric acid (Fig. 6.2-17).

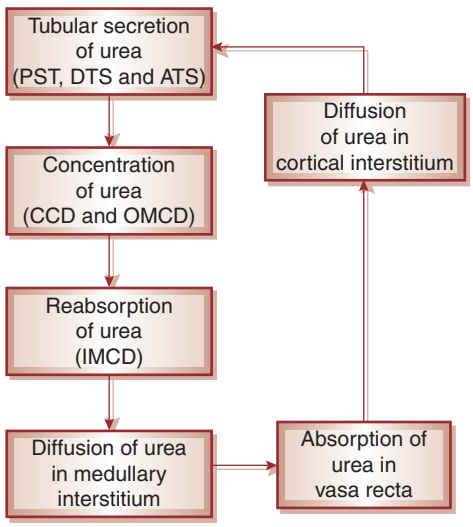


Fig. 6.2-16 Steps involved in urea recycling.

- Mid proximal tubule (S2 segment) secretes a moderate amount of uric acid equivalent to 50% of glomerular filtrate).
- Late proximal tubule (S3 segment) reabsorbs moderate amount of uric acid (equivalent to 40% of glomerular filtrate). This is called post-secretory reabsorption.

Mechanism of uric acid reabsorption

The uric acid is reabsorbed by two mechanisms:

Passive reabsorption occurs through *paracellular* pathway.

Secondary active transport, which occurs through transcellular pathways. The carrier protein involved is called *urate transport protein.* Across the basolateral membrane, the urate moves out using another anion exchanger. The same anion exchangers are employed for the urate secretion also.

RENAL HANDLING OF PARA-AMINO HIPPURIC ACID

The cells of proximal tubule, in addition to reabsorbing solutes and water, also secrete organic anions and organic cations, which include some end products of metabolism circulating in plasma, exogenous organic compounds and certain drugs (Table 6.2-4).

Mechanism of secretion of organic anions can be considered by the example of PAH secretion. Para-amino hippuric acid is an exogenous weak organic acid that is neither

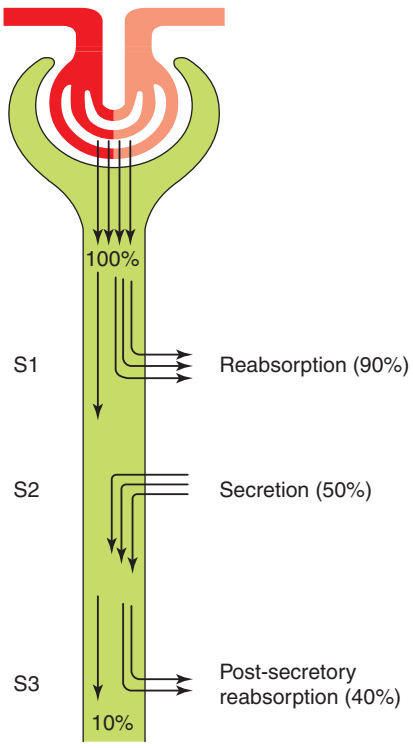


Fig. 6.2-17 Tubular transport of uric acid: reabsorption, secretion and post-secretory reabsorption.

Table 6.2-4 Some organic anions and cations secreted by proximal tubule cells

Anions secreted by proximal tubule	
Endogenous anion	**Drugs**
PAH	Acetazolamide
cAMP	Furosemide
Bile salts	Penicillin
Oxalate	Probenecid
Prostaglandins	Salicylate (aspirin)
Water	NSAIDs
Cations secreted by proximal tubule	
Endogenous cation	**Drugs**
Creatinine	Atropine
Dopamine	Cimetidine
Epinephrine	Morphine
Norepinephrine	Quinine
	Procainamide
	Verapamil

stored nor metabolized and is excreted virtually unchanged in the urine. Because 10% of PAH is bound to plasma proteins, so it is cleared from the plasma both by glomerular filtration and by tubular secretion through the kidney.

As with glucose, the filtered load of PAH increases in direct proportion to the plasma PAH concentration.

Secretion of PAH occurs from the peritubular capillary blood into the tubular fluid via carriers in the proximal tubule by a transport maximum T_m limited process.

- The carrier which secretes PAH is non-selective, i.e. it can transport most of the organic anions.
- At low plasma concentration of PAH, the secretion rate increases as the plasma concentration increases. But, once the carriers are saturated, further increase in the plasma PAH concentration does not cause further increase in the secretion rate T_m.
- The PAH is taken into the cell across the basolateral membrane against its chemical gradient, in exchange for α-ketoglutarate (αKG) via a PAH-αKG *antiport mechanism*. The resultant high intracellular concentration of PAH provides the driving force for PAH exit across the luminal membrane into the tubular fluid via a PAH-anion (A^-) antiporter and possibly a voltage-driven PAH transporter.

Note. PAH clearance can be used to measure renal plasma flow (see page 437).

Chapter 6.3

Concentration, Dilution and Acidification of Urine

CONCENTRATION AND DILUTION OF URINE: A MECHANISM TO REGULATE URINE VOLUME AND OSMOLALITY

INTRODUCTION

The kidneys possess unique property of regulating the volume and osmolality of the urine by concentrating and diluting it as per need of the body.

Purpose of concentration and dilution of urine. The main purpose is to maintain the osmolality and volume of the body fluids within a narrow range, which is accomplished by kidneys in concert with other systems by regulating the excretion of water and NaCl, respectively.

The kidney can produce urine with osmolality as low as 30 mOsm/kg H_2O to as high as 1400 mOsm/kg H_2O by changing the water excretion as high as 23.3 L/day to as low as 0.5 L/day, respectively (Table 6.3-1).

Principal factors. Principal factors responsible for mechanism of concentration and dilution of urine are:

- Antidiuretic hormones and
- Hyperosmolality and osmolality gradient in medullary interstitium of kidneys.

The details about the antidiuretic hormone (ADH) are described on page 546.

MEDULLARY HYPEROSMOLALITY AND MEDULLARY GRADIENT

The interstitial fluid of the medulla is critically important in concentrating the urine because the osmotic pressure of this fluid provides the driving force for reabsorbing water from both the descending thin segment (DTS) and the collecting duct.

Normal osmolality of plasma and other body fluids is about 300 mOsm/kg H_2O. The interstitial fluid of the renal cortex has the same osmolality as that of plasma, and virtually all osmoles attributable to NaCl. The osmolality of the renal medulla is higher than the plasma (i.e. *hyperosmolar*) and that it goes on increasing progressively from about 300 mOsm/kg H_2O at corticomedullary junction to about 1200 mOsm/kg H_2O at papilla (medullary gradient), where a maximally concentrated urine is excreted (Fig. 6.3-1). This hyperosmolality and medullary gradient is generated and maintained by the so-called counter current system.

COUNTER CURRENT SYSTEM

A counter current system refers to a system in which the inflow runs parallel to, counter to, and in close proximity to the outflow for some distance. The counter current flow system is formed by U-shaped tubules. The effect of counter current system can be best understood by studying the effect of a heater on a straight water pipe and a pipe bent in

Table 6.3-1 Effects of concentration and dilution mechanism on volume and osmolality of urine. In each case, the osmotic load excreted is same (700 mOsm/day)

Character of urine formed	GFR (mL/min)	Percentage of filtered water reabsorbed	Urine volume (L/day)	Urine concentration (mOsm/kg H_2O)	Gain or loss of water in excess of solute (L/day)
Urine isotonic to plasma	125	98.7	2.4	290	–
Concentrated urine (vasopressin: maximal antidiuresis)	125	99.7	0.5	1400	1.9 gain
Diluted urine (No vasopressin: complete diuresis diabetes insipidus)	125	87.1	23.3	30	20.9 loss

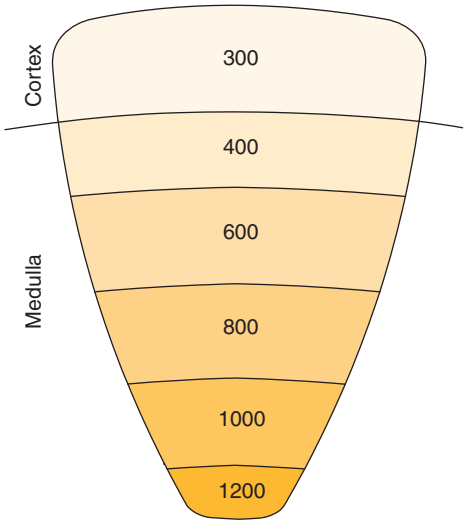


Fig. 6.3-1 Osmolality gradient of renal medullary interstitium (values are in mOsm/kg H_2O).

an U-shape (Fig. 6.3-2). Let us suppose the heater raises the temperature of the flowing fluid through the straight tube at a constant flow rate by 10°C from 30 to 40°C (Fig. 6.3-2A). Now if the pipe is bent in an U shape and the two limbs are brought in close proximity, the temperature at the outlet is again 40°C, but a gradient of temperature is set up along the pipe in such a way that at the bend of pipe the temperature is not raised from 30 to 40°C but from 30 to 100°C (Fig. 6.3-2B). This is because the outgoing fluid warms the incoming fluid and sets up the counter current system.

In the kidney, the structures which form the counter current system are loop of Henle and the vasa recta. The counter current system of kidney consists of two components:

- Counter current multiplier, which is formed by the operation of loop of Henle and is responsible for production of hyperosmolality and a gradient in renal medulla and
- Counter current exchanger, which is formed by the operation of the vasa recta and is responsible for the maintenance of the medullary gradient and hyperosmolality.

Counter current multipliers

The working of a counter current multiplier, which operates in the loop of Henle and generates hyperosmolarity and medullary gradient, can be best understood by describing it as two processes:

- Origin of single effect and
- Multiplication of the single effect.

Origin of single effect in outer medulla

As shown in Fig. 6.3-3, the thick ascending limb (TAL) of loop of Henle is located in the outer medulla, and not in the inner medulla. TAL is impermeable to water. So NaCl and other solutes are actively absorbed in this segment without water, as a result the tubular fluid osmolality decreases to less than the plasma. This is why this segment is called diluting segment. The NaCl reabsorbed in this segment raises the osmolality of outer medullary interstitium by about 200 mOsm. This separation of solute and water by the TAL leading osmolality difference between that tubular fluid and interstitial fluid is called single effect and is the main driving force behind the counter current multiplier.

Origin of single effect in inner medulla

The origin of single effect for the development of medullary gradient and hyperosmolality occur in the inner medullary interstitium due to:

- Passive transport of sodium ions from the ascending thin segment into the interstitium (see page 393).
- Active transport of sodium from the inner medullary collecting duct (see page 394).
- Diffusion of urea from the collecting duct into the medullary interstitium (see page 399).

Role of urea. Urea plays an important role in the development of medullary osmotic gradient, especially when concentration of ADH is high in the blood. Under such circumstances, the inner medullary collecting duct absorbs large amount

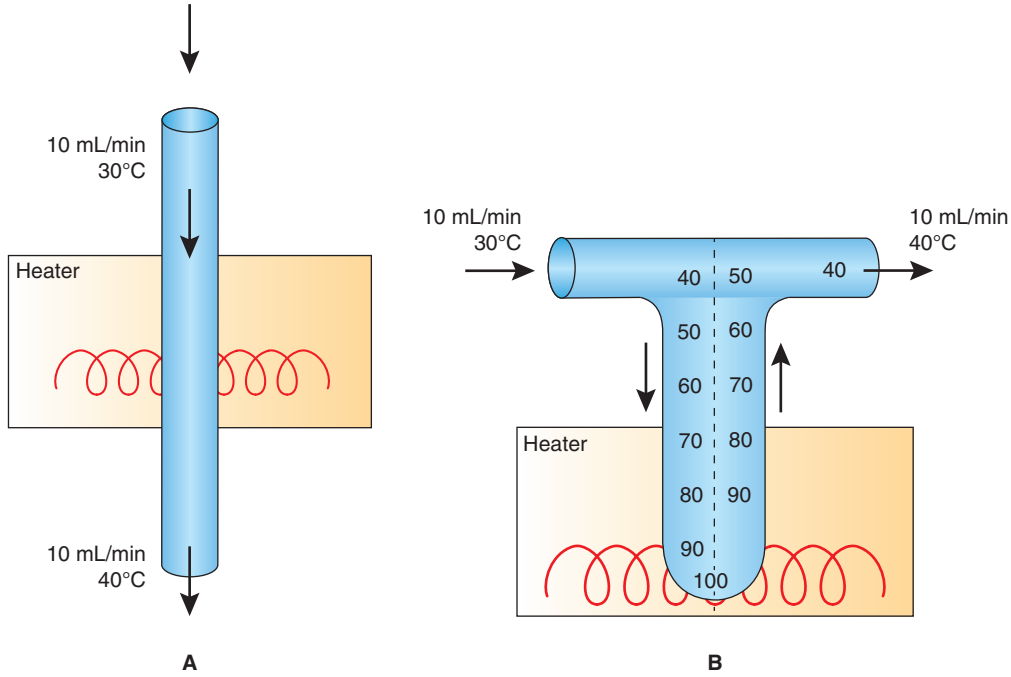


Fig. 6.3-2 Understanding the principle of counter current system. Effect of heating on water flowing at constant rate: A, from a straight pipe and B, from a U-shaped bent pipe (counter current effect).

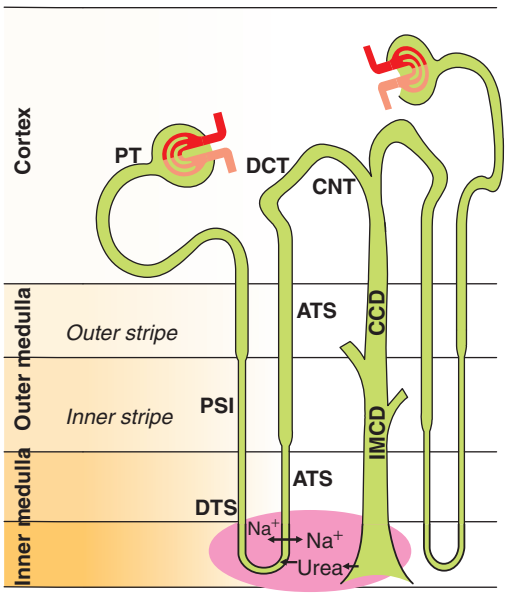


Fig. 6.3-3 Location of different components of cortical nephron and juxta cortical nephron with particular reference to origin of single effect in outer versus inner medulla.

Table 6.3-2 Permeability and transport in various segments of the nephron

Segment of nephron	Permeability and transport of			
	H_2O	Urea	NaCl	Na^+
Loop of Henle	4+	+	±	0
• Thin descending limb	0	+	4+	0
• Thin ascending limb	0	±	±	4+
• Thick ascending limb	±	±	±	3+
Distal convoluted tubule	±	±	±	3+
Collecting duct				
• Cortical collecting duct (CCD)	*3+	0	±	2+
• Outer medullary collecting duct (OMCD)	*3+	0	±	1+
• Inner medullary collecting duct (IMCD)	*3+	3+	±	1+

*These values are in the presence of ADH. In the absence of ADH these values are positive.

of urea, which plays an important role in the counter current system.

Multiplication of the single effect

The hyperosmolality and medullary gradient is in fact generated by the multiplication of the single effect by the counter current multiplier. The main characteristics of the components of counter current multiplier which play role in multiplication of single effect are (Table 6.3-2):

- High permeability of the descending thin segment to water.
- Impermeability to water but high permeability to NaCl of thin ascending limb of loop of Henle.

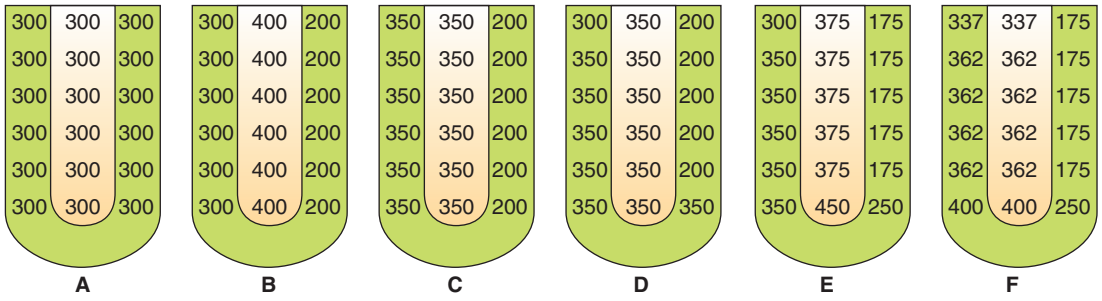


Fig. 6.3-4 Operation of loop of Henle as counter current multiplier producing a gradient of hyperosmolality in the medullary interstitium. A, B, C, D, E and F are hypothetical steps involved in the process of generation of the gradient (see text for details).

- Impermeability to water and ability to actively absorb solutes by TAL.
- In renal medulla, all other tubular structures (except ascending limb) are in osmotic equilibrium. The descending limb, therefore acquires the increased osmolality of the surrounding fluid. The effect is multiplied as new iso-osmolar filtrate arrives at the descending limb and forces the concentrated tubular contents towards the tip of loop of Henle (hair pin band).

The process of multiplication of single effect can be best understood by dividing the whole process in hypothetical steps, leading to a normal equilibrium condition. These steps are summarized (Fig. 6.3-4):

- *Initially*, let us assume the osmolality of fluid in descending limb and ascending limb of loop of Henle and medullary interstitium is 300 mOsm/kg H_2O (Fig. 6.3-4A).
- Further, let us assume that the TAL actively pumps out 100 mOsm/kg H_2O of NaCl into the medullary interstitium. Since the TAL is impermeable to water, this effect will lower the osmolality of fluid in TAL to 200 mOsm/kg H_2O and raise the osmolality of the adjacent interstitium to 400 mOsm/kg H_2O (Fig. 6.3-4B). Establishment of this osmotic gradient of 200 mOsm/kg H_2O, as described earlier, is called single effect.
- The portion of DTS which is located in the outer medulla is moderately permeable to Na^+ and highly permeable to water.
- Due to the osmotic gradient created the water moves out and Na^+ from the interstitium (osmolality 400 mOsm/kg H_2O) moves into the DTS (osmolality 300 mOsm/kg H_2O) till equilibrium is reached with osmolality of 350 mOsm/kg H_2O (Fig. 6.3-4C).
- The fresh iso-osmolar filtrate at 300 mOsm/kg H_2O trickles down into the descending loop and pushes some of the hyperosmolar fluid (350 mOsm/kg H_2O) into the ascending limb (Fig. 6.3-4D).
- In the meanwhile, hypotonic fluid flows into the distal tubule and isotonic and subsequently, hypertonic fluid (Fig. 6.3-4E, F) flows into the ascending thick limb.
- As the process with above steps keeps repeating and the final result is a gradient of osmolality from top to bottom of the loop and surrounding interstitium.

Counter current exchanger

Vasa recta (the counter current exchanger). If the vasa recta would have been a straight blood vessel, the osmotic gradient in the medullary pyramid would not last long, as the Na^+ and urea in the interstitial spaces would have been removed by the circulation (Fig. 6.3-5A). However, because of the hair pin (U-shaped) anatomical arrangement of the vasa recta operate as counter current exchanger and retains these solutes in the medullary interstitium (Fig. 6.3-5B). Thus, the counter current exchanger formed by the vasa recta is responsible for the maintenance of the hyperosmolality medullary gradient generated by the counter current multiplier.

Operation at the level of descending vasa recta. As shown in Fig. 6.3-5B, when the descending vasa recta dips down in the medulla having a progressively increasing osmolality (from 300 to 1200 mOsm/kg H_2O), the solutes diffuse into its lumen and water diffuses out so that the blood flowing in it (with osmolality 300 mOsm/kg H_2O) equilibrates with the medullary interstitium and thus its osmolality also increases progressively.

Operation at the level of ascending vasa recta. The vasa recta, then loops around and ascends towards cortex (Fig. 6.3-5B). As at the beginning of the ascending vasa recta, its blood has attained the osmolality of 1200 mOsm/kg H_2O and when it passes through an interstitium where osmolality is progressively decreasing from 1200 to 300 mOsm/kg H_2O; the solutes move out and water diffuses in, and thus the blood in it once again equilibrates with the interstitium around it.

Effect of operation of counter current exchanger on the medullary interstitium. By the above described operations the hypertonicity of the medulla is maintained. Since, solutes (sodium and urea) are exchanged for water between the ascending and descending limbs of vasa recta, this system has been named counter current exchanger.

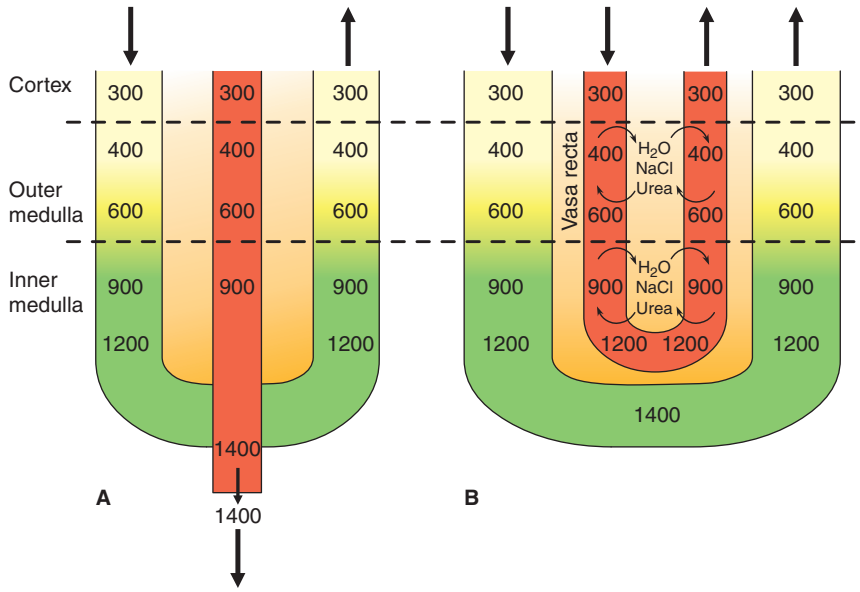


Fig. 6.3-5 Operation of vasa recta as counter current exchangers in the kidney: A, effect on medullary osmolality if vasa recta would be a straight vessel without hairpin arrangement and B, as it is a U-shaped vessel.

Another factor which ensures retention of sodium in the medullary interstitium is very slow rate of blood flow through the medullary parenchyma.

MECHANISM OF URINE DILUTION AND CONCENTRATION

PRODUCTION OF DILUTED URINE

Conditions in which dilute urine is formed

Dilute urine is called hypoosmotic urine, in which urine osmolality is less than blood osmolality. It is produced under following circumstances:

- When circulating levels of ADH are low (e.g. after water drinking), central diabetes insipidus (see page 549),
- When ADH is ineffective (e.g. nephrogenic diabetes insipidus).

Principal factors governing formation of dilute urine

As mentioned earlier, the principal factors governing formation of dilute and concentrated urine are hyperosmolality medullary gradient and the presence or absence of ADH. The renal medullary osmotic gradient is smaller in the absence of ADH. This is because ADH stimulates both counter current multiplication and urea cycling.

PRODUCTION OF CONCENTRATED URINE

Conditions in which concentrated urine is formed

Concentrated urine is also called *hyperosmotic urine,* in which urine osmolality is more than that of blood. It is produced when circulating ADH levels are high, e.g.

- Water deprivation,
- Haemorrhage and
- Syndrome of inappropriate antidiuretic hormone, i.e. SIADH (see page 548).

Principal factors governing formation of concentrated urine

The high level of ADH is the main factor governing the formation of concentrated urine because it:

- Increases the size of the hyperosmolality medullary gradient.
- Augments the urea cycling from the inner medullary collecting ducts into the medullary interstitial fluid.

ADH stimulates NaCl reabsorption in the thick ascending limb and therefore increases the size of medullary gradient by *counter current multiplier.*

ADH increases the H_2O permeability of principal cells of late distal tubule and collecting duct through aquaporin –2 (For details about mechanism of action see page 547).

Note. It is important to note that the tubular response to ADH is not an all-or-none phenomenon, but a graded response. Therefore, varying grades of plasma ADH concentrations can produce proportionate increase in the permeability of collecting ducts to H_2O. Consequently, depending on the status of body water and plasma osmolality, considerable variations in the rate of urine flow and urinary osmolality normally occur in different parts of the day. After an overnight fast, the morning urine samples tend to be relatively more concentrated.

URINE VOLUME AND OSMOLALITY CHANGES IN RESPONSE TO WATER INTAKE AND WATER DEPRIVATION

Water diuresis versus osmotic diuresis

Water diuresis

Water diuresis refers to an increased urinary output following excessive intake of water or hypotonic solution. It occurs due to absence of ADH in the plasma. Steps involved in its occurrence are summarized in Fig. 6.3-6.

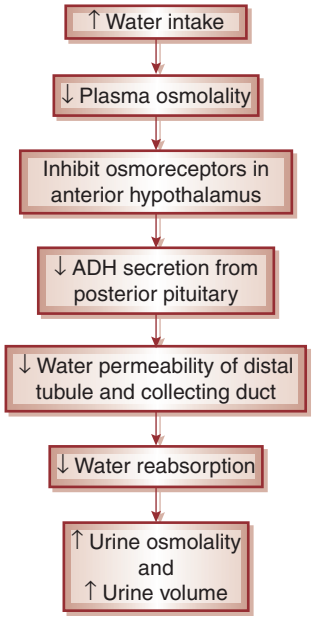


Fig. 6.3-6 Steps involved in urine osmolality changes in response to the increased water intake.

Characteristic features

- Water diuresis begins about 15 min after ingestion of the water and reaches its maximum in 40 min.
- Urine output may be increased to 20 L/day.
- Urine formed is diluted, osmolality is always below 50 mOsm/kg H_2O.
- Specific gravity of urine is always below 1.010.

Osmotic diuresis

Osmotic diuresis refers to an increased urine output because of an osmotic effect. Presence of large quantities of unreabsorbed solutes in the proximal tubules exerts an appreciable osmotic effect.

Characteristic feature and pathophysiology of osmotic diuresis are shown in Table 6.3-3.

Water deprivation

Water deprivation is followed by a sequence of changes which consume water for the body needs. As a consequence urine volume is decreased but urine osmolality is increased (Fig. 6.3-7).

Water intoxication

Excessive ADH secretion leads to water retention (intoxication). For details see page 548.

ASSESSMENT OF RENAL DILUTING AND CONCENTRATING ABILITY

Assessment of renal dilution and concentration process can be made by performing the following tests:

- Measurement of urine osmolality,
- Measurement of urine specific gravity,

Table 6.3-3 Differences of characteristic features and pathology of water and osmotic diuresis

Features	Water diuresis	Osmotic diuresis
Urine characteristics		
I. Volume	> 20 L/day	> 20 L/day
II. Osmolality	50 mOsm/kg H_2O	> 300 mOsm/kg H_2O
III. Specific Gravity	Always < 1.010	Always > 1.010
Pathophysiology and causes	It occurs due to absence/ reduced secretion of ADH. The conditions are: • Diabetes insipidus • Excessive water drinking	It occurs due to large quantity of unabsorbed solutes which exerts osmotic effect. The conditions are: • ↑ filtered load of Na^+, glucose (diabetes mellitus), urea etc. And exceed tubular maximum (Tm) • Mannitol or other polysaccharide administration (substances filtered but not absorbed by tubules) • Excessive NaCl administration
Water absorption in different segments of tubules	*PCT*: Normal *Loop of Henle*: Normal *DCT and collecting duct*: Decreased (due to lack of ADH).	*PCT*: Decreased *Loop of Henle*: Decreased (because concentration gradient is not set up due to decreased absorption in PCT). *DCT and collecting duct*: Decreased (due to increased load even though ADH secretion is normal).

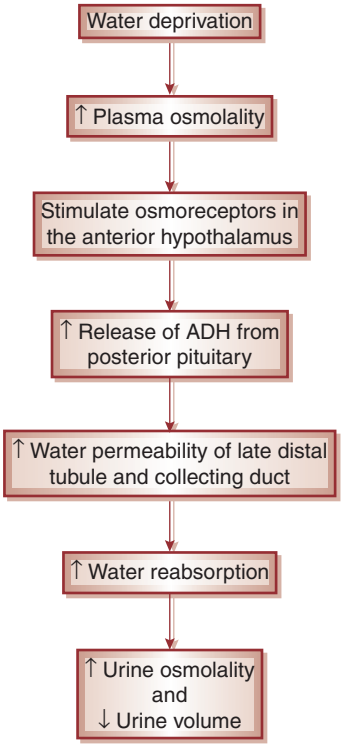


Fig. 6.3-7 Steps involved in urine osmolality changes in response to the water deprivation.

- The urine concentration test,
- The urine dilution test and
- Estimation of free water clearance (C_{H_2O}).

All these tests are described in detail in the subsection on "Kidney function tests", see page 434.

ACIDIFICATION OF URINE

The pH of urine is variable depending upon the concentration of H^+ ions. Under normal circumstances, the pH of urine is acidic (~6.0). This clearly indicates that kidneys have contributed to the *acidification of urine,* when it is formed from the plasma (pH 7.4). In other words, H^+ ions generated in the body in the normal circumstances are eliminated by the acidified urine. Thus, the role of kidneys in the maintenance of acid–base balance of the body (blood pH) is highly significant.

The kidneys regulate the blood pH by three main mechanisms:

- Reabsorption of the filtered HCO_3^-
- Generation of $NaHCO_3^-$ of the alkali reserve of the body and
- Excretion of acid in the form of titrable acid and ammonium ions. All these mechanisms are accomplished through the process of H^+ secretion by the nephron.

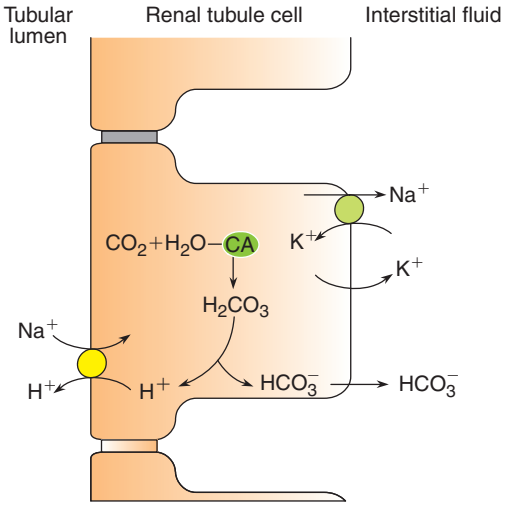


Fig. 6.3-8 Cellular mechanism for secretion of H^+ by proximal tubular cell in the kidney (for details see text).

HYDROGEN ION SECRETION

The tubular cells of the proximal tubule, distal tubule and collecting ducts are capable of secreting H^+.

Mechanism of H^+ secretion by proximal tubule. Steps involved are (Fig. 6.3-8):

- ***Formation of carbonic acid.*** Carbonic acid (H_2CO_3) is first formed in the cells of proximal tubules from CO_2 and H_2O by a reaction that is catalyzed by the intracellular *carbonic anhydrase.* Therefore, the carbonic anhydrase inhibitors depress the secretion of acid by the proximal tubule.
- ***Dissociation of carbonic acid*** (H_2CO_3) then occurs in H^+ and HCO_3^-.
- ***Secretion of H^+ into the lumen*** occurs via *Na^+–H^+* exchange mechanism in the luminal membrane. This is an example of secondary active transport, in which first Na^+ is extruded actively from the cell into the interstitium by Na^+–K^+–ATPase, and intracellular Na^+ is lowered. This is followed by the entry of Na^+ into the cell from the lumen coupled with H^+ secretion into the lumen by *Na^+–H^+*–antiporter (Fig. 6.3-8).
- ***The secreted H^+, in the lumen, combines with the filtered HCO_3^-*** and helps its reabsorption (as described below). Therefore, this process does not result in net secretion of H^+.
- ***HCO_3^- formed in the cell*** (from dissociation of H_2CO_3) diffuses into the interstitial fluid. Thus, for each H^+ secreted one Na^+ ion and one HCO_3^- ion enter the interstitial fluid. The later adds up to the alkali reserve of the body.

Mechanism of H^+ secretion by distal tubules and collecting ducts. In the distal tubule and collecting ducts, H^+ secretion

occurs independent of Na^+. Two mechanisms are involved in secretion of H^+ by the intercalated cells in these parts of tubules are:

- ATP-driven proton pump is mainly responsible for the secretion of H^+ in the distal tubules and collecting ducts.
- H^+, K^+–ATPase is also responsible for secretion of some of the H^+ coupled with reabsorption of K^+ in these parts of renal tubules.

Fate of H^+ secreted in the renal tubule. The secretion of H^+ in the renal tubule can continue only if the H^+ is immediately buffered in the luminal fluid. The tubular cells can secrete H^+, up to a luminal fluid pH of about 4.5, i.e. an H^+ concentration in the urine that is 1000 times the concentration in the plasma. In the absence of buffering of H^+ in the lumen, this pH would be reached rapidly stopping further H^+ secretion. The pH 4.5 is thus the *limiting pH*. However, the free H^+ secreted in the renal tubules are immediately buffered (permitting more acid to be secreted):

- In the proximal tubule, the secreted H^+ is buffered by the filtered HCO_3^- (i.e. consumed in reabsorption of filtered HCO_3^-, vide infra) and
- In the distal tubule and connecting ducts, the secreted H^+ ions are buffered by Na_2HPO_4 and NH_3 and are excreted as titrable acid and ammonium ion (NH_4^+) (vide infra).

REABSORPTION OF FILTERED HCO_3^-

The concentration of HCO_3^- in the plasma and consequently in the glomerular filtrate is about 24 mEq/L. The reabsorption of the filtered HCO_3^- is critically important for the prevention of its loss in the urine and thus for the maintenance of acid–base balance in the body. Under normal circumstances, virtually all the filtered HCO_3^- is reabsorbed by different segments of nephron (Fig. 6.3-9) and none appears in the urine. The mechanisms involved in reabsorption of filtered HCO_3^- in different segments are summarized.

Proximal tubule reabsorbs approximately 80% of the filtered HCO_3^-. Steps of cellular mechanism involved are:

- H^+ secreted in the lumen of proximal tubule (as described above) (Fig. 6.3-8) combines with the filtered HCO_3^- to form carbonic acid (H_2CO_3) (Fig. 6.3-10).
- The H_2CO_3 is rapidly converted to CO_2 and H_2O. This reaction is catalyzed by the brush border carbonic anhydrase.
- The CO_2 diffuses into the tubular cell along the concentration gradient. In the tubular cell, the CO_2 again combines with H_2O to form H_2CO_3 which then dissociates into H^+ and HCO_3^- followed by the secretion of H^+ in tubule and diffusion of HCO_3^- in the interstitial fluid as described above (Figs 6.3-8 and 6.3-10). Thus, for each mole of HCO_3^- reabsorbed from the lumen, one mole of HCO_3^- diffuses from the tubular cell into the blood, even though it is not the same mole that disappeared from the tubular fluid. Further, it is important to note that while there occurs net reabsorption of filtered HCO_3^-, the process neutralizes the H^+ secreted into lumen, meaning thereby that there is no net secretion of H^+ in the lumen and consequently, pH of fluid in the proximal tubule is changed very little.

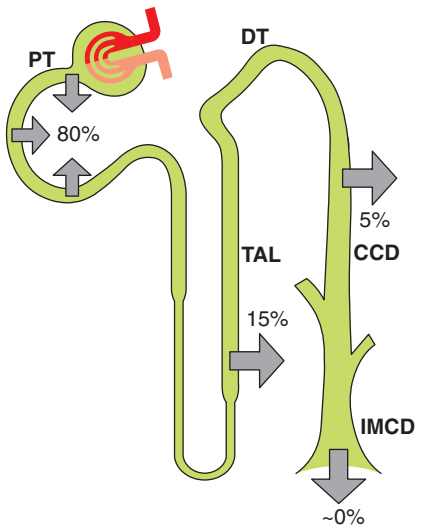


Fig. 6.3-9 Reabsorption of filtered HCO_3^- load along various segments of nephron.

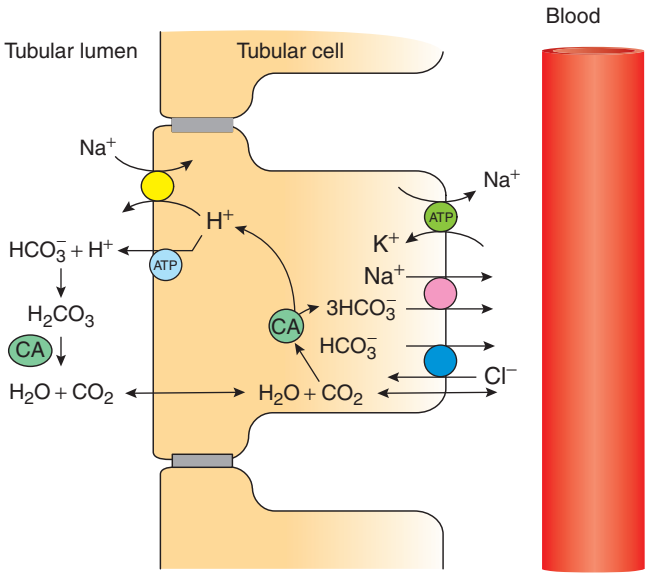


Fig. 6.3-10 Cellular mechanism involved in reabsorption of filtered HCO_3^- in a proximal tubular cell.

Loop of Henle reabsorbs 15% of the filtered HCO_3^-, mainly in the region of TAL. The mechanism involved is exactly the same as described for proximal tubule, except that brush border carbonic anhydrase is not present in the apical membrane of TAL cells.

Distal tubules and collecting ducts reabsorb only 5% of the filtered HCO_3^-, which escapes absorption in the proximal tubule and TAL. So, some of the H^+ secreted by the intercalated cells of these parts of tubule is utilized in the reabsorption of HCO_3^-, while most of the H^+ secreted in these segments is excreted with non-HCO_3^- urinary buffers (described in later discussion).

Regulation of HCO_3^- reabsorption

Various factors that regulate HCO_3^- reabsorption (i.e. H^+ secretion) can be divided into two groups of primary and secondary factors:

Primary factors (those directed at maintaining acid–base balance) involved in regulation of HCO_3^- reabsorption include:

1. Plasma HCO_3^- level. Increase in the plasma HCO_3^- increases the filtered load of HCO_3^- resulting in increased HCO_3^- reabsorption. However, if the plasma concentration becomes very high (above 28 mEq/L) (e.g. metabolic alkalosis), the filtered load will exceed the reabsorptive capacity (the renal threshold for HCO_3^-), the HCO_3^- appears in the urine and urine becomes alkaline.

Conversely with decrease in plasma HCO_3^-, the filtered load is decreased and this results in decreased HCO_3^- secretion. Under such circumstances more H^+ becomes available to combine with other buffer anions. Therefore, lower the plasma HCO_3^- concentration drops, the more acidic the urine becomes and the greater is NH_4^+ content.

2. pCO_2 level when increased results in increased rates of HCO_3^- reabsorption, as the supply of intracellular H^+ for secretion is increased. The reverse happens when pCO_2 level is decreased. These effects of changes in pCO_2 are the physiologic basis for the renal compensation for respiratory acidosis and alkalosis.

Secondary factors (these not directed at maintaining acid–base balance) involved in the regulation of HCO_3^- reabsorption are:

1. Extracellular fluid (ECF) volume. ECF volume expansion (positive Na^+ balance) secondarily results in less H^+ secretion (through Na^+–H^+ antiport) and thus decreased HCO_3^- reabsorption. Conversely, ECF volume contraction (negative Na^+ balance) secondarily results in increased H^+ secretion (through Na^+–H^+ antiport) and thus increased HCO_3^- reabsorption. The aldosterone and angiotensin II are also involved in the changes in Na^+-linked H^+ secretion with changes in the ECF volume.

2. Changes in aldosterone and angiotensin II secondarily affect the HCO_3^- reabsorption by their effect on Na^+ reabsorption and associated H^+ secretion through Na^+–H^+ antiporter.

3. Parathyroid hormone (PTH) also inhibits HCO_3^- reabsorption by proximal tubules. PTH is mainly involved in the maintenance of Ca^{2+} and phosphate balance (see page 573). However, PTH also inhibits the Na^+–H^+ antiporter in the apical membrane of proximal tubule cells.

4. Plasma K^+ levels also influence the secretion of H^+ by the proximal tubules, with hypokalaemia stimulating and hyperkalaemia inhibiting secretion.

GENERATION OF NEW HCO_3^-

As discussed above, the kidneys play an important role in the maintenance of acid–base balance of the body by completely reabsorbing the filtered HCO_3^-. However, in reality HCO_3^- reabsorption alone does not replenish the HCO_3^- lost during the titration of non-volatile acids which are daily added to the plasma, from the diet and produced by metabolism. Therefore, to maintain acid–base balance, the kidneys replace this lost HCO_3^- with new HCO_3^- by following processes:

- Excretion of H^+ as titrable acid and
- Excretion of H^+ as NH_4.

Excretion of H^+ as titrable acid

Excretion of H^+ as titrable acid refers to the excretion of secreted H^+ along with the primary urinary buffer the dibasic phosphate (HPO_4^{2-}). This reaction occurs in the distal tubules and collecting ducts, because it is here that the phosphate which escapes proximal reabsorption is greatly concentrated by the reabsorption of water. The mechanism involved in and net excretion of H^+ with dibasic phosphate urinary buffer is (Fig. 6.3-11):

- H^+ and HCO_3^- are produced in the cell from CO_2 and H_2O.

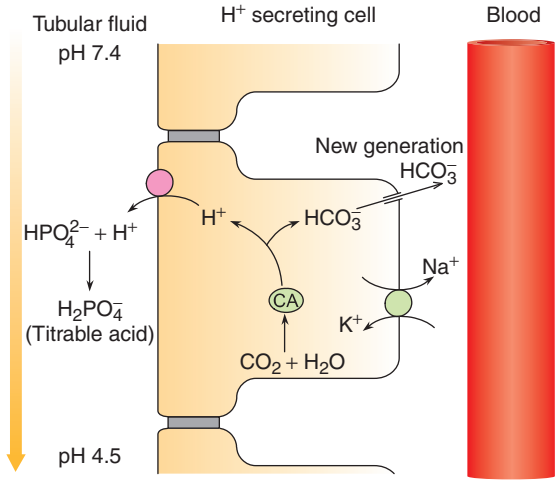


Fig. 6.3-11 Mechanism for excretion of H^+ as titrable acid and synthesis of new HCO_3^-.

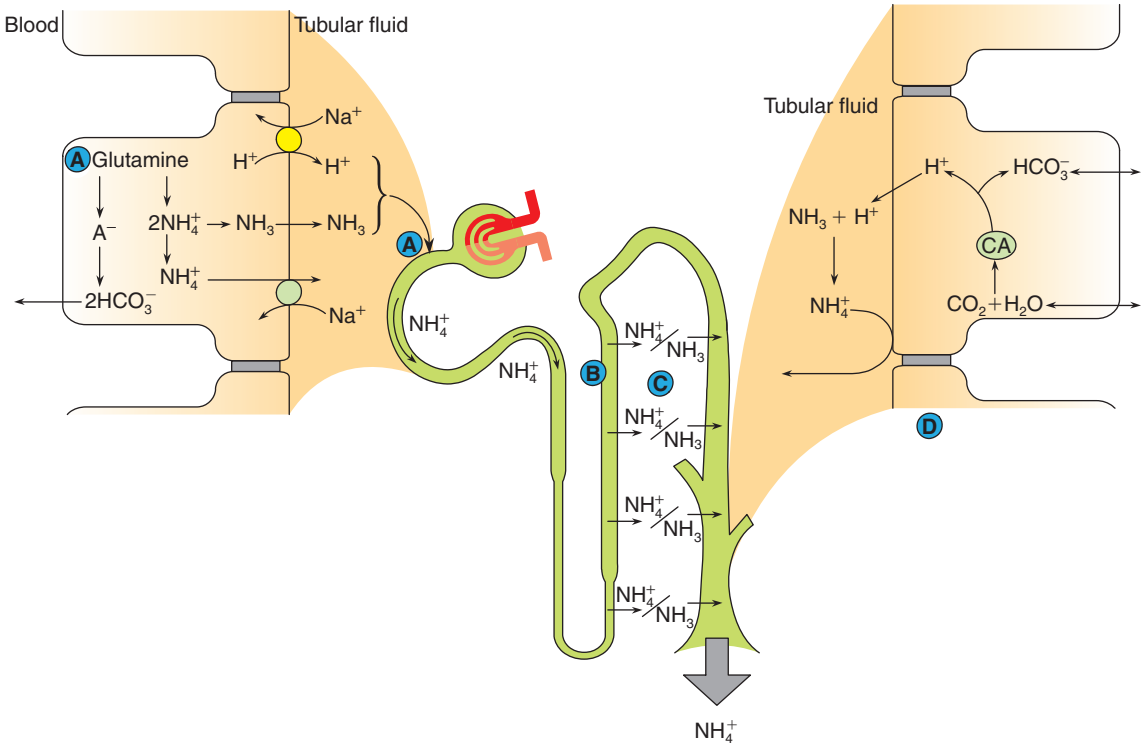


Fig. 6.3-12 Excretion of H^+ as NH_4^+ and generation of new HCO_3^- can be considered in four stages: A, synthesis of NH_4^+ and HCO_3^- from glutamine in the proximal tubule; B, reabsorption of NH_4^+ across the thick ascending limb; C, accumulation of NH_4^+ in the medullary interstitium in equilibrium with NH_3^+ and D, anionic diffusion and diffusion trapping in the collecting ducts.

- The new HCO_3^- is reabsorbed into the blood.
- H^+ secreted into the lumen (mainly of by H^+–ATPase) combines with filtered HPO_4^{2-} to form $H_2PO_4^-$, which is excreted as a titrable acid.

As a result of H^+ excretion in the form of a titrable acid, the pH of urine is progressively decreased (from 7.4, that of blood). The *acidification of the urine* may lower its pH to a minimum of 4.5, i.e. H^+ concentration of urine is approximately 1000 times the concentration of H^+ in the plasma. Thus, the titrable acidity is a measure of acid excreted in the urine by the kidney.

Excretion of H^+ as ammonium ion

Excretion of H^+ as NH_4^+ is another mechanism of excretion of secreted H^+ and formation of new HCO_3^-. The amount of H^+ excreted as NH_4^+ depends upon both the amount of NH_3 synthesized by renal cells and the urine pH. The process by which the kidneys excrete NH_4 is complex and can be described to have four stages (for the purpose of understanding only) (Fig. 6.3-12):

A. Synthesis of NH_4^+ and new HCO_3^- in proximal tubule. Ammonium (NH_4^+) is produced in the cells of proximal tubules from the metabolism of *glutamine.* Each molecule of glutamine is metabolised into two molecules each of NH_4^+ and HCO_3^-.

- HCO_3^- diffuses across the basolateral membrane into the peritubular blood as new HCO_3^-.
- NH_4^+ is secreted into the lumen via Na^+–H^+ antiporter. Some NH_4^+ is converted into NH_3^+ and H^+. NH_3^+ diffuses into the lumen where it combines with the secreted H^+ to form NH_4^+.

B. Reabsorption of NH_4^+ across thick ascending limb. NH_4^+ then moves along the tubular fluid. In the TAL of loop of Henle, a significant amount of NH_4^+ is reabsorbed via two mechanisms:

- *Transcellularly*, via $1Na^+$–$1K^+$–$2Cl^-$ symporter with NH_4^+ substituting for K^+ and
- *Paracellularly* driven by the lumen positive transepithelial voltage in this segment.

C. Accumulation of NH_4^+ in medullary interstitium. The NH_4^+ reabsorbed across the TAL accumulates in the medullary interstitium, where it exists in chemical equilibrium with NH_3^+.

D. Anionic diffusion and diffusion trapping in collecting ducts. The cells of collecting duct are not permeable to NH_4^+, but permeable to NH_3^+. From the medullary interstitium, NH_3^+ diffuses into the lumen of collecting ducts by a process called non-ionic diffusion and is protonated to NH_4^+ by combining with H^+ secreted by the cells of the collecting duct. Since the cells of the collecting ducts are impermeable to NH_4^+, so NH_4^+ is trapped in the lumen of the collecting duct *(diffusion trapping)* and is excreted in the urine. Thus, for every NH_4^+ excreted in the urine, a new HCO_3^- is returned to the systemic circulation.

Chapter 6.4

Regulation of Body Fluid Osmolality, Composition and Volume

INTRODUCTION

The control of body fluid osmolality, composition and volume, and thus the water and electrolyte balance in the body is concerted function of the kidneys, blood, skin, lungs, digestive system, certain hormones and neural mechanisms. However, the kidneys play major role in these homeostatic mechanisms. For the ease of understanding, this chapter has been divided into following sections:

- The body fluid compartments,
- Control of body fluid osmolality,
- Regulation of extracellular fluid volume and composition, and
- Water and electrolyte disturbances.

BODY FLUID COMPARTMENTS

VOLUMES AND COMPOSITION OF BODY FLUID COMPARTMENTS

The total body water (TBW) and the body fluid compartments have been described in Chapter 1.1 (see page 4). The salient features are summarized here:

- Total body water, intracellular fluid (ICF) and extracellular fluid (ECF), respectively form 60%, 40% and 20% of the total body weight (60–40–20 rule).
- Percentage of TBW is highest in newborns and adult males and lowest in adult females and in adults with a large amount of adipose tissue.
- The major cation of ECF is Na^+, while that of ICF is K^+.
- The major anions of ECF are Cl^- and HCO_3^-, while that of ICF are protein and organic phosphates [adenosine triphosphate, adenosine diphosphate and adenosine monophosphate].

CONTROL OF BODY FLUID OSMOLALITY

At steady state, the major fluid compartments of the body, i.e. ECF and ICF are in osmotic equilibrium and thus have the same osmolality. This equilibrium is maintained by free water shifts between the ECF and the ICF compartments. Therefore, a measurement of plasma osmolality provides a measure of both ECF and ICF osmolality. It is important to note that though the osmolality of ECF and ICF osmolality is similar, there is marked difference in the concentration of electrolytes (cations and anions) between the ECF and the ICF (see page 6).

Normal plasma osmolality ranges from approximately 280–295 mOsm/kg H_2O. Sodium and its associated anions make the largest contribution (~90%) to plasma osmolality.

Computation of plasma osmolality, for practical purposes can be, done from the concentration (mmol/L) of Na^+, K^+, urea and glucose:

Plasma osmolality = 2 (Na^+) + 2 (K^+) + urea + glucose.

The factor of two is used for Na^+ and K^+ ions to account for the associated anion concentration. Since plasma Na^+ is the most predominant contributor to osmolality, the above calculation can be simplified (provided plasma glucose and urea are in the normal range) as follows:

Plasma osmolality (mmol/kg) = 2 × Plasma Na^+ (mmol/L).

Plasma (Na^+) and ECF. It is important to realize that Na^+ and its associated anions (mainly Cl^-) are mainly confined to the ECF. Therefore, the retention of water in the ECF is directly related to the osmotic effect of these ions. *Thus, the amount of Na^+* in the ECF ultimately determines its volume. When evaluating abnormal plasma (Na^+) in an individual, it is tempting to suspect a problem in Na^+ balance. However, the problem most often elates to water balance, not Na^+ alterations in the volume of ECF, not its osmolality.

WATER BALANCE IN THE BODY

The kidneys possess tremendous capacity to regulate the body water balance. In a healthy individual, this is achieved by balancing the daily water input and output.

Water input

Water is added to the body fluids by:

1. Ingestion of water in the form of fluid and as constituent of foodstuffs (Table 6.4-1). The water intake is highly variable, which may range from 0.5 to 2.0 L/day depending upon the social and personal habits and environmental conditions. In general, people living in hot climate drink more water. Ingestion of water is mainly controlled by the thirst centre. Increase in the plasma osmolality stimulates thirst centre and promotes water ingestion.

2. Endogenous production of water during oxidation of foodstuffs adds about 300 mL of water to the body fluids per day (Table 6.4-1).

Water output

A variable amount of water is lost from the body in urine, faeces, sweat and as insensible loss (Table 6.4-1).

1. Insensible loss of water (about which the individual is unaware) occurs by evaporation from the cells of skin and respiratory passages.

2. Water loss in sweat. Water loss by the production of sweat from skin can vary from 100 mL/day in routine at room temperature of 23°C to 1400 mL in hot weather to 5000 mL following prolonged exercise.

Table 6.4-1 Water balance in the body represented by daily water input and output in adults at room temperature (23°C)

Water input (mL/day)		Water output (mL/day)	
Source	**Volume**	**Route**	**Volume**
Ingested water	1200*	Insensible	700
In food	1000	Sweat	100
Metabolic water	300	Faeces	200
		Urine	1500**
Total	2500		2500

*Fluid ingestion varies from 1000 to 2000 mL/day, obligatory water ingestion is 400 mL/day.
**Urine flow varies depending upon the water ingested, obligatory urine volume is 500 mL/day.

3. Water loss in faeces. Most of the water entering the gastrointestinal tract (GIT) is reabsorbed by the intestine. About 200 mL/day is lost through faeces in a healthy individual (Table 6.4-1). Faecal loss of water is tremendously increased in diarrhoea. GIT water losses can also occur with vomiting.

4. Water loss in urine. About 1500 mL of water is eliminated from the body in urine. Water losses through kidney are highly variable.

It is important to note that water loss in sweat, faeces and evaporation from the lungs and skin is *not well regulated*. However, the renal loss though variable but is very well regulated to maintain the water balance. For water balance the water output is precisely matched with the water intake by the kidneys:

- The kidneys produce a small amount of concentrated urine (hyperosmotic with respect of plasma), when the intake is low or losses are more, and conversely,
- The kidneys produce large amount of dilute urine (hypoosmotic with respect to plasma), when the water intake is high.

Thus, in a normal individual, depending primarily on the ADH concentration the urine osmolality may vary from 50 to 1200 mOsm/kgH_2O with a corresponding urine volume of 18–0.5 L/day.

Disorders of water balance alter body fluid osmolality. Changes in body fluid osmolality are manifested by a change in plasma (Na^+):

- *Positive water balance* (intake > excretion) results in a decrease in body fluid osmolality and hyponatraemia.
- *Negative water balance* (intake < excretion) results in an increase in body fluid osmolality and hypernatraemia.

Factors controlling water balance

The input of water (i.e. thirst) and output of water (i.e. urinary excretion) are both controlled by plasma osmolality and blood volume.

1. Plasma osmolality controls water balance by stimulating thirst centre and ADH secretion through the osmoreceptors as mentioned below. Significant changes in ADH secretion occur by a small (2%) change in the plasma osmolality.

2. Blood volume. Under normal circumstances, the water balance of the body is mainly regulated through osmoreceptors. However, a significant (more than 10%) decrease in blood volume also stimulates thirst and ADH secretion from the posterior pituitary. If blood volume is markedly decreased, ADH release is stimulated even if plasma osmolality is low. Decreased blood volume is sensed by low pressure receptors in the atria and the pulmonary vessels.

For further details see pages 258.

MECHANISMS CONTROLLING BODY FLUID OSMOLALITY

The control of body fluid osmolality, i.e. defence of the tonicity of the ECF is primarily the function of the vasopressin secretion and thirst mechanisms:

When the effective osmotic pressure of plasma rises, the *osmoreceptors* located in the anterior hypothalamus are stimulated. The stimulation of receptors results from their shrinkage caused by cellular dehydration. Cellular dehydration may occur because of:

- Deficiency of total body water, or
- Excessive intake of NaCl, which causes the water to shift from ICF to ECF.

The *stimulated osmoreceptors* in turn cause (Fig. 6.4-1):

- Increase in thirst which regulates water intake and
- Increase in vasopressin release which regulates water excretion by the kidneys.

When the effective osmotic pressure of plasma decreases, i.e. when plasma becomes hypotonic, the vasopressin secretion is decreased (increasing water excretion in excess of solute) and the thirst is decreased (decreasing water intake). Thus the main mechanisms which are related with regulation of body fluid osmolality are:

- Role of antidiuretic hormone (see page 547),
- Role of thirst mechanism (see page 745) and
- Renal mechanisms for dilution and concentration of urine (see page 402).

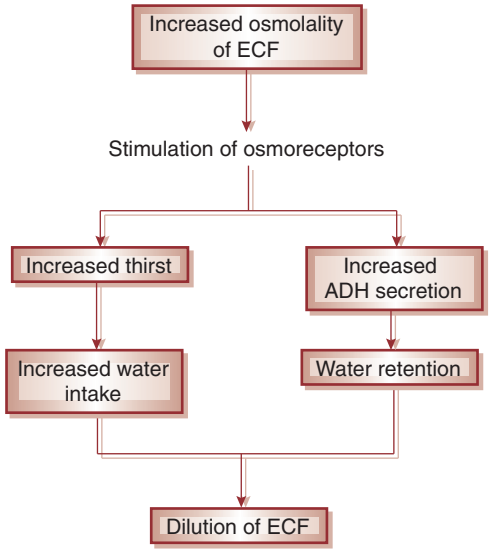


Fig. 6.4-1 Mechanisms regulating body fluid osmolality.

REGULATION OF EXTRACELLULAR FLUID VOLUME AND COMPOSITION

The regulation of ECF volume is primarily mediated through the regulation of the amount of osmotically active solute in it. The major solutes of the ECF are the salts of Na^+; of these NaCl is the most abundant. Since the kidneys are a major route of NaCl excretion from the body, as such, they play an important role in the regulation of ECF. The kidneys get signals from the *volume sensing system* to make appropriate adjustment in NaCl excretion. The volume sensors generate signals in response to changes in *effective circulatory volume* (ECV). The volume sensor signals then control the volume of ECF by controlling the renal excretion of NaCl and water. Therefore, the process of regulation of ECF volume can be described in following subsections:

- Concept of effective circulating volume and volume-sensing system,
- Volume sensor signals, neural and hormonal control of NaCl excretion and sodium balance.

EFFECTIVE CIRCULATORY VOLUME AND VOLUME SENSORS

CONCEPT OF EFFECTIVE CIRCULATING VOLUME

Effective circulatory volume (ECV). In physiologically conceptual term, the ECV refers to the portion of ECF volume that is present in the arterial system under particular pressure and is effectively perfusing the tissues. About 0.7 L of vascular volume (i.e. 20% of plasma, or 5% of ECF or 1.7% of TBW or 1% of the body weight) forms the ECV. The ECV is regulated by the volume sensors which are located entirely

within the vascular tree. Sensation of ECV is related to 'fullness', i.e. 'volume and pressure' in the vascular tree.

In the normal state, variations in ECV are reflected as parallel variation in the ECF volume and also in the vascular volume, arterial blood pressure and cardiac output. In other words, a decrease in ECF, vascular volume, arterial pressure or cardiac output will be sensed in the body as a decrease in ECV.

Maintenance of ECV and regulation of Na^+ balance are closely related. Therefore, Na^+ loading produces expansion of ECV and Na^+ loss leads to depletion of ECV. In other words, when ECV is decreased, the renal NaCl excretion is reduced. This adaptive response restores the ECV to normal and thereby maintains adequate tissue perfusion. Conversely, an increase in ECV results in an enhanced renal NaCl excretion termed the *natriuresis*. Again this is an adaptive response to restore the ECV to its normal point.

ECV SENSORS (VOLUME-SENSING SYSTEM)

The ECV sensors, commonly known as volume sensors or volume receptors, refer to the receptors which are located in the vascular system and respond to the degree of stretch of the vessel wall and not to the volume of the vessel (hence also called as baroreceptors). Various volume sensors known are:

I. Vascular volume receptors

A. Low pressure volume receptors are located in:
- Cardiac atria and
- Pulmonary vasculature.

B. High pressure volume receptors include:
- Carotid sinus (see page 256),
- Aortic arch (see page 256) and
- Juxtaglomerular apparatus of kidney (see page 380).

II. Hepatic sensors, though not as important as vascular volume receptors, but do play a role in regulating renal NaCl excretion by reflexly regulating renal sympathetic discharge. The hepatic sensors also appear to be involved in the regulation of gastrointestinal Na^+ absorption. For example, when the Na^+ concentration of the portal vein blood is increased, there is reflex reduction in the jejunal NaCl absorption.

NEURAL AND HORMONAL REGULATION OF RENAL SODIUM CHLORIDE EXCRETION

Both neural and hormonal volume sensor signals arise in response to the afferents from the above described volume sensors and regulate the renal excretion of NaCl (Table 6.4-2).

Table 6.4-2 Summary of neural and hormonal control of renal NaCl and water excretion

Neural control

↑ *Sympathetic activity decreases NaCl excretion by:*
- ↓ Glomerular filtration rate
- ↑ Renin secretion
- ↑ Tubular NaCl reabsorption

Hormonal control

↑ *Renin angiotensin-aldosterone secretion decreases NaCl excretion by:*
- ↑ Proximal tubule absorption and
- ↑ ADH secretion by angiotensin-II
- ↑ Tubular reabsorption by aldosterone

↑ *ANP secretion increases NaCl excretion by:*
- ↑ Glomerular filtration rate
- ↓ Renin secretion
- ↓ Aldosterone secretion
- ↓ NaCl and water reabsorption by the collecting duct
- ↓ ADH secretion and action of ADH on the collecting duct

↑ *ADH secretion decreases H_2O excretion by:*
- ↑ H_2O absorption by the collecting duct

NEURAL REGULATION

As described in Chapter 6.1 (page 382), the sympathetic nerve fibres innervate the afferent and efferent arterioles of the glomerulus as well as nephron cells.

Renal sympathetic stimulation, which is induced by the vascular low-and high-volume sensors, in conditions of negative Na^+ balance (i.e. decreased ECV) leads to decrease in NaCl excretion by following mechanisms:

1. Reduction in glomerular filtration rate (GFR) occurs due to vasoconstriction of the afferent arterioles induced by sympathetic stimulation. Reduced GFR leads to a decrease in the filtered load (Filtered load = GFR × plasma Na^+ concentration). Reduction in filtered load do help in Na^+ conservation, however, changes in filtered load of Na^+ are not reflected in parallel changes in the urinary Na^+ excretion because of the following effects:

- Glomerular tubular balance and
- Tubuloglomerular feedback (see page 384).

2. Increased Na^+ reabsorption, along the nephron, is directly produced by stimulation of α-adrenergic receptors. Proximal tubule is the most important segment influenced by the sympathetic nerve stimulation.

3. Stimulation of renin secretion from the cells of afferent arterioles is produced by the activation of β-adrenergic receptors. As described below, it results in an increased plasma concentration of angiotensin II and aldosterone.

Renal sympathetic inhibition, which is induced by the vascular low- and high-pressure volume receptors in conditions of positive Na^+ balance (i.e. increased ECV) leads to increased NaCl excretion by the reverse effect of the above described mechanisms.

HORMONAL REGULATION

1. Renin–angiotensin–aldosterone system. Renin–angiotensin–aldosterone system when stimulated with volume depletion, results in decreased NaCl excretion; conversely, when suppressed with volume expansion enhance NaCl excretion.

Renin is secreted by smooth muscles of afferent arterioles of kidney in response to:

- Reduced perfusion pressure,
- Increased renal sympathetic discharge (as mentioned above) and
- Decreased delivery of NaCl to the macula densa cells (Tubuloglomerular feedback mechanism) (see page 384).

Renin converts angiotensinogen (produced by liver) into angiotensin I which is converted to angiotensin II by angiotensin converting enzyme (Fig. 6.4-2).

Angiotensin II subserves following physiological functions:

- Stimulates aldosterone secretion by the adrenal cortex,
- Increases blood pressure by the arteriolar vasoconstriction,
- Stimulates ADH secretion and thirst, and
- Enhances NaCl reabsorption by the proximal tubules.

Aldosterone enhances NaCl reabsorption from the thick ascending limb of loop of Henle by stimulating Na^+–K^+–$2Cl^-$ symporter at apical membrane and Na^+–K^+–ATPase at the basolateral membrane.

2. Atrial natriuretic peptide is released from the atrial myocytes by the atrial stretch caused by volume expansion (increased ECV). Its actions are opposite to that of renin–angiotensin–aldosterone system and result in an increased urinary excretion of NaCl by following mechanisms:

- *Increased GFR* by vasodilation of afferent and vasoconstriction of efferent arterioles.
- Inhibition of renin secretion by the afferent arterioles.
- *Inhibition of aldosterone secretion* by inhibiting renin secretion and also by its direct effect on the adrenal cortex cells.
- *Inhibition of ADH secretion* by posterior pituitary and inhibition of ADH action on the collecting duct.
- *Inhibition of NaCl reabsorption* by the collecting duct by inhibiting aldosterone secretion, by inhibiting ADH secretion and also by its direct effect on the collecting duct cells.

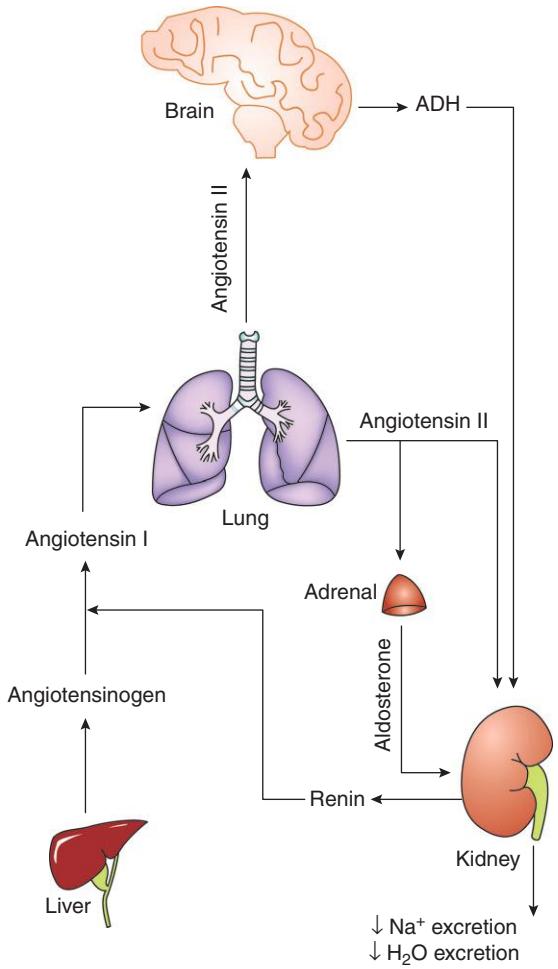


Fig. 6.4-2 Mechanism of decreased Na^+ excretion by the renin–angiotensin–aldosterone system (for explanation see text).

3. Antidiuretic hormone secretion by the posterior pituitary is increased with volume depletion (decreased ECV) leading to retention of water by the kidneys and thus re-establishing euvolaemia. Reverse occurs in volume expansion (see pages 407 and 408).

An integrated neural and hormonal regulation is summarized in Fig. 6.4-3.

DISTURBANCES OF VOLUME AND CONCENTRATION OF BODY FLUID

Broadly, disturbances of fluid volume include:

- *Dehydration* (i.e. fluid loss) and overhydration (i.e. fluid gain): Depending upon the osmolality relative to plasma, a fluid may be:
- *Hypotonic fluid*, i.e. the fluid having osmolality less than that of plasma, such solutions cause the cells to swell and if sufficiently dilute, to burst (lyse).

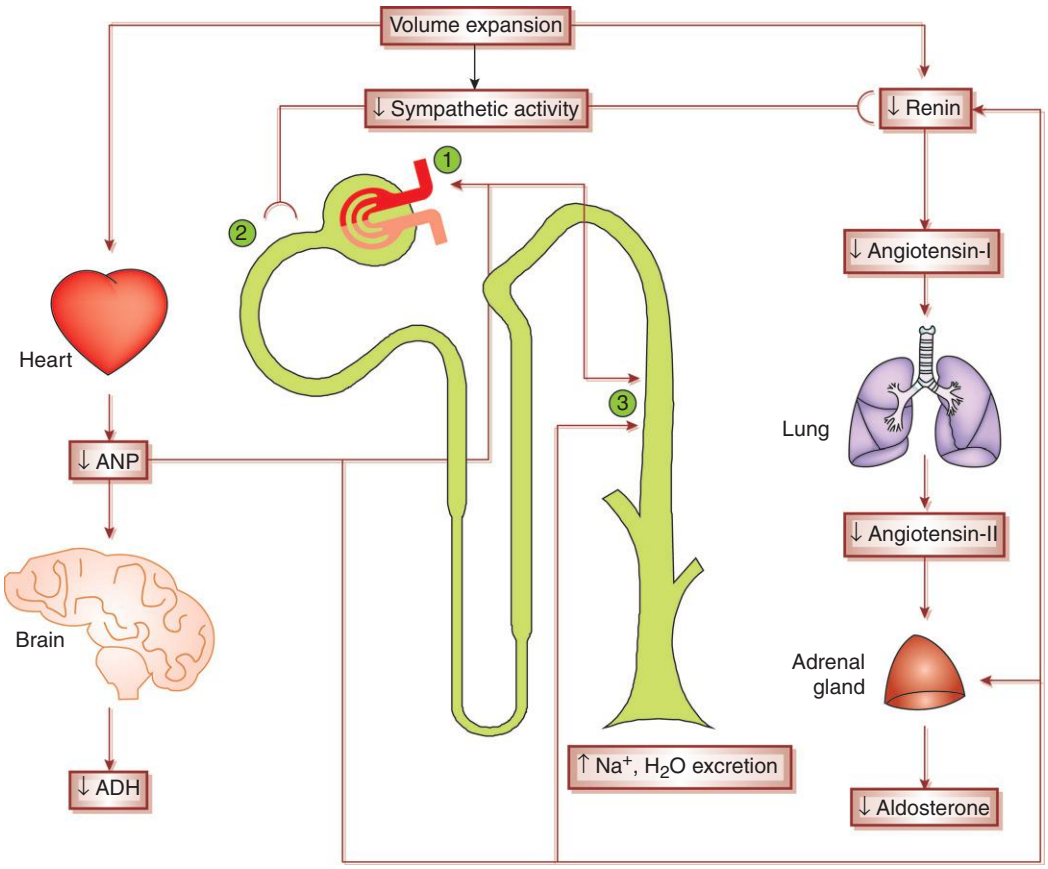


Fig. 6.4-3 An integrated neural and humoral responses to volume expansion: increased glomerular filtration rate (1); decreased sodium reabsorption in proximal tubule (2); and decreased sodium absorption in collecting duct (3). **Note** a response just reverse of the above occurs in response of volume contraction.

Table 6.4-3 Summary of disturbances of volume and concentration of body fluids

Type of disturbance	Volume (L)		Osmolality (mOsm/L)		Causes	Consequences	Corrective response
	ICF	ECF	ICF	ECF			
DEHYDRATION							
Iso-osmotic contraction	----	↓	---- (Fig. 6.4-4B)	----	• Water loss due to: – Diarrhoea, – Vomiting, – Haemorrhage, – Burns and – Ascites	– Plasma protein concentration *increases* – Haematocrit increases – Arterial BP falls	Decrease plasma volume → inhibition of volume sensors → reflexly **restoration of plasma volume** due to decrease excretion of Na^+ and water **Note.** Thirst sensation is quenched by drinking isotonic salt solution
Hyperosmotic contraction	↓	↓	↑ (Fig. 6.4-4D)	↑	• Water loss due to: – Decreased water intake – Excessive sweating	– Increase plasma protein concentration due to decreased ECF volume – Haematocrit increases	↑ ECF osmolality → stimulates osmoreceptors → decreased plasma volume → **restoration of plasma volume and osmolality** → inhibit volume receptors

Hyperosmotic contraction				– Diabetes (mellitus and insipidus) – Alcoholism – Tracheostomy patients (if loss > 500 mL)		
Hypo-osmotic contraction	↑	↓	↓ ↓ (Fig. 6.4-4F)	• Loss of NaCl or hypertonic fluid from body, e.g. vomiting • Adrenocortical insufficiency (Addison's disease)	– Increase plasma protein concentration causes increase oncotic pressure therefore, shift of fluid from plasma to interstitial fluid. – However in grave salt depletion plasma volume decreases. – Thirst is inhibited because thirst cells swell up – Haematocrit increases due to decreased ECF – Arterial BP decreases	– The ECF volume decreased but thirst is absent – Salt craving or salt appetite stimulates person to consume more NaCl → ECF osmolality restoration → shift of water from ICF to ECF → shrinkage of thirst centre cells → thirst stimulation → drinking of water → **normalization of plasma volume and osmolality**
OVERHYDRATION						
Iso-osmotic expansion	---	↑	---- ---- (Fig. 6.4-4A)	• Infusion of isotonic fluids (0.9% NaCl)	– Plasma protein concentration decreases – Haematocrit decreases – Arterial pressure increases (due to increase in ECF volume)	– Change in plasma volume sensed by volume receptors → excretion of large volume of hypotonic urine (water diuresis) → **normalisation of ECF volume**
Hyperosmotic expansion	↓ (Due to fluid shift from ICF to ECF)	↑	↑ ↑ (equalisation with ECF) (Fig. 6.4-4C)	• Administration of excessive amount of hypertonic saline	– Plasma protein concentration decreases – Haematocrit decreases – Arterial pressure increases (due to increase in ECF volume)	– Increase plasma osmolality promotes water retention – increase in plasma volume – thirst and ADH secretion is suppressed (*increase plasma volume oversides osmolality*) → excretion of excessive hypotonic urine → **plasma volume normalised** – ANP secretion increases → promote Na excretion → **osmolality normalised**
Hypo-osmotic expansion	↑ (shift of water from ECF to ICF)	↑ (due to water retention)	↓ ↓ (Fig. 6.4-4E)	• SIADH (syndrome of inappropriate ADH secretion), and • ingestion of large volume of water	– Water shifts from ECF to ICF → decreased ICF osmolality – Plasma protein concentration decreases – Haematocrit remains unchanged (because water shifts from plasma into RBCs)	Vascular volume receptors and osmoreceptors sense volume changes → excretion of large amount of hypotonic urine → **volume and osmolality normalised**

ICF = intracellular fluid; ECF = extracellular fluid; ↑ = increased; ↓ = decreased; --- = no change.

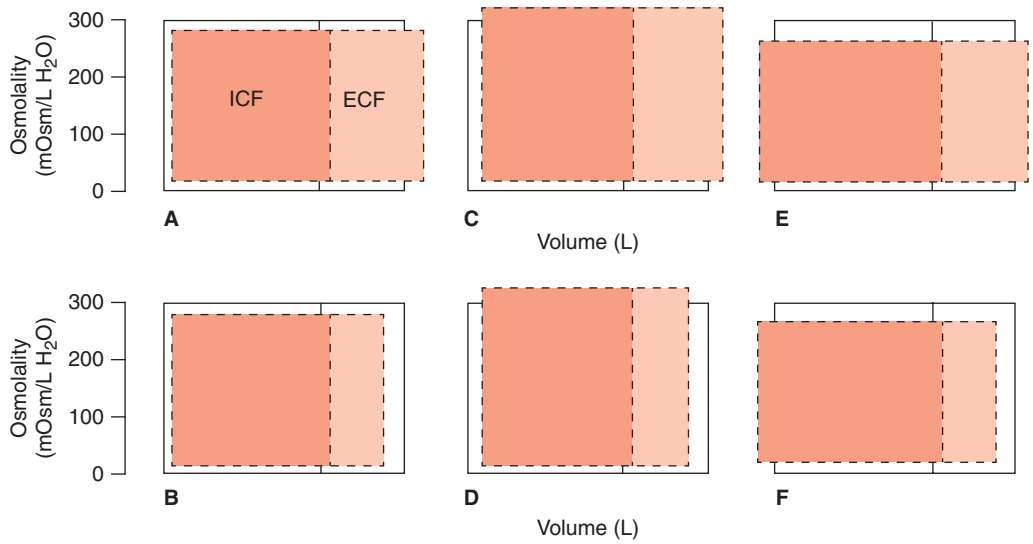


Fig. 6.4-4 Volume and osmolality of ECF and ICF and shift of water between two compartments during disturbances of fluid volume; A, iso-osmotic volume expansion; B, iso-osmotic volume contraction; C, hyperosmotic volume expansion; D, hyperosmotic volume contraction; E, hypo-osmotic volume expansion and F, hypo-osmotic volume contraction. Volume and osmolality of normal ECF and ICF are indicated by solid lines. Changes in volume and osmolality as consequences of various situations are indicated by dashed lines.

- *Hypertonic fluid,* i.e. the fluid having osmolality more than that of plasma such solutions cause cells to shrink (i.e. to undergo crenation).

Disturbances of fluid volume and concentration can be classified as:

- Iso-osmotic volume expansion,
- Iso-osmotic volume contraction,
- Hyperosmotic volume expansion,
- Hyperosmotic volume contraction,
- Hypo-osmotic volume expansion and
- Hypo-osmotic volume contraction.

The causes, consequences and corrective responses of disturbances of volume and concentration in dehydration and overhydration are summarized in Table 6.4-3 (Fig. 6.4-4).

Chapter 6.5

Physiology of Acid–Base Balance

GENERAL CONSIDERATIONS

ACIDS AND BASES

Acids and bases. Acid refers to a substance that acts as proton (H^+) donor while base refers to a substance that accepts proton (H^+). Examples of a few acids and their corresponding base are:

Acid	*Base*
HCl	$H^+ + Cl^-$
H_2CO_3	$H^+ + HCO_3^-$

Alkalies refer to the metallic hydroxides, e.g. NaOH and KOH. These compounds do not directly satisfy the criteria of bases. However, they dissociate to form metallic ion and OH^- which being the base accepts H^+ ions. Therefore, for all practical purposes, alkalies are considered bases.

Strong acid and bases. Acid or base having strong tendency to dissociate into ions is called strong acid or strong base; and the acid or base having weak tendency to dissociate into ions is called weak acid or weak base. In general, a strong acid has a weak base, while a weak acid has a strong base. For example, strong acid HCl has weak base Cl^- and weak acid HCN has a strong base CN^-.

Ampholytes refer to the substances that can act both as acids and bases. Water is the best example of ampholyte.

Concept of pH and H^+ concentration

H^+ ion concentration. The acidic or basic nature of a solution is measured by H^+ ion concentration. Since the concentration of H^+ ions in the biological fluids is exceedingly low, the conventional units such as mEq/L or moles/L, etc. are not commonly used to express H^+ concentration. Therefore, ***pH*** is the term suggested to express H^+ ion concentration. pH is defined as the negative logarithm of H^+ concentration.

$$pH = -\log [H^+]$$

It is important to note that pH and H^+ are inversely related. For example, pH of plasma with a H^+ ion concentration of 0.00004 mEq/L is 7.4, while pH of HCl with H^+ ion concentration of 150 mEq/L is 0.8.

Neutral pH, acidic pH and alkaline pH. Pure water has an equal concentration of H^+ and OH^- ions, i.e. 10^{-7} M each. Thus pure water has pH of 7, which is neutral. Therefore, solutions with pH less than 7 are considered acidic and those with more than 7 are considered alkaline.

H^+ concentration and pH of biological fluids

The H^+ concentration and pH of some biological fluids are depicted in Table 6.5-1.

MAINTENANCE OF BLOOD pH

GENERAL CONSIDERATIONS

BLOOD AND PLASMA pH

- The term blood pH always refers to the plasma pH.
- Normal plasma pH is 7.4 (H^+ concentration –40 mEq/L), which is higher than the intracellular pH of the erythrocyte (7.2).

Table 6.5-1 H^+ concentration and pH of biological fluids

Fluid	H^+ concentration	
	mol/L	pH
1. Pure water	1×10^{-7}	7.0
2. Blood		
Normal mean	3.98×10^{-8}	7.4
Normal range	4.36×10^{-8} to 3.6×10^{-7}	7.36–7.44
Acidosis (severe)	1.26×10^{-7}	6.9
Alkalosis (severe)	2.00×10^{-8}	7.7
3. Cerebrospinal fluid		
CSF (Normal range)	4.36×10^{-8} to 3.6×10^{-7}	7.36–7.44
4. Gastric juice (pure)	1×10^{-1}	1.0
5. Pancreatic juice	1×10^{-8}	8.0
6. Urine		
Normal average	1×10^{-6}	6.0
Maximum acidity	3.16×10^{-5}	4.5
Maximum alkalinity	1×10^{-8}	8.0
7. Intracellular fluid (ICF)	1.58×10^{-7}	6.8

- Plasma pH compatible with life varies from 7.7 to 6.9.
- At rest, normal pH of mixed venous blood is 7.38 compared with 7.41 of arterial blood because of the uptake of CO_2 by blood as it perfuses the tissues.
- Normally, the pH of extracellular fluid (ECF) is maintained between a narrow range of 7.35 and 7.45.

DIETARY AND METABOLIC PRODUCTION OF ACIDS AND BASES

The daily consumed diet contains many acids and alkalies. In addition, cellular metabolism produces a number of acidic and alkaline substances that have an impact on the body pH.

Acid production by the body

The metabolic activities of the body are accompanied by production of two types of acids:

1. Volatile acids. CO_2 is the volatile acid produced from the aerobic metabolism of cells. It is also major end product in the oxidation of carbohydrates, fats and amino acids. CO_2 accounts for over 12,000 mEq/L of H^+ per day. It is considered acid because CO_2 combines with H_2O [by a reaction catalyzed by carbonic anhydrase, i.e. (CA)] to form weak acid H_2CO_3, which dissociates into H^+ and HCO_3^- by the following reaction:

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

It is called volatile, because it is a gas, and under normal circumstances almost all the CO_2 is excreted by lungs.

2. Non-volatile acids, also called fixed acids, contribute about 50–100 mEq H^+ per day, depending upon the diet. These include:

- Sulphuric acid
- Phosphoric acid,
- Hydrochloric acid,
- Organic acids like Lactic acid, Acetic acid and β-hydroxybutyric acid. Uric acid produced in the metabolism of nucleoproteins.

Production of bases by the body

In a normal circumstance, a negligible amount of bases is formed in the body because:

- *HCO_3^-* produced by the metabolism of organic anions (e.g. citrate) offsets non-volatile acid production to some degree.
- *Ammonia* produced in the amino acid metabolism is converted to urea, hence its contribution as a base in the body is insignificant.

DEFENCES AGAINST CHANGES IN H^+ CONCENTRATION

There are three lines of defence to regulate the body's acid–base balance and maintain the blood pH (around 7.4):

I. Buffer systems of body fluid. These:
- Form first line of defence,
- Act instantaneously in ECF and
- Immediately combine with H^+ to prevent changes in H^+ concentration and from a temporary measure to control changes in the H^+ concentration.

II. Respiratory mechanism to regulate acid–base balance
- Forms second line of defence,
- Acts within a few minutes,
- Acts via respiratory centre to regulate removal of CO_2 (and therefore H_2CO_3) and
- Forms a short-term measure to regulate changes in the H^+ concentration.

III. Renal mechanism to regulate acid–base balance
- Forms third line of defence,
- Takes days or weeks,
- Slow, but most powerful and effective in regulating pH,

- Acts by reabsorbing filtered HCO_3^-, generating new HCO_3^- and excreting H^+ as titrable acid and ammonium ion and thus
- Provides a permanent solution to acid–base balance.

BUFFER SYSTEM: PRIMARY DEFENCE

Buffers. A buffer is a solution, consisting of a weak acid and its salt with strong base that prevents a change in pH when H^+ ions are added to or removed from a solution. It must be born in mind that a buffer cannot remove H^+ ions from the body. It temporarily acts as a shock absorbant to reduce the free H^+ ions. The H^+ have to be ultimately eliminated by the renal mechanism.

When acid is added to a buffer solution, its H^+ ion concentration is increased and the reaction is forced towards right leading to an increase in undissociated molecules, therefore, increase in H^+ concentration is less.

When base is added to a buffer, reaction shifts towards left; more H^+ ions are released from the buffer to combine with base, thereby limiting the decrease in H^+ concentration.

Henderson–Hasselbalch equation which is used to calculate the pH in a buffer system can be derived as: The general equation for a buffer is

$$HA \rightleftharpoons H^+ + A^-,$$

A^- represents any anion from a buffer (i.e. H^+ acceptor) and HA the undissociated acid from a buffer (i.e. H^+ donor). By the law of mass action, at equilibrium the product of concentration of dissociates in the chemical reaction divided by concentration of product of the reaction are constant.

$$K = \frac{[H^+][A^-]}{[HA]} \quad (1)$$

(K = Dissociation constant of the acid HA)

The equation to represent free H^+ ion in a solution can be rewritten as

$$[H^+] = K \times \frac{[HA]}{[A^-]} \quad (2)$$

we know that $pH = \log \frac{1}{[H^+]}$

By taking the reciprocals and logarithms (for log, multiplication becomes addition).

$$\log \frac{1}{[H^+]} = \log \frac{1}{K} + \log \frac{A^-}{[HA]}$$

As $\log \frac{1}{K} = pk$,

The equation (3) may be rewritten as

$$pH = pK + \log \frac{[A^-]}{[HA]}$$

From this equation, it is evident that buffering capacity of a buffer system is greatest when the amount of anions $[A^-]$ and undissociated acid [HA] is same, i.e.

$$\frac{[A^-]}{[HA]} = 1, \quad \text{or} \quad \left[\log \frac{[A^-]}{[HA]}\right]$$

thus, pH = pK. Therefore, most effective buffers in the body are those with pK close to the pH in which they operate.

Isohydric principle states that when there is a change in H^+ concentration in the ECF, balance of all the buffer systems changes at the same time. According to this principle, all buffers in a common solution are in equilibrium with the same H^+ concentration, i.e.

$$[H^+] = K_1 \times \frac{HA_1}{A_1} = K_2 \times \frac{HA_2}{A_2} = K_3 \times \frac{HA_3}{A_3}$$

where

- K_1, K_2 and K_3 are dissociation constants of the three acids,
- HA_1, HA_2 and HA_3 are undissociated acids and
- A_1, A_2 and A_3 are the concentrations of free negative anions.

CLASSIFICATION OF THE BUFFER SYSTEMS

Buffer systems in the body can be classified by different methods:

A. Bicarbonate versus non-bicarbonate buffers

1. *Bicarbonate buffer* forms 53% of the buffering in the whole body. Out of it:
   - Plasma HCO_3^- contributes 35% and
   - Erythrocyte HCO_3^- contributes 18%.
2. *Non-bicarbonate buffers* form remaining 47% of the buffering in the whole body. With a contribution from:
   - Haemoglobin and oxyhaemoglobin 35%
   - Plasma proteins 7%
   - Organic phosphate 3%
   - Inorganic phosphate 2%

B. Extracellular versus intracellular buffers

1. *Bicarbonate* (HCO_3^-) is the major extracellular buffer, which is produced from CO_2 and H_2O.
2. *Phosphate* is a minor extracellular buffer. Phosphate is most important as urinary buffer, excretion of H^+ as $H_2PO_4^-$ is called titrable acid.
3. *Plasma proteins* form the non-bicarbonate buffer in the blood and are responsible for 7% of the total buffering of blood.
4. *Haemoglobin,* though found intracellularly, it is more conventionally regarded as part of an extracellular system (as described later).

C. Intracellular buffers

1. *Organic phosphate,* e.g. AMP, ADP, ATP and 2,3-diphosphoglycerate (DPG).
2. *Proteins* of the skeletal muscles.

3. *HCO_3^-* present in an intracellular fluid of skeletal and cardiac muscles.

MAJOR BUFFER SYSTEMS OF THE BODY

The major buffer systems involved in the maintenance of body pH are:

- Bicarbonate buffer
- Phosphate buffer
- Protein buffer

1. Bicarbonate buffer system

The carbonic acid–sodium bicarbonate (H_2CO_3–$NaHCO_3$) is the most predominant buffer system of the extracellular fluid, particularly the plasma H_2CO_3 in the body is formed by CO_2 and H_2O

$$CO_2 + H_2O \xrightarrow[\text{anbydrase}]{\text{carbonic}} H_2CO_3$$

This reaction is catalyzed by the enzyme carbonic anhydrase, which is present in the RBCs, walls of the lungs, alveoli and epithelial cells of renal tubules.

Dynamics of bicarbonate buffer system

Carbonic acid dissociates into hydrogen and bicarbonate ions.

According to the Henderson–Hasselbalch equation for this system

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^- \quad (1)$$

The pK for the system in an ideal solution is low (about 3) and the amount of H_2CO_3 is small and hard to measure accurately. However, in the body, H_2CO_3 is in equilibrium with CO_2, i.e.

$$H_2CO_3 \rightleftharpoons CO_2 + H_2O$$

If the pK is changed to pK_1 and CO_2 is substituted for H_2CO_3, the pK_1 is 6.1

$$pH = 6.10 + \log \frac{[HCO_3^-]}{[CO_2]} \quad (2)$$

Since, the amount of dissolved CO_2 is proportionate to the partial pressure of CO_2 and the solubility co-efficient of CO_2 in mmol/L/mm Hg is 0.0301, the clinically relevant form of this equation is as follows:

$$pH = 6.10 + \log \frac{[HCO_3^-]}{0.0301\ [PCO_2]} \quad (3)$$

$[HCO_3^-]$ cannot be measured directly but can be calculated from the values of pH and pCO_2, which can be measured with suitable accuracy using pH and pCO_2 glass electrodes.

Blood pH and the ratio of HCO_3^- to H_2CO_3. The pH of arterial plasma with normal CO_2 tension (pCO_2) of 40 mm Hg, and normal HCO_3^- concentration of 24 mmol/L can be calculated from the above equation (3) as below:

$$\begin{aligned} pH &= 6.10 + \log \frac{24\ \text{mmol/L}}{.0301 \times 40\ \text{mm Hg}} \\ &= 6.10 + \log \frac{24}{1.2} \\ &= 6.1 + \log 20 \\ &= 6.1 + 1.3 \\ &= 7.4 \end{aligned}$$

The pK_1 of this system (6.1) is still low, relative to the pH of the blood (7.4), but the system is one of the most effective buffer systems in the body because the amount of dissolved CO_2 is controlled by respiration, and plasma concentration of HCO_3^- is regulated by the kidney. Therefore, pH of ECF can be precisely controlled. As this system consist of H_2CO_3 (weak acid), which only partially dissociates into H^+ and HCO_3^-.

Addition of strong acid, e.g. HCl is followed by buffering of H^+ by the following reaction:

$$\uparrow H^+ + HCO_3^- \rightarrow \text{More } H_2CO_3 \rightarrow CO_2 + H_2O$$
$$\downarrow$$
$$\text{Elimination of } CO_2 \leftarrow \uparrow \text{Respiration}$$

Addition of strong base (e.g. NaOH), which is converted into a weak base ($NaHCO_3$), as shown:

$$NaOH + H_2CO_3 \rightarrow NaHCO_3 + H_2O \rightarrow Na^+ + HCO_3^-$$

- As a consequence, the concentration of H_2CO_3 decreases and that of HCO_3^- increases. Therefore, more CO_2 combines with H_2O to form new carbonic acid.

$$H_2O + CO_2 \rightleftharpoons H_2CO_3$$

- Decrease in CO_2 (as above) inhibits the respiration, leading to correction of CO_2 deficiency.
- Increase in HCO_3^- is corrected by an increased renal excretion of HCO_3^-.

Effective buffering of H^+. It is evident that at a blood pH 7.4, the ratio of bicarbonate to carbonic acid is 20%. Thus, the bicarbonate concentration is much higher (20 times) than the carbonic acid in the blood. This is referred to as *alkali reserve* and is responsible for effective buffering of H^+.

2. Phosphate buffer system

Inorganic orthophosphate buffer system

Inorganic orthophosphate buffer system is formed by sodium dihydrogen phosphate and disodium hydrogen phosphate (NaH_2PO_4~Na_2HPO_4), which exist in at a plasma pH of 7.4 in a concentration ratio of 1:4.

Sites of operation of NaH_2PO_4–Na_2HPO_4 buffer

1. In ECF (plasma and interstitial fluid), the HPO_4^{2-}/$H_2PO_4^-$ buffer exists in a small concentration (0.66 mmol/L) and thus contributes little to the buffering capacity of plasma. However, it is important to note that this buffer pair with a pK of 6.8 would be a more effective buffer than HCO_3^-/CO_2 system (pK=6.1) if it were present in an appreciable concentration.

2. In intracellular fluid (ICF), the $HPO_4^{2-}/H_2PO_4^-$ forms an important buffer pair because:

- Its concentration in ICF is high (6 mmol/L) and
- Its pK (6.8) is much closer to pH of ICF (6.9).

3. In renal tubules. In the proximal convoluted tubule approximately 75% of filtered HPO_4^{2-} is absorbed, only 25% of the filtered HPO_4^{2-} is available for buffering in the distal convoluted tubule and the collecting ducts. The HPO_4^{2-}/$H_2PO_4^-$ forms an effective extracellular buffer because:

- Phosphate becomes greatly concentrated in the tubular fluid due to reabsorption of H_2O and
- pH of tubular fluid and urine is more acidic than the pH of ECF, i.e. is close to pK of phosphate buffer.

The $HPO_4^{2-}/H_2PO_4^-$ system is a major elimination route for H^+ via the urine. For details see page 410.

Mechanism of action of $HPO_4^{2-}/H_2PO_4^-$

- This non-bicarbonate buffer system can buffer both non-carbonic and carbonic acid.
- It equilibrates as

$$H_2PO_4^- \rightleftharpoons H^+ + HPO_4^{2-}$$

- *On addition of strong acid,* e.g. HCl, it forms a weak acid

$$HCl + Na_2HPO_4 \rightleftharpoons \underset{\text{(weak acid)}}{NaH_2PO_4} + NaCl$$

- *On addition of strong base,* e.g. NaOH, it forms a weak base

$$NaOH + NaH_2PO_4 \rightleftharpoons \underset{\text{(weak acid)}}{Na_2HPO_4} + H_2O$$

Organic phosphate buffer system

Organic phosphates (such as AMP, ADP, ATP and 2,3-diphosphoglycerate, i.e. (2,3-DPG) exist in quantitatively significant amount in the ICF (8.4 mmol/L), giving this compartment the capacity to effectively buffer both non-carbonic and carbonic acid, as well as alkali.

3. Protein buffer system

The protein buffer system of the blood is constituted by the plasma proteins and haemoglobin combinedly. The buffering capacity of proteins is dependent on the pK of ionizable groups of amino acids. The *imidazole group of histidine* (pK 6.7) is the most effective contributor of protein buffers:

Plasma proteins buffer system

Plasma proteins buffer system accounts for 15% of the buffering capacity of the whole blood. Plasma proteins are effective buffers because both their free carboxyl and free amino groups dissociate

$$RCOOH \rightleftharpoons RCOO^- + H^+;$$

$$pH = pK_1RCOOH + \log\frac{[RCOO^-]}{[RCOOH]}$$

$$RNH_3 \rightleftharpoons RNH_2^- + H^+;$$

$$pH = pK_1RNH_3 + \log[(RNH_2)/(RNH_3)]$$

Because of their amphoteric nature, plasma proteins can combine with acids and bases as:

- *In acidic pH,* the NH_2 group of the proteins acts as base and accept proton and is converted to NH_3.
- *In alkaline pH,* the –COOH group of the proteins acts as an acid and can donate a proton and thus becomes COO^-.
- *At normal pH* of blood, proteins act as acids and combine with cations (mainly sodium).

Haemoglobin buffer system

Haemoglobin buffer system (Hb/HbO_2) accounts for 35% of the total buffering capacity of the whole blood. It mainly buffers the fixed acids, besides being involved in the transport of gases (O_2 and CO_2).

Haemoglobin: intracellular versus extracellular buffer concept. Haemoglobin, though found intracellularly, is more conventionally regarded as a part of extracellular buffer system because:

- Haemoglobin is confined to the erythrocytes, which is a cellular component of ECF,
- Haemoglobin is readily available for the buffering of extracellular acids and
- Haemoglobin is the primary non-carbonate buffer of the body.

Buffering system in haemoglobin is provided by dissociation of an imidazole group of histidine residues:

- *Hb is a major buffer in blood, although between pH 7* and *7.7, it contributes relatively less to buffering capacity.* This is because:
- Haemoglobin molecules are present in large amounts. One litre of whole blood contains about 150 g (2.3 mmol) of haemoglobin.
- Haemoglobin molecule contains 38 histidine residues.

Deoxyhaemoglobin (Hb) is better buffer than oxyhaemoglobin (HbO_2) because the imidazole groups of Hb dissociate less than those of HbO_2, making Hb a weaker acid.

RESPIRATORY MECHANISM FOR pH REGULATION

Second line of defence against acid–base disorders is formed by the respiratory mechanism, which provides a short-term but rapid control. It acts via respiratory centre located in the medulla to regulate removal of CO_2 and therefore carbonic acid (H_2CO_3) concentration in the blood.

Role of respiratory centres

Respiratory centres are influenced by both CO_2 as well as H^+ concentration: through central and peripheral chemoreceptors (see page 342).

Respiratory response occurs in response to metabolic acid–base disorders only and consists of:

1. *Hyperventilation.* It occurs in response to the metabolic acidosis and results in lowering of pCO_2 to match the decreased (HCO_3^-).
2. *Hypoventilation* occurs in response to metabolic alkalosis and results in raising the pCO_2 to match the increased (HCO_3^-).

RENAL MECHANISM FOR pH REGULATION

The kidneys regulate pH through three main processes:

- Reabsorption' of filtered HCO_3^-,
- Generation' of new HCO_3^- and
- H^+ excretion in the form of titrable acid and NH_4^+.

For details see description of acidification of urine (page 408).

ACID–BASE DISORDERS

Acidosis refers to a decline in blood pH, while *alkalosis* refers to a rise in blood pH. As described above, our body has been provided with an efficient system for the maintenance of acid–base equilibrium with a result that the pH of blood is almost constant (7.4). The blood pH compatible to life is 6.8–7.8, beyond which life cannot exist.

Acid–base disorders can be classified into two groups:

I. *The simple acid–base disorders* include:
- Metabolic acidosis,
- Metabolic alkalosis,
- Respiratory acidosis and
- Respiratory alkalosis.

II. *Mixed acid–base disorders* include:
- Metabolic acidosis and respiratory acidosis,
- Metabolic acidosis and respiratory alkalosis,
- Metabolic alkalosis and respiratory alkalosis and
- Metabolic alkalosis and respiratory acidosis.

SIMPLE ACID–BASE DISORDERS

The physiological aspects of single acid–base disorders are summarized in Table 6.5-2 and described:

Metabolic acidosis

Physiological disturbance that produces metabolic acidosis is either increased net non-volatile acid load or loss of base (HCO_3^-).

Table 6.5-2 Summary of characteristics of simple acid–base disorders

Disorder	Primary disturbance	Arterial plasma (Approximate values)			Defence mechanism		
		pH *(Normal 7.4)*	HCO_3^- (mEq/L) *(Normal 24)*	pCO_2 (mm Hg) *(Normal 40)*	Buffering	Respiratory compensation	Renal compensation
Metabolic acidosis	↓ Plasma HCO_3^-	↓ (7.28)	↓ (18)	→ (40)	ECF & ICF	Hyperventilation (↓ pCO_2)	↑ H^+ excretion ↑ New HCO_3^- reabsorption
Metabolic alkalosis	↑ Plasma HCO_3^-	↑ (7.5)	↑ (30)	→ (40)	ECF & ICF	Hypoventilation (↑ pCO_2)	↓ H^+ excretion ↓ New HCO_3^- reabsorption
Respiratory acidosis	↑ pCO_2	↓ (7.34)	↑ (25)	↑ (48)	ICF	None	↑ H^+ excretion ↑ New HCO_3^- reabsorption
Respiratory alkalosis	↓ pCO_2	↑ (7.53)	↓ (22)	↓ (27)	ICF	None	↓ H^+ excretion ↓ New HCO_3^- reabsorption

ICF = Intracellular fluid; ECF = extracellular fluid; ↑ = increased; ↓ = decreased; → = normal.

Primary disturbance in metabolic acidosis is *decreased plasma* [HCO_3^-] producing a *low plasma pH.*

Causes of metabolic acidosis include:

1. *Addition of non-volatile acids* to the body can occur in:
   - Diabetic ketoacidosis causing accumulation of aceto-acetic acid and β-OH-butyric acid.
   - Lactic acidosis in hypoxia.
2. *Loss of non-volatile alkali* from the body occurs in:
   - Diarrhoea (G1 loss of HCO_3^-),
   - Type 2 renal tubular acidosis (renal loss of HCO_3^-).
3. *Failure of the kidney to excrete sufficient net acid to replenish HCO_3^-* used to titrate the net daily acid load, as may occur in:
   - Chronic renal failure (failure to excrete H^+ as titrable acid and NH_4).
   - Type I distal renal tubular acidosis (failure to excrete titrable acid and NH_4).

Uncompensated metabolic acidosis characterized by a low plasma pH and low plasma HCO_3^- is expressed as

$$\downarrow pH = pK + \log \frac{HCO_3^- \downarrow}{pCO_2}$$

Compensatory mechanisms

When metabolic acidosis is produced by the non-renal factors, the respiratory and renal compensatory mechanisms tend to minimize the change in pH of blood. In renal failure, only respiratory compensation is possible.

1. Respiratory compensation. Increased H^+ stimulates the respiratory centre through peripheral chemoreceptors and produces *hyperventilation* (Kussmaul *breathing*), which in turn decreases the arterial pCO_2 value and minimizes the degree of acidosis. Respiratory compensatory mechanism is prompt but short term.

Compensated metabolic acidosis characterized by near normal pH with pCO_2 to compensate the HCO_3^- is expressed as

$$pH = pK + \log \frac{HCO_3^- \downarrow}{pCO_2 \downarrow}$$

2. Renal compensation is slow, but an effective mechanism to control metabolic acidosis. It consists of:

- Increased excretion of fixed H^+ as titrable acid and NH_4.
- Increased reabsorption of new HCO_3^-, which replenishes the HCO_3^- used in buffering the added fixed H^+.
- In chronic metabolic acidosis, an adaptive increase in NH_3 synthesis helps in the excretion of excess H^+.

Hyperchloraemic versus normochloraemic metabolic acidosis

Normochloraemic metabolic acidosis is characterized by decreased plasma (HCO_3^-) and low plasma pH with normal serum Cl^- levels. As described above, the metabolic acidosis caused by conditions adding non-volatile acid to the body or due to renal failure to excrete H^+ is normochloraemic type of metabolic acidosis.

Hyperchloraemic metabolic acidosis is characterized by decreased plasma (HCO_3^-) and low plasma pH with an increase in plasma (Cl^-).

This type of metabolic acidosis occurs in:

- *Diarrhoea*, in which HCO_3^- is lost from the gut in exchange of Cl^- and
- *Type 2 renal tubular acidosis*, in which failure to reabsorb HCO_3^- by kidneys is accompanied by excessive reabsorption of Cl^-.

Anion gap concept

Anion gap helps to differentiate the hyperchloraemic metabolic acidosis from the normochloraemic metabolic acidosis.

According to the law of electroneutrality, the total concentration of cations and anions in serum are equal. Routine serum electrolyte determinations measure essentially all cations but only a fraction of the anions. This apparent disparity between the total cation concentration and the total anion concentration is termed the anion gap (Fig. 6.5-1). It is a virtual measurement and does not represent any specific ionic constituent as shown (Fig. 6.5-1A):

$$[Na^+ + K^+ + Ca^{2+} + Mg^{2+}] = [HCO_3^- + Cl^- + \underset{\text{anions}}{\text{Unmeasures}}]$$
(Anion gap)

$$\therefore \text{Anion gap [AG]} = [Na^+ + K^+ + Ca^{2+} + Mg^{2+}] - [HCO_3^- + Cl^-]$$

Often the anion gap is calculated with Na^+ as the major cation, as K^+, Ca^{2+} and Mg^{2+} has a relatively minor quantitative contribution. The equation then becomes (Fig. 6.5-1B):

$$\begin{aligned} AG &= [Na^+] - [HCO_3^- + Cl^-] \\ &= [142\,mEq/L] - [25\,mEq/L + 105\,mEq/L] \\ &= 142\,mEq/L - 130\,mEq/L \\ &= 12\,mEq/L \end{aligned}$$

The anion gap, which has a normal value of 12 ± 4 mEq/L, reflects the concentration of those anions which are actually present but are not determined routinely (i.e. other than HCO_3^- and Cl^-) and include polyanionic plasma proteins (primarily albumin), inorganic phosphates, sulphate and ions of organic acids.

Anion gap in metabolic acidosis. As mentioned above, in metabolic acidosis, the serum (HCO_3^-) decreases, since it is utilised in buffering the fixed acid. For electroneutrality, the concentration of another anion must increase to replace

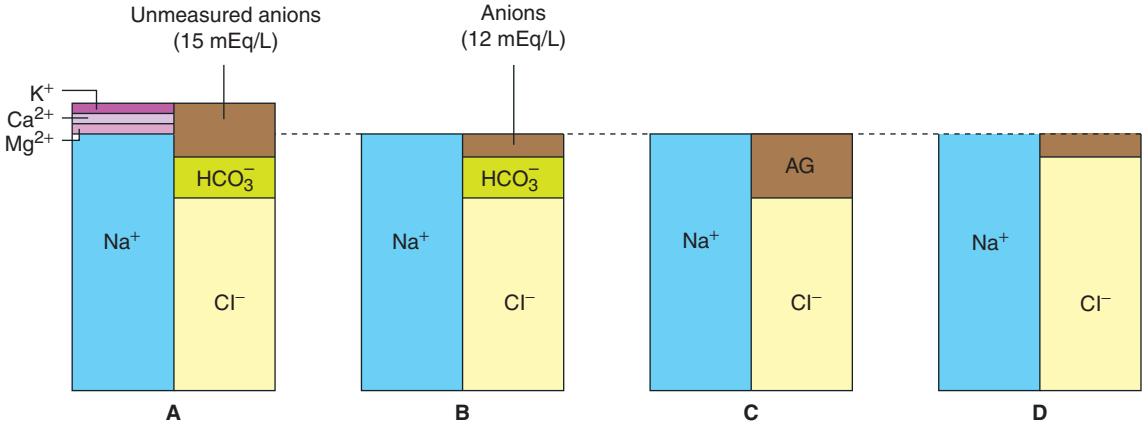


Fig. 6.5-1 Concept of anion gap: A, normal ionogram showing the major cations (Na^+) and the minor cations (K^+, Ca^{2+} and Mg^{2+}). The major anions are HCO_3^- and Cl^- and the unmeasured anion constitute the anion gap; B, simplified ionogram showing Na^+ as the only cation and making the anion column fall as the Na^+ column; C, ionogram depicting increased anion gap in normochloraemic metabolic acidosis and D, ionogram depicting normal anion gap but increased [Cl^-] in hyperchloraemic metabolic acidosis.

HCO_3^-. That anion can be Cl^- or the unmeasured anions (which constitute anion gap).

- *In normochloraemic metabolic acidosis*, the concentration of unmeasured anions is increased to replace HCO_3^- and hence the serum anion gap is increased (Fig. 6.5-1C).
- *In hyperchloraemic metabolic acidosis*, the concentration of Cl^- is increased to replace the HCO_3^-, so the serum anion gap is normal (Fig. 6.5-1D).

Metabolic alkalosis

Physiological disturbance that produces metabolic alkalosis is either addition of non-volatile alkali or loss of H^+ from the body.

Primary disturbance in metabolic alkalosis is *increased plasma* HCO_3^- producing a high plasma pH.

Causes of metabolic alkalosis include:

1. *Addition of non-volatile alkali* to the body, e.g. ingestion of antacids.
2. *Volume contraction alkalosis* may occur with haemorrhage and thiazide diuretics.
3. *Loss of H^+* from the body (a common cause) may occur in:
   - Vomiting (H^+ is lost from the stomach) and
   - Hyperaldosteronism (increased H^+ secretion by distal tubule).

Uncompensated metabolic alkalosis characterized by a high plasma pH and high plasma HCO_3^- is expressed as

$$\uparrow \text{pH} = \text{pK} + \log \frac{\text{HCO}_3^- \uparrow}{\text{pCO}_2}$$

Compensatory mechanisms

1. Respiratory compensation. Increased pH (or decreased H^+) inhibits the respiratory centre through peripheral chemoreceptors and produces *hypoventilation,* which in turn elevates pCO_2 and thus normalizes plasma pH.

Compensated metabolic alkalosis characterized by near normal pH with CO_2 to compensate the HCO_3^- is expressed as

$$\text{pH} = \text{pK} + \log \frac{\text{HCO}_3^- \uparrow}{\text{pCO}_2 \uparrow}$$

Note. It is important to note that the magnitude of respiratory compensation is limited by the fact that hypoventilation results in decreased arterial pO_2, which stimulates the respiratory centres via peripheral chemoreceptors.

2. Renal compensation for metabolic alkalosis consists of:

- Decreased H^+ secretion by the renal tubules and
- Increased HCO_3^- excretion as the filtered load of HCO_3^- exceeds the ability of renal tubule to reabsorb it. The urinary loss of HCO_3^- decreases the plasma level of HCO_3^-, thereby restoring the pH of blood to near normal.

Respiratory acidosis

Primary disturbance in respiratory acidosis is *increased pCO_2*, which by mass action causes an increase in H^+ and thus lowers the blood pH.

Causes. The pCO_2 is increased due to decreased gas exchange across the alveoli because of following causes:

- Drug-induced (opiates, sedatives, anaesthetics) depression of respiratory centres,
- Weakening of respiratory muscles as in the Guillain–Barre syndrome, polio, amyotrophic lateral sclerosis and multiple sclerosis,
- Airway obstruction,

- Impaired gas diffusion, as may occurs in cardiovascular diseases or lung disease (e.g. adult respiratory distress syndrome, chronic obstructive pulmonary disease).

Uncompensated respiratory acidosis, characterized by low plasma pH and high pCO_2 is expressed as

$$\downarrow pH = pK + \log \frac{HCO_3^-}{\uparrow pCO_2}$$

Compensatory mechanisms

Note. There is no respiratory compensation for respiratory acidosis.

Buffering in respiratory acidosis, in contrast to metabolic acidosis, occurs almost entirely in the intracellular compartment.

Renal compensation. Increased pCO_2 supplies more H^+ to renal tubule cells for secretion which leads to:

- Increased excretion of H^+ as a titrable acid and NH_3
- Increased reabsorption of new HCO_3^-

The resulting increase in serum HCO_3^- helps to normalize the pH. Thus, acidosis is mostly but not completely compensated by the renal mechanism.

Compensated respiratory acidosis characterized by near normal pH with increased plasma HCO_3^- to compensate the increased pCO_2 is expressed as

$$pH = pK + \log \frac{HCO_3^- \uparrow}{pCO_2 \uparrow}$$

Respiratory alkalosis

Primary disturbance in the respiratory alkalosis is decreased pCO_2 associated with low (H^+) and thus an elevated plasma pH.

Causes. The pCO_2 is decreased due to increased gas exchange in the lungs because of increased ventilation as seen in following conditions:

- Pneumonia and pulmonary embolus (ventilation rate is increased secondary to hypoxaemia),
- High altitude (ventilation rate is increased secondary to hypoxaemia),
- Psychogenic hyperventilation may occur as a response to anxiety or fear, and
- Salicylate intoxication (hyperventilation occurs due to direct stimulation of medullary respiratory centres).

Uncompensated respiratory alkalosis, characterized by high plasma pH and low pCO_2, is expressed as

$$\uparrow pH = pK + \log \frac{HCO_3^-}{\downarrow pCO_2}$$

Compensatory mechanisms

Note. There is no respiratory compensation for respiratory alkalosis.

Buffering in respiratory alkalosis, in contrast to metabolic alkalosis, occurs almost entirely in the intracellular compartment.

Renal compensation. Decreased pCO_2 causes a deficit of H^+ in the renal cells for secretion which leads to:

- Decreased excretion of H^+ as titrable acid and NH_4^+,
- Decreased reabsorption of new HCO_3^- and
- Decreased reabsorption of the filtered HCO_3^-.

The resulting decrease in serum (HCO_3^-) helps to normalize the pH. In this way, the alkalosis is mostly but not completely compensated by the renal mechanism.

Compensated respiratory alkalosis, characterized by near normal pH with decreased plasma HCO_3^- to compensate the decreased pCO_2 is expressed as

$$pH = pK + \log \frac{HCO_3^- \downarrow}{pCO_2 \downarrow}$$

Summary of characteristics of simple acid–base disorders. Various characteristics of simple acid–base disorders are summarized in Table 6.5-2.

ANALYSIS AND CLINICAL EVALUATION OF ACID–BASE DISORDERS

THREE-STEP APPROACH FOR ANALYSIS OF ACID–BASE DISORDERS

Three-step approach for analysis of acid–base disorders is summarized in Fig. 6.5-2. It consists of following three steps:

Step I: Estimate pH to know acidosis (pH<7.4) or alkalosis (pH>7.4).

Step II: Detect primary disturbance to know whether the disorder is metabolic (primary disturbance of HCO_3^-) or respiratory (primary disturbance of pCO_2).

Step III: Analysis of compensatory response can be done from the values of plasma HCO_3^- and pCO_2.

GRAPHIC ANALYSIS OF CHANGES IN pH, pCO_2 AND HCO_3^-

Acid–base nomogram

Acid–base nomogram (Fig. 6.5-3) is the graphical display of changes in pCO_2 (curved lines in Fig. 6.5-3), plasma HCO_3^- and pH of arterial blood in the respiratory and metabolic acid–base disorders. This nomogram is useful in predicting compensatory responses to simple acid–base disorder. While the shaded areas of nomogram show the 95%

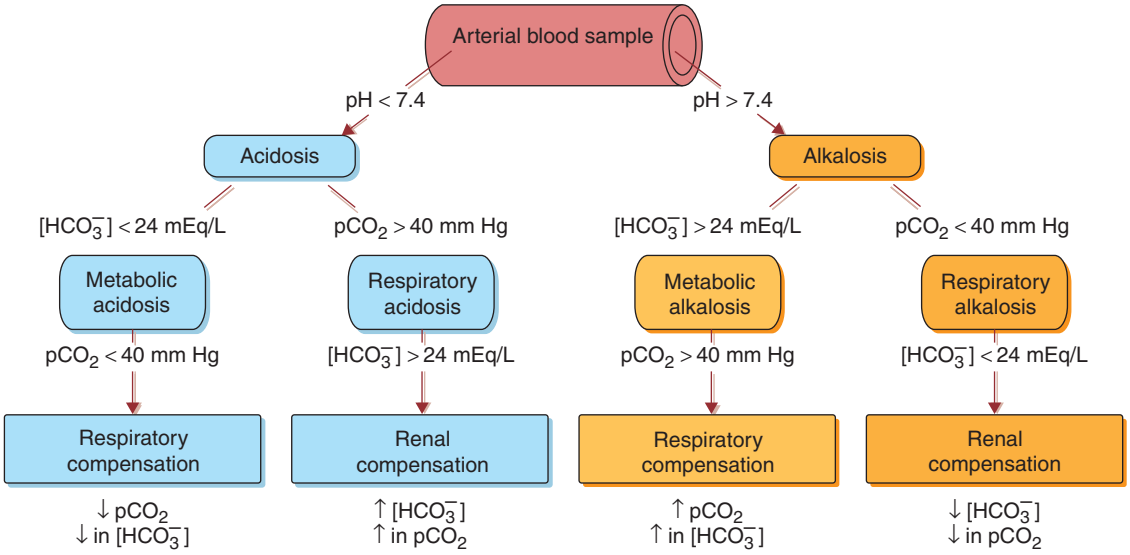


Fig. 6.5-2 Algorithm of three-step approach for analysis of acid-base disorder.

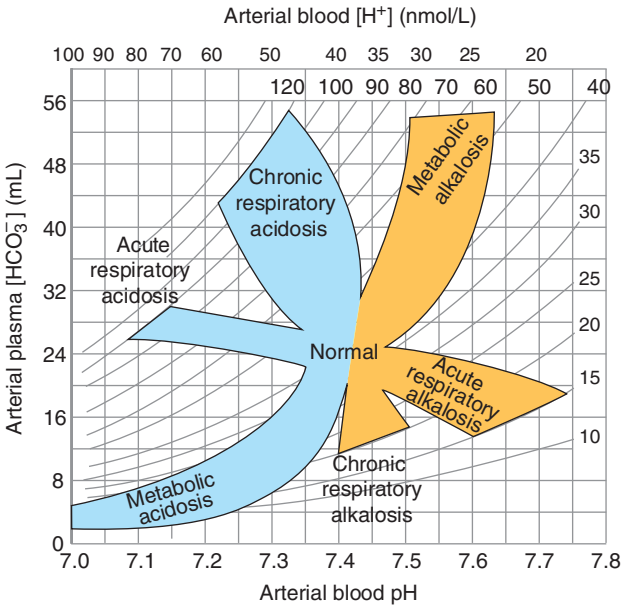


Fig. 6.5-3 Acid–base nomogram (for explanation see text).

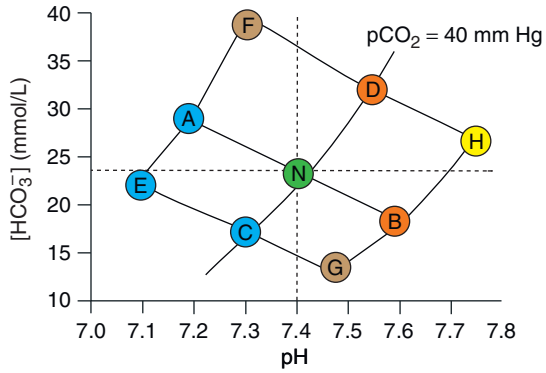


Fig. 6.5-4 Interpretation of acid–base abnormalities using the pH–[HCO_3^-] diagram (Davenport diagram). For explanation see text.

confidence limits for normal compensation in simple disturbances, finding acid–base values within the shaded and do not necessarily rule out a mixed disturbance. Note that the shifts in HCO_3^- and pH as acute respiratory acidosis and alkalosis are compensated producing their chronic counterparts.

Davenport diagram: graphic display of true plasma pH, HCO_3^- and pCO_2 in metabolic acidosis and alkalosis

Figure 6.5-4 is the typical graphical display of true plasma pH, HCO_3^- and pCO_2 in uncompensated and compensated metabolic acidosis and metabolic alkalosis. It shows the relationship between pH and HCO_3^- at a constant pCO_2 and hence also called *pCO_2* isobar. Thus, acid–base imbalances are determined graphically, with reference to the intercept of pCO_2 isobar of 40 mm Hg (line CND,) and the normal buffer line (Line ANB, Fig. 6.5-4). The intercept of these two curves marks the point of normality (N) which is associated with a pH of 7.4 (the abscissa) and a [HCO_3^-] of 24 mmol/L (the ordinate). Thus, the point N is the triple intercept that defines the pH, [HCO_3^-] and pCO_2 of true arterial plasma of a normal individual.

Interpretation of acid–base abnormalities using pH, HCO_3^- diagram is made as:

- *Point A*, represents uncompensated respiratory acidosis,
- *Point B*, represents uncompensated respiratory alkalosis,
- *Point C*, represents uncompensated metabolic acidosis,
- *Point D*, represents uncompensated metabolic alkalosis,
- *Point E*, represents respiratory acidosis + metabolic acidosis,

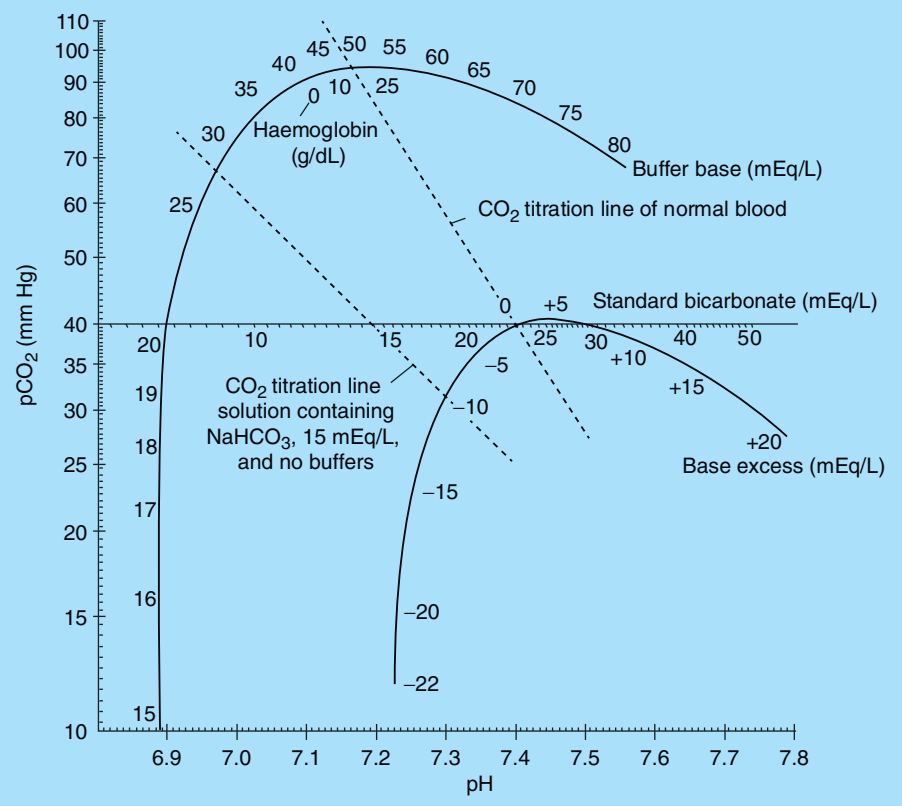


Fig. 6.5-5 Siggard–Anderson curve nomogram to plot the acid–base characteristics of arterial blood.

- *Point F*, represents respiratory acidosis + metabolic alkalosis,
- *Point G*, represents respiratory alkalosis + metabolic acidosis and
- *Point H*, represents respiratory alkalosis + metabolic alkalosis.

Siggard–Anderson curve nomogram

Siggard–Anderson (SA) curve nomogram (Fig. 6.5-5) has pCO_2 plotted on a log scales on the vertical axis and pH on the horizontal. This nomogram is helpful in the clinical situation to plot the acid–base, a characteristic of arterial blood.

Protocol for using SA nomogram

- Arterial capillary blood is drawn anaerobically and pH is measured. pH of the same blood after equilibration with each of two gas mixtures containing known amount of CO_2 are determined.
- pH value at known pCO_2 levels are plotted and connected to provide CO_2 titration line from the blood sample.
- pH of the blood sample before equilibration is plotted on this line and actual pCO_2 of sample read off the vertical line.

Following values can be determined:

- Standard HCO_3^- content of sample, i.e. measure of the alkali reserve of the blood.
- Buffer base (normal value 48 mEq/L).
- Base excess is represented by the point at which CO_2 calibration line intersects the lower curved scale on nomogram. Base excess is positive in alkalosis and negative in acidosis.

Chapter 6.6

Applied Renal Physiology Including Renal Function Tests

PATHOPHYSIOLOGY OF COMMON RENAL DISORDERS
- Common urinary symptoms
- Renal failure
- Nephrotic syndrome

DIURETICS
- Classification
- Site of action, mechanism of action and major effects

RENAL FUNCTION TESTS
- **Analysis of urine**
- **Analysis of blood**
- **Renal clearance tests**
- **Radiology and renal imaging**
- **Renal biopsy**

DIALYSIS AND RENAL TRANSPLANTATION
- Dialysis
- Haemodialysis
- Peritoneal dialysis
- Renal transplantation

PATHOPHYSIOLOGY OF COMMON RENAL DISORDERS

The applied aspects of common renal disorders which need some elaboration are:

- Common urinary symptoms,
- Renal failure and
- Nephrotic syndrome

COMMON URINARY SYMPTOMS

Polyuria, nocturia and urinary frequency. Normal urine output per day is 800–2500 mL. Therefore, a reasonable criterion to satisfy the definition of polyuria is excretion of 3.0 L of urine daily, provided the patient is not on high fluid diet.

Nocturia means excessive amount of urine passed at night.

Urinary frequency means the increase in the number of times the patient goes for urination. Polyuria is differentiated from increased frequency by measuring the 24 h urine output.

Common causes of polyuria are:

I. *Physiological* (primary polydipsia or excessive water drinking), which can be
- Psychogenic, or
- Drug induced (chlorpromazine, anticholinergics).

II. *Pathological* (defective water conservation by the kidney):
- Diabetes insipidus.
- Solute diuresis, as in chronic renal failure, diabetes mellitus and mannitol infusion.

Dysuria and urgency of micturition. *Dysuria* refers to pain or burning during micturition. *Urgency of micturition* is the exaggerated sense or urge to micturate. It is due to either irritative or inflammatory disorders of the urinary bladder. This is often associated with an increased frequency of urination.

Incontinence. This refers to inability to retain urine in the bladder. It results from the neurological or mechanical disorders of the complicated system that controls normal micturition.

Common causes of incontinence are:

- Neurogenic incontinence due to disturbances of neural control of micturition,
- Stress incontinence, e.g. in post-menopausal parous women,
- Mechanical incontinence, e.g. damage to the urethral sphincters.
- Overflow incontinence, e.g. in obstruction due to benign prostatic enlargement,
- Psychogenic incontinence, as in anxious children and
- Functional incontinence is seen in very old persons who have mental derangement.

Enuresis refers to the involuntary passage of urine at night or during sleep. It is also called night bed-wetting or nocturnal enuresis. It is normal in children up to 2–3 years of age. In some children it continues for long.

Oliguria refers to the urine output less than 500 mL/day in an average adult. It invariably occurs in acute on chronic renal failure or acute renal failure.

Anuria is said to occur when patient does not pass any urine or passes less than 50 mL of urine/day. In physiological sense, the term anuria means less formation or absence of formation of urine by the kidney.

RENAL FAILURE

Renal failure refers to the deterioration of renal functions resulting in a decline in the glomerular filtration rate (GFR) and rise in urea and non-nitrogenous substances in the blood.

It is of two types:

- Acute renal failure and
- Chronic renal failure.

Acute renal failure

Acute renal failure refers to a sudden decline in GFR over a period of days or weeks associated with the rapid rise in blood urea.

Chronic renal failure

Chronic renal failure refers to a slow, insidious, irreversible deterioration of renal functions resulting in the development of clinical syndrome of uraemia, manifested by excretory, metabolic, neurological, haematological and endocrinal abnormalities.

Acute versus chronic renal failure. Differentiating features of acute and chronic renal failure are summarized in Table 6.6-1.

NEPHROTIC SYNDROME

Nephrotic syndrome refers to a massive proteinuria (more than 3.5 g/day), mainly albuminuria and its associated consequences which include:

- Hypoalbuminaemia,
- Oedema,
- Hyperlipidaemia,
- Lipiduria and
- Hypercoagulability.

Pathophysiology. A wide variety of disease processes including immunological disorders, toxic injuries, metabolic abnormalities, biochemical defects and vascular disorders involving glomeruli contribute to the development of nephrotic syndrome. The sequence of events involved in pathophysiology of nephrotic syndrome is summarized in Fig. 6.6-1.

DIURETICS

The diuretics are the drugs which primarily cause a net loss of Na^+ *(natriuresis)* associated with water loss (secondary to natriuresis) and thus increase the rate of urine flow.

Classification

Depending upon their efficacy, the diuretic drugs can be classified as:

1. *High-efficacy diuretics* (inhibitors of Na^+–K^+–$2Cl^-$ transport), also called loop diuretics, e.g. Furosemide.

Table 6.6-1 Distinguishing features of acute and chronic renal failure

Feature	Acute renal failure	Chronic renal failure
Onset	Sudden over days or to weeks	Gradual, over months or years
Reversibility	Invariably reversible	Usually irreversible
Causes	May be pre-renal or post-renal	Mostly renal may be extra renal
Urinary volume	Oliguria and anuria	Polyuria and nocturia
Signs and symptoms of uraemia	Of recent onset	Of more than 3 months duration
Characteristic features	○ Sudden reduction in GFR ○ Rapid rise in blood pressure, urea and creatinine level ○ High urine osmolality (> 400 mOsm/kg water)	Of chronicity, i.e. uraemic symptoms of long duration e.g. water retention, small-sized kidneys, anaemia, hypertension and so on
Renal failure casts (broad casts) in urine	Absent	Present
Specific gravity of urine	High	Low and fixed
Past history of renal disease	Absent	Present
Dialysis	Required for short period	Repeated chronic maintenance dialysis required
Renal transplantation	Usually not required	Usually, is the final answer

2. *Medium efficacy diuretics* (inhibitors of Na^+–Cl^- symport)
   (i) Thiazide diuretics, e.g. chlorothiazide,
3. *Weak or adjunctive diuretics*
   (i) Carbonic anhydrase inhibitors, e.g. acetazolamide
   (ii) Potassium sparing diuretics, e.g. spironolactone, triamterene and amiloride
   (iii) Osmotic diuretics, e.g. mannitol, isosorbide and glycerol
   (iv) Xanthines, e.g. theophylline.

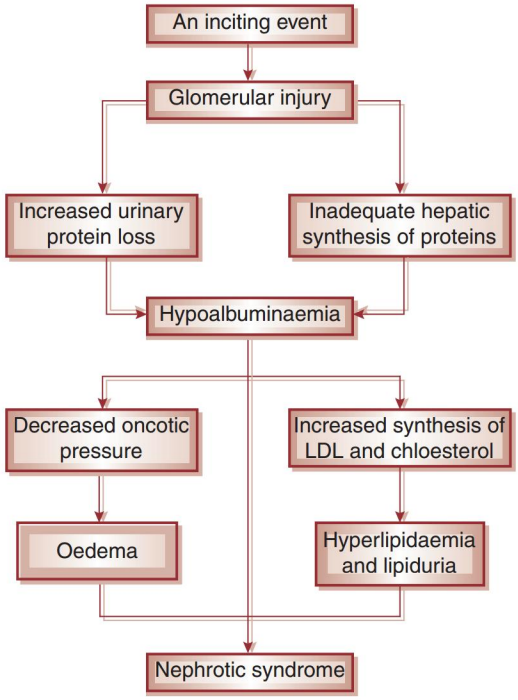


Fig. 6.6-1 Algorithm of pathophysiology of nephrotic syndrome.

Site of action, mechanism of action and major effects

Site of action. The different sites of a nephron, where the diuretics act are summarized in Table 6.6-2.

Mechanism of action and major effects of the diuretics are summarized in Table 6.6-2.

Other substances which act as diuretics

In addition to the above described drugs, following other substances also act as diuretics e.g.

- *Water.* It acts by inhibiting antidiuretic hormone (ADH) secretion
- *Alcohol.* It also acts by inhibiting ADH secretion
- *Antagonist of V_2 vasopressin inhibitors* inhibit action of vasopressin on the collecting ducts
- *Glucose* (e.g. in diabetes mellitus). It acts as an osmotic diuretic
- *Caffeine.* It increases GFR and probably decreases renal tubular Na^+ reabsorption.

RENAL FUNCTION TESTS

Renal function tests are carried out to assess the functional capacity of the kidneys. The main aims of these tests in clinical

Table 6.6-2 Site of action, mechanism of action and major effects of various classes of diuretics

Class of diuretic	Site of action	Mechanism of action	Major effects
High efficacy diuretics Loop diuretics, e.g. – Furosemide – Bumetanide – Ethacrynic acid	TAL	Inhibition of Na^+–K^+–$2Cl^-$	↑NaCl excretion ↑K^+ excretion ↑Ca^{2+} excretion ↓Ability to concentrate urine ↓Ability to dilute urine
Medium efficacy diuretics Thiazide diuretics, e.g. – Chlorothiazide – Hydrochlorothiazide	Early distal tubule	Inhibition of Na^+–Cl^- symport	↑NaCl excretion ↑K^+ excretion ↓Ca^{2+} excretion ↓Ability to dilute urine
Weak or adjunctive diuretics			
(i) Carbonic anhydrase inhibitor, e.g. – Acetazolamide	Proximal tubule	Inhibition of carbonic anhydrase	↑HCO_3^- excretion
(ii) Potassium sparing. diuretics, e.g. – Spironolactone – Triamterene – Amiloride	Late distal tubule and collecting duct	Inhibition of Na^+ reabsorption, inhibition of K^+ secretion	↑Na^+ excretion ↓K^+ excretion ↓H^+ excretion
(iii) Osmotic diuretics, e.g. – Mannitol – Isosorbide	Proximal tubule	Retain water iso-osmotically and inhibits NaCl reabsorption	↑Na^+ excretion ↑K^+ excretion ↑Cl^- excretion

medicine are the detection of renal impairment as early as possible in its course and the quantitative measure of change in function with time. However, it must be remembered that about two-thirds of renal tissue must be functionally damaged to show any abnormality by these tests. Renal function tests can be divided into following groups:

- Analysis of urine
- Analysis of blood
- Renal clearance tests
- Radiology and renal imaging
- Renal biopsy

ANALYSIS OF URINE

Analysis of urine helps, of course, to a limited degree, to assess kidney functioning. In patients with suspected renal disorder, the urine analysis should be performed for volume, specific gravity, osmolality, pH, abnormal constituents, microscopic examination and bacteriological finding.

1. Volume. Normal urine output per day is 800–2500 mL. Abnormalities of urine volume include: polyuria, oliguria and anuria (see page 432).

2. Colour. The normal light yellow colour of the urine is due to the presence of urochrome pigment (a compound of urobilin and urobilinogen with peptide). On keeping the urine in test tube for some time, the colour deepens due to oxidation of urobilinogen into urobilin. Abnormalities of urine colour include:

- *Brownish yellow,* due to the presence of conjugated bilirubin in patients with hepatic and post-hepatic jaundice.
- *Cloudy appearance* is seen in strongly alkaline urine due to precipitation of calcium phosphate and due to precipitation of urates.
- *Frothy appearance* is an indicative of proteinuria.
- *Red-dark brown* tinge of urine is seen in porphyria.

3. Osmolality and specific gravity. Normal urinary osmolality varies from 50 to 1200 mOsm/kg and specific gravity from 1.003 to 1.030, depending upon the state of hydration of the body. If the early morning urine sample after an overnight fast has an osmolality of <600 mOsm/kg H_2O (and specific gravity >1.018), then the patient has a normal urine concentrating ability. Certain abnormalities are:

- Fixed urinary osmolality of 300 mOsm/kg H_2O (specific gravity 1.010) is an evidence of fairly advanced urinary failure.
- Persistently low urinary osmolality (less than 100 mOsm/kg H_2O) even after 8 h of fluid deprivation is diagnostic of diabetes insipidus.

4. Urine pH. Normal pH of urine varies from 4.5 to 8.0. Urine is normally slightly acidic, except for a short post-prandial alkaline tide. Intake of a high protein non-vegetarian diet shifts the urinary pH towards acidic side, while vegetarian diet shifts it towards alkaline side.

5. Chemical analysis for abnormal urinary constituents may reveal:

(i) *Proteinuria.* Normally, up to 150 mg of proteins are excreted daily in urine. Excretion of >150 mg/day of protein is called proteinuria. It occurs in following conditions:
- In congestive heart failure
- After prolonged standing (orthostatic proteinuria)
- Renal diseases–*Glomerular proteinuria* occurs in diseases in which permeability of glomerular membrane is increased. Massive glomerular proteinuria is seen in the nephrotic syndrome. Other causes of glomerular proteinuria are acute glomerulonephritis and pyelonephritis.

(ii) *Glycosuria* refers to the presence of glucose in the urine. Glycosuria may be due to diabetes mellitus, renal disorders (renal glycosuria), GIT disorder (alimentary glycosuria). Other sugars like galactose and fructose, may also be present in urine in certain inborn errors of metabolism.

(iii) *Ketonuria* refers to the presence of ketone bodies (acetoacetic acid, β hydroxybutyric acid and acetone) in the urine. Ketonuria occurs in the patients suffering from ketosis due to severe diabetes mellitus or prolonged starvation.

(iv) *Bilirubinuria* refers to appearance of bilirubin in the urine of patients with elevated conjugated bilirubin levels, in hepatic or post-hepatic jaundice. Normally, 1–3.5 mg of urobilinogen is excreted daily in the urine. Its excessive excretion in the urine is one of the characteristic features of haemolytic jaundice.

(v) *Haemoglobinuria,* i.e. presence of haemoglobin in the urine indicates intravascular haemolysis, as seen in black-water fever due to falciparum malarial infection.

(vi) *Haematuria,* i.e. presence of blood in the urine is seen in acute glomerulonephritis and renal stone disease.

6. Microscopic examination. Examination of centrifuged sediment of urine may show casts, cells and crystals.

(i) *Casts* are proteinaceous plugs formed by the coagulation of Tamm–Horsfall protein within the renal tubules and washed out by the flow of tubular fluid. They have cylindrical shape, broken ends and various shapes corresponding to the tubule in which they formed. Casts may be cellular or non-cellular.

In cellular casts, certain cells are coagulated with the protein material. *Non-cellular casts* are hyaline and granular casts.

(ii) *Crystals* are usually present in normal urine and thus have no pathological significance. Commonly seen are

crystals of calcium oxalate, calcium phosphate, calcium ammonium–magnesium phosphate (triple phosphate) or uric acid. Uric acid crystals and cysteine crystals, when present in large amounts have some diagnostic significance.

(iii) *Cells* found on the microscopic examination may be RBCs, leucocytes, tubular epithelial cells, and squamous epithelial cells.

7. Bacteriological examination of urine. The midstream sample of urine is examined for pus cells and bacteria. It normally contains 1–2 WBCs or pus cells/HPF. Bacteriuria and pyuria indicate urinary tract infection.

ANALYSIS OF BLOOD

Estimation of blood levels of the substances that are excreted by the kidneys throw some light on the functional status of kidney, although these tests are less sensitive than the clearance tests.

1. Blood urea level (normal 20–40 mg/dL) is an index of glomerular function. The blood urea levels begin to rise after about 50% glomerular damage has occurred.

2. Plasma creatinine concentration (normal 0.6–1.5 mg/dL) is more reliable than blood urea, as the later is subjected to variations by dietary proteins, hydration and tissue breakdown.

3. Serum proteins levels. (Normal: total protein 6.7–8 g/dL; albumin, 3–5 g/dL; globulins, 2–3 g/dL and A/G ratio, 1.7:1) are reduced if there is significant proteinuria with renal failure. In nephrotic syndrome the albumin levels decrease and globulin levels increase, leading to reversal of A/G ratio.

4. Serum cholesterol levels (normal 150–200 mg/dL) are increased in the nephrotic syndrome.

5. Serum electrolyte levels (normal: Na^+, 152 mEq/L; K^+, 5 mEq/L; Ca^{2+}, 9–11 mg/dL; PO_4^{3-}, 3–4.5 mg/dL; SO_4^{2+}, 0.5–1.5 mEq/L and Mg^{2+}, 1.5–2.5 mEq/L) are of value in variety of renal disorders. For example, chronic renal failure is mostly accompanied by high potassium and phosphate but low sodium and calcium levels in blood.

RENAL CLEARANCE TESTS

The renal clearance can be defined as the volume of plasma that is cleared of a substance in 1 min by excretion of the substance in the urine. It is a 'virtual volume'. The unit of renal clearance (C) is mL/min and is calculated from the following formula:

$$C = \frac{UV}{P},$$

where

C = Renal clearance,
U = Urine concentration of the substance,
V = Rate of flow of urine and
P = Plasma concentration of the substance.

Principles governing renal clearance of a substance, described on page 390 are summarized briefly:

- Substances that are freely filtered, but neither reabsorbed nor secreted (e.g. *inulin)* have renal clearance rate equal to GFR and hence are called glomerular markers.
- Substances that are freely filtered, but are partially reabsorbed in the tubules have renal clearance rate less than GFR.
- Substances that are freely filtered, but are completely reabsorbed (e.g. Na^+, glucose, amino acids, Cl^- and HCO_3^-) have lowest renal clearance rate.
- Substances that are filtered and also secreted by the tubules but not reabsorbed (e.g. *PAH* and Diodrast) have the highest renal clearance rate. Such substances are thus entirely excreted by a single passage of blood through kidneys. Clearance of such substances represents the range of blood flow.

Renal clearance as kidney function test. Renal clearance of a substance is correlated more directly with the status of kidney function. It shows a deviation from normal (earlier) in the course of renal damage.

Renal clearance tests, therefore can be employed to assess the different functions of a nephron, e.g.

- To assess glomerular filtration,
- To assess tubular secretory capacity,
- To assess renal plasma flow (RPF) and renal blood flow (RBF).
- To assess osmotic and free water clearance.

RENAL CLEARANCE TESTS TO MEASURE GFR

Glomerular filtration rate can be accurately measured by the renal clearance of inulin, urea and creatinine.

1. Inulin clearance test

Inulin is a dye (chemically a fructo-polysaccharide) that does not exist naturally in the body. Inulin clearance (C_{in}) is a measure of GFR because the volume of plasma completely cleared of inulin per unit time equals the volume of plasma filtered per unit time. Inulin clearance gives the measure of GFR because of its following characteristics:

- It is freely filtered by the glomeruli and neither reabsorbed nor secreted by the tubules,
- It is biologically inert and non-toxic,
- It is neither metabolized nor stored in the kidney,
- Its concentration can be easily estimated in the laboratory.

Method. To perform this test, a single bolus dose of inulin injected intravenously which is followed by a continuous intravenous infusion at a rate which compensates for its loss in urine. This endurance is important to achieve a fairly constant level of the plasma concentration of inulin. The inulin clearance (C_{in}) and thus, the GFR is then calculated from the values of:

- Urine concentration of inulin (U_{in}) = 35 mg/mL,
- Urine flow rate (V) = 0.9 mL/min and
- Plasma concentration of inulin (P_{in}) = 0.25 mg/mL as:

$$C_{inulin}(\text{or GFR}) = \frac{U_{in}V}{P_{in}} = 35 \times \frac{0.9}{0.25} = 126\,\text{mL/min}$$

Clinical applications (significance) of inulin clearance. In addition to its use as a measure of GFR, the inulin clearance rate is also used as an indicator of plasma clearance mechanisms. A comparison of the clearance of a given substance (C_X) with the clearance of inulin (C_{in}) provides information about the renal transport processes used to remove the substance from the plasma (Fig. 6.6-2).

2. Creatinine clearance test

Though creatinine clearance test is less accurate than the inulin clearance test for measurement of GFR, but in clinical practice the former is preferred over the later because the later is more cumbersome as it requires a continuous intravenous infusion. Creatinine is an endogenous substance having a fairly constant plasma value (P) of about 0.6–1.5 mg/dL. It is filtered by the glomeruli and only marginally secreted by the tubules. The value of creatinine clearance is close to GFR, hence its measurement is a fairly good method of measuring GFR.

Method. In the traditional method, creatinine content of 24 h urine collection and the plasma concentration in a sample collected at midpoint of the urinary collection period are estimated. The creatinine clearance (C) is then calculated by the usual formula:

$$C = \frac{UV}{P}$$

Normal value of creatinine clearance ranges from 80 to 110 mL/min in an adult and declines with age in healthy individuals. Because creatinine clearance is an index of GFR, it reflects the normal decline in GFR with age, although plasma creatinine concentration remains constant because of decreased muscle mass.

Creatinine clearance as kidney function test in disease. The plasma creatinine concentration varies inversely with GFR, and the product of GFR and plasma creatinine concentration is constant. Thus, a fall in GFR may be the earliest clinical sign of renal disease (i.e. a decline in functional renal mass).

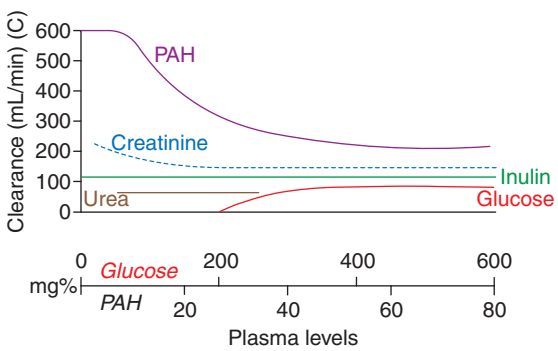


Fig. 6.6-2 Clearance of various substances plotted against their plasma concentration.

3. Urea clearance test

Urea is the end product of protein metabolism. After being filtered by the glomeruli, it is partly reabsorbed by the renal tubules. Hence, urea clearance is less than the GFR and further it is influenced by the protein content of the diet. For these reasons, urea clearance is not as sensitive as the creatinine clearance, for assessing renal function.

RENAL CLEARANCE TESTS TO ASSESS TUBULAR SECRETORY CAPACITY

The tubular secretory capacity can be assessed by the renal clearance of substances that are actively secreted by the tubular cells. As described on page 400, secretion of PAH (para-aminohippuric acid), occurs into the tubular fluid via carriers in the proximal tubule by a transport maximum (T_m) limited process. The T_m PAH is about 80 mg/min, therefore,

- When P_{PAH} is low, PAH is almost completely cleared from plasma by a combined process of glomerular filtration and tubular secretion.
- When P_{PAH} is above 20 mg/dL, the transepithelial secretory process is saturated and T_m PAH is reduced. Then quantity of PAH secreted remains constant and is independent of P_{PAH}.
- Once T_m of PAH is reached, the clearance of PAH (C_{PAH}) becomes progressively more a function of glomerular filtration, hence the C_{PAH} approaches C_{in}, and the constant amount of PAH secreted becomes a smaller fraction of the total amount excreted. Because the T_m (PAH) is nearly constant, it is used clinically to estimate tubular secretory capacity (T_s).

RENAL CLEARANCE TEST TO ASSESS RENAL PLASMA FLOW

The renal plasma flow can be calculated by applying the *Fick's principle* to the kidneys. According to this principle,

the amount of a substance excreted by the kidney per unit time (UV) is equal to the RPF multiplied by the arteriovenous difference in its plasma concentration:

$$UV = RPF\ (P_a - P_v)$$

$$\text{or } RPF = \frac{UV}{P_a - P_v},$$

where

P_a = Concentration of the substance in renal arterial plasma (mg/mL)
P_v = Concentration of the substance in renal venous plasma (mg/mL)
U = Concentration of the substance in urine (mg/mL)
V = Volume of urine excreted (mL/min).

PAH clearance is used to measure RPF

Method. PAH is continuously infused at low doses, so as to keep its plasma concentration constant. The RPF is calculated as

$$\frac{U_{PAH}V}{P_{a(PAH)} - P_{v(PAH)}}$$

- At low plasma concentration of PAH, all the PAH is excreted into the urine and none is returned to the circulation via renal vein. As a result, the PAH concentration in renal vein is zero and can be eliminated. The equation now becomes

$$RPF = \frac{U_{PAH}V}{P_{a(PAH)}}$$

- Since $\frac{U_{PAH}V}{P_{PAH}}$ is equal to clearance of PAH (C_{PAH}), so the equation can be written as

$$RPF = C_{PAH}$$

- About 10% of total renal plasma flow perfuses the non-excretory portions of the kidney. Therefore, RPF calculated from the clearance of PAH is referred to as the effective RPF (ERPF) because only 90% of the plasma PAH is extracted. Therefore, the equation can be written as

$$ERPF = C_{PAH}$$

- Normally, concentration of PAH in urine (U_{PAH}): 14 mg/mL
  - Urine flow (V): 0.9 mL/min
  - Concentration of PAH in plasma (P_{PAH}): 0.02 mg/mL
  - Therefore, ERPF = 14 × 0.9/0.02 = 630 mL/min
- To obtain true RPF, it is necessary to divide C_{PAH} by 0.9

$$\text{True RPF} = \frac{C_{PAH}}{0.9}$$

therefore actual RPF = 630/0.9 = 700 mL/min

- From the value of true RPF, value of RBF can be easily determined, if haematocrit value (Hct) is known as

$$RBF = RPF \times \frac{1}{1 - Hct}$$

Normally, Hct is 45%, therefore,

$$RBF = 700 \times 1/1\text{-}0.45$$
$$= 700 \times 1/0.55 = 1273\ \text{mL/min}$$

- Normal values of ERPF are 650 mL/min/1.73 m^2 body surface area (BSA) in males and 600 mL/min/1.73 m^2 BSA in females. Accordingly, renal blood flow is approximately 1200 mL/min.

RENAL CLEARANCE TEST TO ASSESS 'OSMOTIC' AND 'FREE WATER' CLEARANCE

1. Osmotic clearance (C_{osm})

Osmotic clearance (C_{osm}) is the amount of plasma (in mL) completely cleared of osmotically active solutes that appear in the urine each minute. It measures the rate at which plasma is cleared of osmotic particles and is calculated by the usual renal clearance formula of $C = \frac{U \times V}{P}$, which can be written as

$$C_{osm} = \frac{U_{osm}V}{P_{osm}},$$

where

U_{osm} = urinary osmolality,
V = rate of urine flow mL/min and
P_{osm} = Plasma osmolality.

Normal value of C_{osm} is about 3 mL/min. It is increased in osmotic diuresis and decreased in fasting or diet deficient in proteins.

2. Free water clearance (C_{H_2O})

The quantitative measure of the kidney's ability to excrete water is termed *free water clearance* (C_{H_2O})'. Free water clearance (C_{H_2O}) denotes the volume of pure (i.e. solute free) water that must be removed from, or added to, the flow of urine (in mL/min) to make it iso-osmotic with plasma. In other words, it is a measure of the ability of the kidneys to generate solute-free water. It is not a true clearance because no osmotically free water exists in plasma.

Free water or solute free water is generated in the diluting segments of the kidney (i.e. thick ascending limb and early distal tubule), where NaCl is reabsorbed and free water is left in the tubular fluid.

In the absence of ADH, this solute-free water is excreted and C_{H_2O} is positive.

In the presence of ADH, this solute-free water is not excreted but is reabsorbed by the late distal tubule and collecting ducts and C_{H_2O} is negative.

Calculation of C_{H_2O}

$$C_{H_2O} = V - C_{osm}$$

where

C_{H_2O} = Free water clearance (mL/min),
V = Urine flow rate (mL/min),

$$C_{osm} = \text{Osmolality clearance} = \frac{U_{osm}V}{P_{osm}} \text{ (mL/min)}$$

Example. If the urine flow rate (V) is 10 mL/min, urine osmolality (U_{osm}) is 100 mOsm/kg H_2O and plasma osmolality is 300 mOsm/kg H_2O, then the C_{H_2O} can be calculated as

$$C_{H_2O} = V - C_{osm}$$

therefore,

$$C_{H_2O} = 10\,\text{mL/min} - \frac{100\,\text{mOsm/kg}H_2O \times 10\,\text{mL/min}}{300\,\text{mOsm/kg}H_2O}$$

$$= 10\,\text{mL/min} - 3.33\,\text{mL/min}$$
$$= 6.7\,\text{mL/min}.$$

TESTS FOR TUBULAR FUNCTIONS

The reabsorptive and secretory functions of renal tubules can be tested by the following tests:

1. Urine concentration test. The ability of tubules to concentrate the urine is assessed by measuring the specific gravity of urine either after 12 h of water deprivation or 12 h after injection of vasopressin (ADH). In either case if the specific gravity of urine is above 1.020, the tubular function is considered to be normal.

2. Urine dilution test. In this test, patient is asked to drink 1 L of water and the urine sample is collected every hour for the next 4 h. Normally, at least 750 mL (75%) of urine should be excreted during this period, and at least one of the samples should have osmolality less than 100 mOsm/kg H_2O (or specific gravity below 1.004).

3. Urine acidification test. In this test, patient is given ammonium chloride (NH_4Cl) orally in the dose of 0.1 g/kg BW and pH of the urine is tested in a sample collected after 6 h. Normally, urine pH should be below 5.3 because after metabolism in the liver, the NH_4Cl yields HCl

$$NH_4Cl \rightarrow NH_3 + HCl$$

4. Other methods of study of tubular function, usually employed in research laboratory, include:

(i) *Micropuncture technique* to analyse the tubular fluid at various levels.
(ii) *Microcryoscopic studies* of renal tissue slices at different depths.
(iii) *Microelectrode studies* to measure the membrane potential of the tubular cells.

RADIOLOGY AND RENAL IMAGING

Though not strictly speaking, kidney function tests are quite useful investigations in present day clinical practice to assess anatomical and physiological abnormalities of the kidneys.

1. Plain radiograph of abdomen is useful in detecting calcium-containing (radiopaque) renal stones.

2. Intravenous pyelography is performed by injecting a radiopaque dye like urographin intravenously and taking radiographs of the abdomen at short intervals (1, 5, 10 and 30 min) (Fig. 6.6-3).

3. Ultrasonography is a quick, non-invasive, inexpensive and harmless method to evaluate size, shape, position of kidney and to detect tumour, stones, cysts, etc. of the kidneys, ureter, prostate and urinary bladder.

4. Computed tomography is performed to detect abnormalities in and around the kidneys as mentioned above in ultrasonography.

5. Radionuclide studies are carried out by injecting radioactive compounds which are concentrated and excreted by the kidneys. Radioactivity of the kidneys is recorded by a gamma camera.

RENAL BIOPSY

Renal biopsy is performed percutaneously with the help of needle. The biopsy specimen is subjected to light, electron and immunofluorescence microscopic studies. This technique has increased knowledge and better understanding of glomerular and tubular diseases.

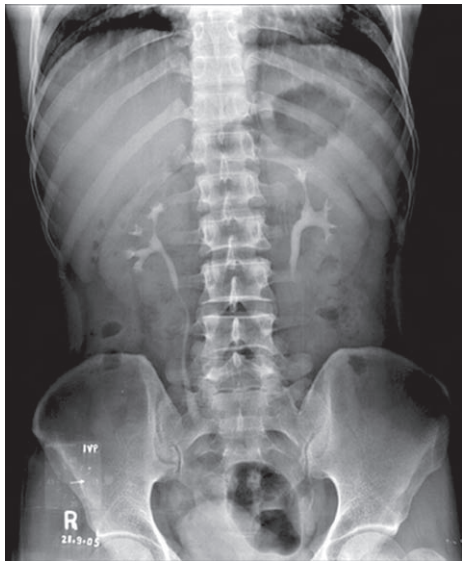


Fig. 6.6-3 Intravenous pyelogram.

DIALYSIS AND RENAL TRANSPLANTATION

DIALYSIS

The term dialysis in physiological sense refers to the diffusion of solutes from an area of higher concentration to the area of lower concentration through a semipermeable membrane. This principle has been used to dialyse the blood of patients with renal failure especially those developing uraemia.

Uraemia develops when more than 75% of nephrons are damaged and is characterized by:

- Accumulation of nitrogenous waste products in the blood,
- Metabolic acidosis and
- Hyperkalaemia.

By dialysis, the dissolved crystalloids of the plasma pass through a semipermeable membrane so that their levels are brought down to lower levels. Two types of dialysis procedures are available:

- Haemodialysis or artificial kidney and
- Peritoneal dialysis.

HAEMODIALYSIS OR ARTIFICIAL KIDNEY

Haemodialysis machine is also called artificial kidney. Haemodialysis is done in a hospitalized patient through intravenous (IV) line for 3–5 h. During haemodialysis, the patient's radial artery is connected to the haemodialysis machine. Inside the haemodialysis machine, the blood is passed through a long and coiled cellophane tube immersed in a dialysis fluid (Fig. 6.6-4). Heparin is used as an anticoagulant while passing the blood through the machine.

Dialyzing fluid. The composition of a dialyzing fluid is similar to that of the plasma, except it is free of waste products like urea, uric acid, etc. The fluid contains less amount of sodium, potassium and chloride ions than in the uraemic blood. But the quantity of glucose, bicarbonate and calcium ions are more in the dialyzing fluid than in the uraemic blood.

During haemolysis, the semipermeable cellophane membrane permits the free diffusion of the constituents of plasma except proteins. In this way, the dialysis of patient's blood removes the toxic waste products and restores normal electrolyte concentration in the plasma. The dialysed blood is returned back to the patient's body through a peripheral vein (Fig. 6.6-4). At a time about 500 mL is passed through the artificial kidney. Haemodialysis is done usually thrice a week in severe uraemia.

Haemodialysis can save the life in many types of *acute renal failure.* The intermittent haemodialysis may prolong the life of many patients with chronic renal failure, which may lead an active life for many useful years.

The dialysis can partially replace the excretory function of the kidneys but does not replace endocrine and metabolic functions.

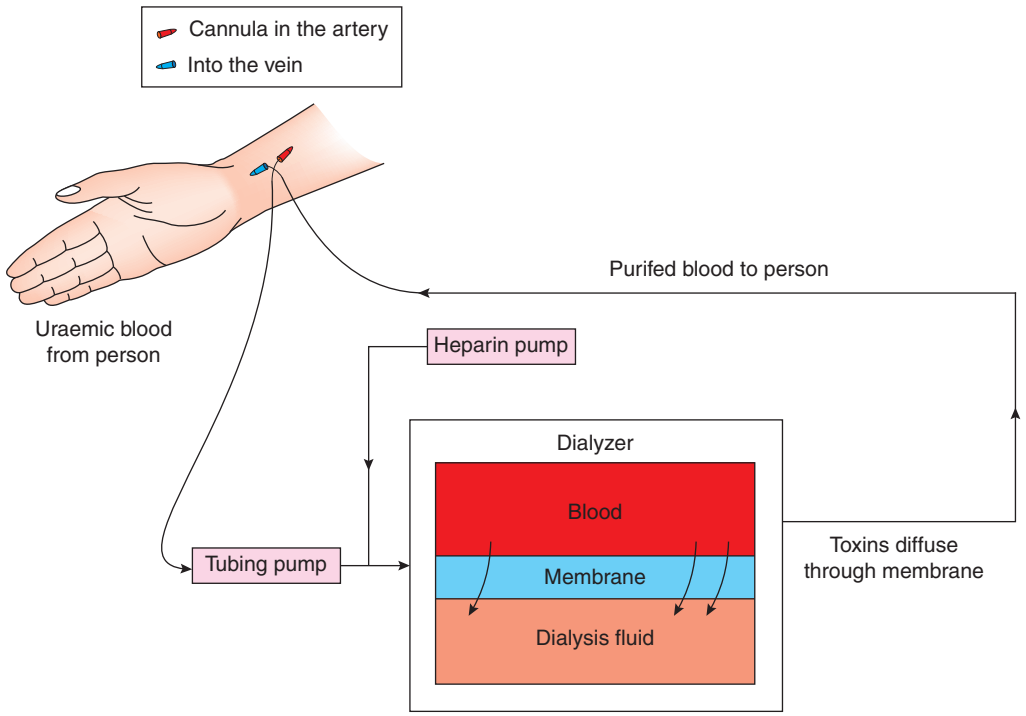


Fig. 6.6-4 Basic principle of haemodialysis.

PERITONEAL DIALYSIS

Peritoneal dialysis is a form of long-term dialysis done by the patients at home or at work. In this type of dialysis, the peritoneum acts as a semipermeable membrane.

Two litres of dialyzing fluid is introduced through a intraperitoneal catheter. It is then kept in the peritoneal cavity for exchange to take place for a period of 15–20 min called dwell time. Fluid is then drained out and measured. A strict input and output chart is maintained. The whole procedure constitutes one cycle. It is done at 6 h intervals (4 cycles/day), even when the patient is ambulatory or mobile. There is no need for hospitalization. It is useful for young children and old patients with cardiovascular disorders. It prolongs survival in patients with chronic renal failure for many years. Peritoneal dialysis is not very suitable for drug poisoning cases.

RENAL TRANSPLANTATION

Renal transplantation is the final answer to all the problems in cases with chronic renal failure. It reverses metabolic and excretory abnormalities. The graft is taken from a cadaver donor, or from a sibling or a parent. Usually left kidney of donor is transplanted to right iliac fossa of recipient. Long-term immune suppression with prednisolone and cyclosporin is needed. It offers complete rehabilitation and is most cost-effective option.

Chapter 6.7

Physiology of Micturition

URINARY BLADDER AND URETHRA

Gross anatomy

External features. The urinary bladder, a hollow muscular viscus, is a temporary reservoir for urine. The main body of empty bladder is pyramidal having an apex and a base. The lowest part of the bladder is called neck, which continues as urethra.

Interior of the bladder. In an empty bladder, the greater part of the mucosa shows irregular folds due to its loose attachments to the muscular coat. The interior of the base (posterior surface) of the bladder presents a triangular area, the *trigone* where the mucosa is smooth due to its firm attachment.

Internal urethral orifice is located at the apex (inferior angle) of the trigone. The ureters open into the bladder at superior angles of the trigone (Fig. 6.7-1). The ureters pierce the bladder wall obliquely, and this provides a valve-like action, which prevents a reverse flow of urine towards the kidneys as the bladder fills.

Structure of the bladder

The wall of the urinary bladder consists of three layers: an outer serous layer, a thick coat of smooth muscle, and the inner mucous membrane.

Mucous membrane is lined by the transitional epithelium. Its characteristic features are:

- It stretches when the bladder distends,
- It forms a complete barrier to the passage of fluid and electrolytes. Therefore, urine stored in the bladder remains unchanged in chemical composition.

Muscular layer is formed by smooth muscle fibres, which constitute the detrusor muscle. Contraction of this muscle coat is responsible for emptying of the bladder.

Urethra and its sphincters

Male urethra is about 20 cm in length, is divided into three parts: prostatic urethra (3 cm), membranous urethra (1.25 cm) and penile urethra (15.75 cm). Membranous urethra is surrounded by the external sphincter.

Female urethra is about 3.8 cm long. It extends from the neck of the bladder to the external meatus. It traverses the external sphincter and lies immediately in front of the vagina.

Sphincters of the urethra

1. Internal sphincter. The circular smooth muscle fibres in the area of the neck of bladder are thickened to form the internal sphincter (sphincter vesicae). The natural tone of the internal sphincter prevents emptying of the bladder

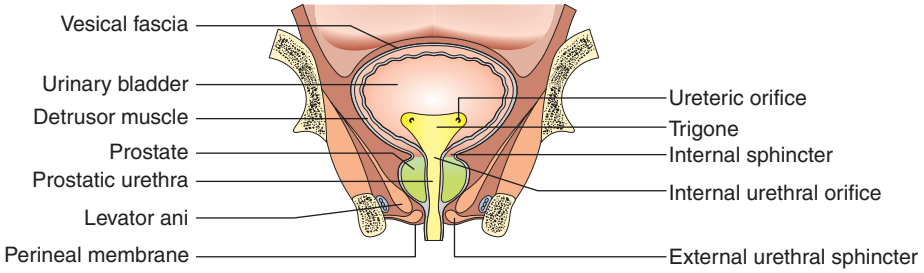


Fig. 6.7-1 Coronal section through the bladder and prostate to show the interior of the bladder, internal urethral sphincter and external urethral sphincter.

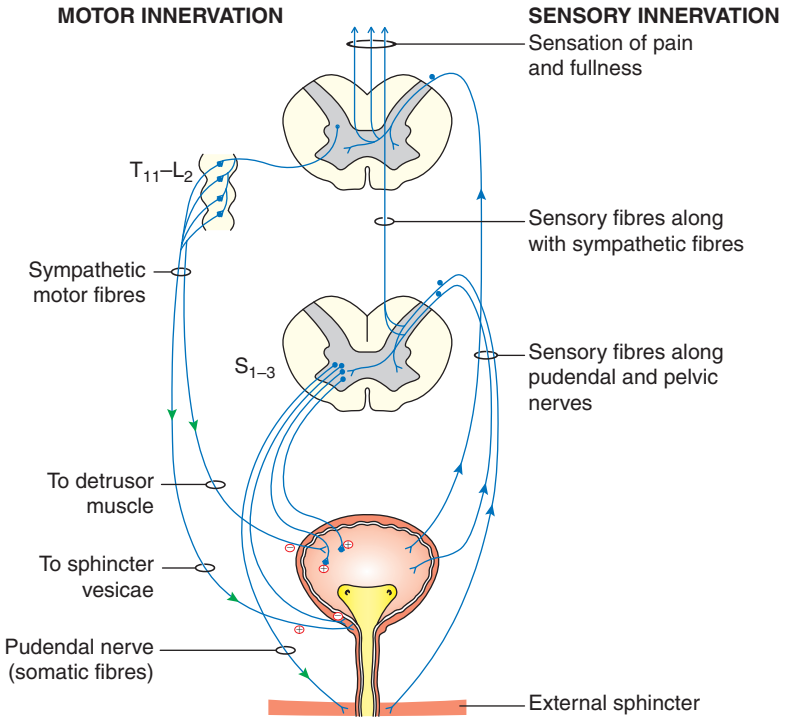


Fig. 6.7-2 Innervation of urinary bladder.

until the pressure in the body of bladder rises above a threshold level.

2. External sphincter. Beyond the bladder neck, it is encircled by a ring of voluntary (skeletal type) muscle known as external sphincter of the bladder. The external sphincter provides voluntary control over micturition.

Innervation of the urinary bladder (Fig. 6.7-2)

Motor innervation

Parasympathetic innervation. The parasympathetic efferent fibres (nervi erigentes) are derived from the second, third and fourth sacral segments (mainly S_2 and S_3). These fibres carry motor impulses to the urinary bladder causing contraction of detrusor muscle and emptying of the bladder. These fibres are inhibitory to the internal sphincter.

Sympathetic innervation. These nerves arise in the 11th thoracic to the second lumbar segments (T_{11}–L_2). These fibres are said to be inhibitory to detrusor muscle and motor to the sphincter vesicae.

Sympathetic activity is not involved in micturition. Increased sympathetic discharge to the bladder occurs during ejaculation and helps to prevent the reflux of sperms from the prostatic urethra into the bladder.

Somatic motor innervation. The somatic pudendal nerve (S_2, S_3 and S_4) supplies the external sphincter which is voluntary.

Sensory innervation

Sensation of bladder distension. Afferents from the detrusor stretch receptors travel to the spinal cord via the pelvic splanchnic nerve (nervi erigentes). From the region of the bladder neck and trigone, the afferents travel via the hypogastric plexus to spinal cord segments T_{11}–L_2.

In the spinal cord, the fibres of awareness of bladder distension run in the posterior column (fasciculus gracilis) to

reach the spinal, pontine and suprapontine micturition centres.

Sensation of bladder pain. The pain fibres are stimulated by excessive distension or spasm of the bladder wall, or by stone, inflammation or malignant disease irritating the bladder. The pain fibres run predominantly in the hypogastric plexus but are also present in the nervi erigentes.

In the spinal cord, the fibres carrying pain sensation run in the lateral spinothalamic tract.

Urethral sensations. Including sensation of imminent voiding associated with maximal bladder filling, reach the spinal cord via the pudendal nerve.

In the spinal cord, fibres carrying urethral sensations travel in the dorsal column.

PHYSIOLOGY OF MICTURITION

Micturition is the process by which urinary bladder empties when filled. The main physiological events in the process of micturition are:

- Filling of urinary bladder and
- Emptying of urinary bladder.

FILLING OF URINARY BLADDER

Transport of urine into urinary bladder through ureters

As urine collects in the renal pelvis, the pressure in the pelvis increases and initiates a peristaltic contraction beginning in the pelvis and spreading along the ureter to force urine towards the bladder.

Capacity of the bladder

Physiological capacity of the bladder varies with age, being 20–50 mL at birth, about 200 mL at 1 year, and can be as high as 600 mL in young adult males. In all cases, the physiological capacity is about twice that at which the first desire to void is felt.

Volume and pressure changes in bladder during filling

The normal bladder is completely empty at the end of micturition and the intravesical pressure is equal to the intra-abdominal pressure. As the bladder is filled up, it adjusts its tone and a fairly large volume of urine can be accommodated with minimal alterations in the intravesical pressure. This is possible because of the phenomenon of adaptation. The *adaptation* occurs because of the inherent property of *plasticity,* the smooth muscles of detrusor and *because of law of Laplace* (see page 306).

Cystometry. This refers to the process of studying the relationship between the intravesical volume and pressure, the cystometrogram refers to a graphical record of this relationship.

Normal cystometrogram shows three phases of filling (Fig. 6.7-3):

Phase Ia. It is the initial phase of filling in which pressure rises from 0 to 10 cm H_2O, when about 50 mL of fluid is collected in the bladder.

Phase Ib. It is the phase of plateau which lasts till the bladder volume is 400 mL. During this phase, the pressure in the bladder does not change much and remains approximately at 10 cm H_2O. This is because of adaptation of urinary bladder by relaxation, as described above.

Phase II. This phase starts beyond 400 mL volume when the pressure begins to rise markedly, triggering the micturition reflex. Normally, the voiding contraction raises the intravesical pressure by about 20–40 cm H_2O. If voiding is avoided (not initiated), the pressure rises from 10 cm H_2O onward, as shown by dotted lines beyond the phase II in Fig. 6.7-3. Beyond 600 mL, the urge to void urine becomes almost unbearable.

EMPTYING OF THE BLADDER

Emptying of the bladder is basically a reflex action called the micturition reflex, which is controlled by supraspinal centres and is assisted by contraction of perineal and abdominal muscles. Therefore, emptying of the urinary bladder focuses on:

- Micturition reflex,
- Voluntary control of micturition and
- Role of perineal and abdominal muscles in micturition.

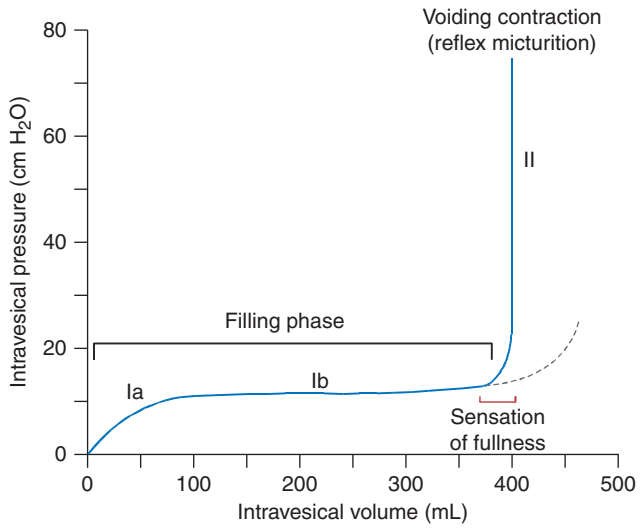


Fig. 6.7-3 Normal cystometrogram.

Micturition reflex

Initiation. Micturition reflex is initiated by the stimulation of the stretch receptors located in the wall of urinary bladder.

Stimulus. Filling of bladder by 300–400 mL of urine in adults constitutes the adequate stimulus for the micturition reflex to occur.

Afferents. The afferents from the stretch receptors in the detrusor muscle and urethra travel along the pelvic splanchnic nerves and enter the spinal cord through dorsal roots to S_2, S_3 and S_4 segments to reach the sacral micturition centre (Fig. 6.7-4).

Sacral micturition centre is formed by the sacral detrusor nucleus and sacral pudendal nucleus.

Efferents. Efferents arising from the sacral detrusor nucleus are the preganglionic parasympathetic fibres, which relay in the ganglia near or within bladder and urethra (Fig. 6.7-4). The post-ganglionic parasympathetic fibres are excitatory to the detrusor muscle and inhibitory to the internal sphincter.

Response. Once micturition reflex is initiated, it is self-regenerative, i.e. initial contraction of the bladder wall further activates the receptors to increase the sensory impulses (afferents) from the bladder and urethra which cause further increase in the reflex contraction of detrusor muscle of the bladder. The cycle thus keeps on repeating itself again and again until the bladder has reached a strong degree of contraction.

Once the micturition reflex becomes powerful enough, this causes another reflex which passes through pudendal nerves to external sphincter to cause its inhibition. If this inhibition is more potent than the voluntary constrictor signals from brain, then urination will not occur. If not so, urination will not occur unless the bladder fills still more and micturition reflex becomes more powerful.

Voluntary control of micturition

Role of supraspinal centres

The micturition reflex is fundamentally a *spinal reflex* facilitated and inhibited by higher brain centres *(supraspinal centres)* and, like defaecation, is subjected to voluntary facilitation and inhibition. In infants and young children, micturition is purely a reflex action. *Voluntary control* is gradually acquired as a learned ability of the toilet training. Once voluntary control is acquired, the supraspinal control

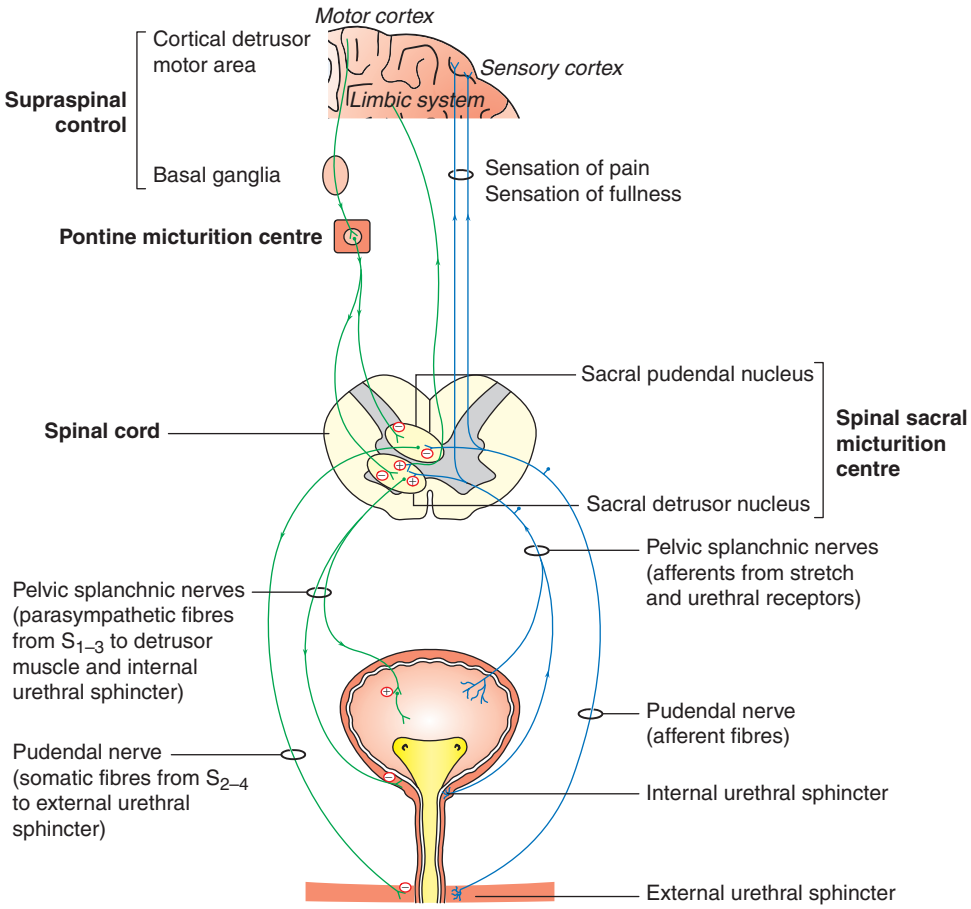


Fig. 6.7-4 Pathway and supraspinal control of micturition reflex.

centres exert final control of micturition by following means:

- The higher centres keep the micturition reflex partially inhibited all the time except when it is desired to micturate.
- When the convenient time to urinate present, the higher centres facilitate the sacral micturition centre (SMC) to initiate a micturition reflex and inhibit the external urinary sphincter so that urination can occur.

Supraspinal control centres which control the micturition reflex (a completely automatic cord reflex) include the pontine micturition centre (PMC) and suprapontine centres.

Pontine micturition centre, corresponds to the locus ceruleus of the rostral pons. Neurons from PMC descend in the reticulospinal tract and exert control over the SMC and thoracolumbar sympathetics. Function of PMC is *co-ordination of detrusor contraction and sphincter relaxation*, which is important for proper micturition.

Suprapontine centres which relay their influence on the sacral micturition centre through the PMC are:

- *Cerebral cortex*
- *Basal ganglion*
- *Limbic system*

Role of perineal and abdominal muscles in micturition

Certain muscular movements, which aid the emptying of bladder, but are not the essential component of micturition process are:

- At the onset of micturition, the levator ani and perineal muscles are relaxed, thereby shortening the post-urethra and decreasing the urethral resistance.
- The diaphragm descends and
- The abdominal muscles contract, accelerating the flow of urine by raising intra-abdominal pressure which in turn secondarily increase the intravesical pressure thereby increasing the flow of urine.

Note. Certain important facts about micturition are:

- A voiding contraction, once initiated, is normally maintained until all the urine has been discharged from the urinary bladder. This is a function of facilitating impulses from the higher centres. However, if required so, the micturition can be voluntarily stopped in between by inhibitory impulses from the higher centres.
- The bladder contracts in all directions like a toy balloon deflating from its neck.
- After urination, the female urethra empties by gravity, whereas the urine remaining in the urethra of male is expelled by several contractions of bulbospongiosus muscle.

Note. In the urinary bladder dysfunction, bladder contractions are insufficient to completely empty the bladder, therefore, some urine is left in the urinary bladder called residual urine.

ABNORMALITIES OF MICTURITION

EFFECT OF INTERFERENCE WITH NERVOUS CONTROL OF BLADDER

1. Transection of sympathetic supply

Following effects are produced:

- In man, the immediate effect would be the relaxation of ureteric reflexes, trigone and internal sphincter.
- Later, the internal sphincter may recover and closes completely, though it gives way easily when a catheter is passed.
- After an initial and inconstant period of frequency of micturition, bladder function is re-established in a comparatively normal way.

2. Effect of deafferentation or atonic bladder

- The destruction of sensory nerve fibres from the bladder to spinal cord prevents transmission of stretch signals from the bladder and therefore, also prevents micturition reflex contractions.
  - In these conditions, the person loses all bladder control despite intact efferent fibres from the cord to the bladder and despite intact neurogenic connections with brain.
  - Instead of emptying periodically, the bladder fills to capacity and overflows a few drops at a time through the urethra. This is called *overflow dribbling.*

Causes of atonic bladder are:

- *Syphilis.* It frequently causes constrictive fibrosis around the dorsal nerve root fibres where they enter the spinal cord and subsequently destroys these fibres.
- *Crushing injuries to spinal cord.* It damages the sensory roots.

3. Effect of denervation

When there is interruption with both afferent and efferent nerves of bladder, the following consequences are observed:

- The bladder is flaccid and distended for a while,
- Gradually, however, the muscle of *decentralized bladder* becomes active, with many contraction waves that expel dribbles of urine out of the urethra and

- The bladder becomes shrunken and the bladder wall hypertrophies.

4. Effect of spinal cord transection

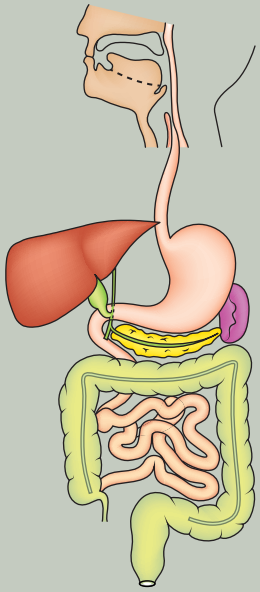
During spinal shock

- Voluntary micturition is completely abolished. The activity of detrusor muscle remains in abeyance for a long period, but sphincter now returns very soon. At this stage, bladder responds to filling in the same manner as the dead organ or an elastic bag. Retention of urine is therefore, complete from an early stage. If no catheter is passed the bladder becomes increasingly overstretched. The sphincter is finally forced open by a high intravesical pressure and small quantities of urine escape at frequent intervals—a condition of *retention with overflow*.
- The capacity is reduced and its walls become hypertrophied. This type of bladder is sometimes called *spastic neurogenic bladder*.

After spinal shock has passed, the voiding reflex returns, although there is no voluntary control. Some paraplegic patients train themselves to initiate voiding by pinching or stroking their thighs, provoking a mild mass reflex.

SECTION 7

Gastrointestinal System



To sustain life, body needs a continual supply of water, electrolytes and nutrients. This function is served by the gastrointestinal or the so-called digestive system.

Gastrointestinal system comprises alimentary canal and other associated organs, such as liver, gall bladder and pancreas. Alimentary canal is a long tube starting at the mouth, passing through pharynx, oesophagus, stomach, small intestine, large intestine, rectum and ending at anus.

FUNCTIONS OF GASTROINTESTINAL SYSTEM

I. Digestive functions. The major function of the gastrointestinal system is to transfer nutrients, minerals and water from external environment to the circulating body fluids for distribution to all the body tissues. This function is accomplished by following processes:

1. Ingestion of food. It involves:

- Placing the food into the mouth. Most of the foodstuffs are taken into mouth as large particles mainly made of carbohydrates, proteins and fats.
- Chewing the food into smaller pieces is carried out with the help of teeth and jaw muscles. This process is called mastication.
- Lubrication and moistening of the food is done by the saliva.
- Swallowing the food (deglutition). It refers to pushing the bolus of food from mouth into the stomach. It is accomplished in three phases: oral phase, pharyngeal phase and oesophageal phase.

2. Digestion of food. It refers to the conversion of complex insoluble large organic molecules (food) into soluble, smaller and simpler molecules which can be easily absorbed. Digestion of food is accomplished with the help of hydrochloric acid and digestive juices containing various enzymes.

3. Absorption of digested food. Absorption of food refers to the movement of digested molecules from the lumen of alimentary canal across its epithelial lining to the blood or lymph. The absorbed water, electrolytes and nutrients are carried away to the various tissues by the circulating blood.

4. Egestion, i.e. excretion of unwanted undigested food by the alimentary canal in the form of faeces is called defaecation.

To understand the digestive function of gastrointestinal system it is imperative to have knowledge about:

- Functional anatomy and organization of the gastrointestinal system,
- Gastrointestinal motility,
- Gastrointestinal blood flow,
- Role of salivary glands, liver, gall bladder and pancreas, and
- Neural and hormonal control of gastrointestinal functions.

II. Non-digestive functions. The main non-digestive function of the gastrointestinal system is its role as an immune system. The lymphoid tissue in the tonsils, adenoids and Peyer's patches constitute an important part of body's immune system. These provide both the humoral and cellular immunity, which is especially effective against the micro-organisms trying to enter the body from the alimentary canal.

“This page intentionally left blank"

Chapter 7.1

Functional Anatomy and General Principles of Functions of Gastrointestinal System

FUNCTIONAL ANATOMY

FUNCTIONAL ORGANIZATION

The digestive system comprises gastrointestinal tract (GIT) and accessory organs of digestion like teeth, tongue, salivary glands, liver and exocrine part of pancreas.

Gastrointestinal tract, also known as alimentary canal, is basically a muscular tube extending from the mouth to the anus (Fig. 7.1-1). At either end, the lumen is continuous with external environment. It measures about 10 m (30 ft) and comprises following parts:

Mouth. Mouth is loosely used term to denote the external opening and for the cavity it leads to. The cavity containing anterior two-thirds of tongue and teeth is the *mouth cavity* or *oral cavity* or *buccal cavity* (Fig. 7.1-2). The oral cavity extends from the lips to the oropharyngeal isthmus, i.e. junction of the mouth with the pharynx. Oral cavity is subdivided into two parts: the vestibule and oral cavity proper (Fig. 7.1-2).

- *Vestibule lies* between the lips and cheeks externally and the gums and teeth internally.
- *Oral cavity proper* lies within the alveolar arches, gums and teeth.

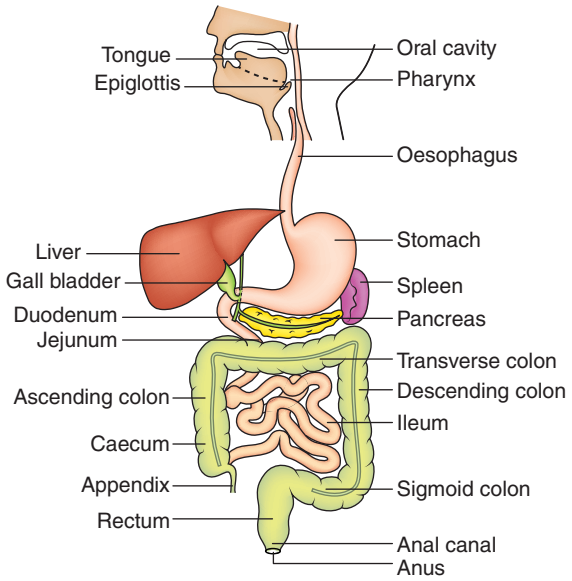


Fig. 7.1-1 The gastrointestinal system.

Tongue, in the digestive system, plays two important roles:

- Tells the taste of food and
- Helps in chewing and swallowing of the food.

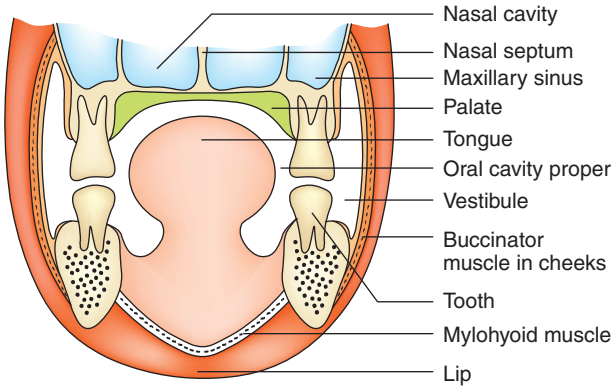


Fig. 7.1-2 Schematic coronal section through oral cavity.

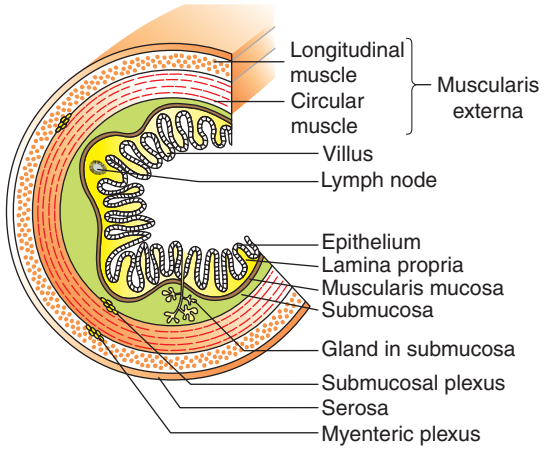


Fig. 7.1-3 Cross-section of the alimentary canal depicting structural characteristics of its wall.

Teeth. Functions of different types of teeth in chewing are:

- *Incisors* provide strong cutting action,
- *Canines* are responsible for tearing action,
- *Premolars and molars* have grinding action.

Pharynx. The pharynx is a median passage that is common to the gastrointestinal and respiratory systems.

Oesophagus. It is a fibromuscular tube about 25 cm long. At its junction to the pharynx, upper oesophageal sphincter is present and its junction with the stomach lower oesophageal sphincter is present.

Stomach. It is a hollow muscular bag connected to the oesophagus at its upper end and to the duodenum at the lower end.

Small intestine. It is a long tubular structure which can be divided into three parts:

- *Duodenum* is the first part of small intestine. It is C-shaped and measures about 25 cm in length,
- *Jejunum,* the middle part of the small intestine is about 25 m long and
- *Ileum,* the last part of small intestine, is about 3.5 m long.

Large intestine. It arches around and encloses the coils of the small intestine and tends to be more fixed than the small intestine. It is divided into following parts (Fig. 7.1-1):

- *Caecum,*
- *Colon,*
- *Anal canal.*
- *Appendix,*
- *Rectum,* and

STRUCTURAL CHARACTERISTICS OF GIT WALL

Different parts of the GIT are specialized for carrying out different functions particularly digestion and absorption, but the basic structural characteristics of the wall of whole GIT are similar. The intestinal wall from inside to outwards consists of following layers (Fig. 7.1-3):

1. Mucosa (mucous layer). It is innermost coat consisting of three layers:

- *Surface epithelium* lining the luminal surface consists of epithelial cells which vary in type from simple squamous to tall columnar depending upon the function of the part of GIT.
- *Lamina propria* is composed of loose connective tissue, which contains numerous glands, small blood vessels, lymphatics and nerve fibres.
- *Muscularis mucosa* is composed of two thin layers of smooth muscle fibres, which help in localised movements of the mucosa.

2. Submucosa. This refers to the layer of connective tissue present outside the mucosa. It contains blood vessels, lymphatics and a network of nerve fibres and nerve cells called submucosal nerve plexus (Meissner's plexus).

3. Muscle coat. It is formed by a thick layer of smooth muscle fibres surrounding the submucosa. The smooth muscle fibres are arranged in two layers:

- *Circular muscle fibres* form the inner layer, and
- *Longitudinal muscle fibres* form the outer layer.

In between the circular and longitudinal muscle fibres is present an extensive network of nerve cells and fibres named *Auerbach's plexus* (myenteric plexus).

4. Serosa (serous layer). This is the outermost layer consisting of a layer of connective tissue. This layer helps in the attachment of gut to the surrounding structures.

INNERVATION OF THE GIT

The innervation of the GIT includes intrinsic and extrinsic system (Fig. 7.1-4).

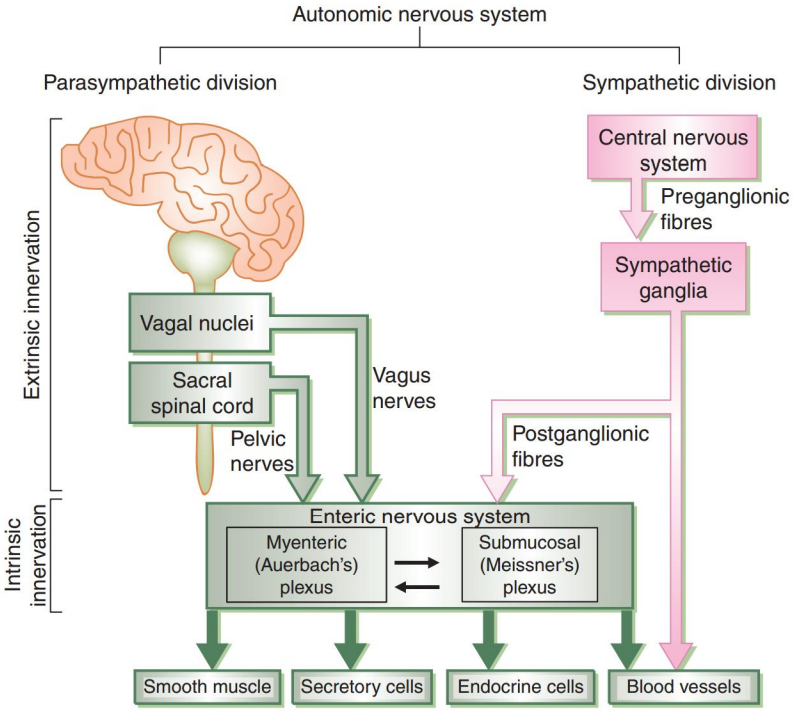


Fig. 7.1-4 Schematic illustration of the innervation of gut.

1. Intrinsic innervation

The intrinsic nervous system, also called as *enteric nervous system,* consists of nerve cells and fibres which originate and are located in the intestinal wall itself. This system supplies the smooth muscles of GIT (i.e. musculature of GIT except upper oesophagus and external anal sphincter which contain striated muscle). This system controls most of the gastrointestinal functions like secretion and motility. The enteric nervous system is composed mainly of two plexuses:

Myenteric plexus or Auerbach's plexus is present in between the circular and longitudinal muscle fibres of muscular coat of the GIT. Stimulation of myenteric plexus causes increase in: tone of the gut wall, intensity of rhythmical contractions of gut wall, rate of contraction and velocity of contraction.

Meissner's plexus or submucosal plexus is present in the submucosal layer. It controls the secretory activity and blood flow to the gut. In contrast to the myenteric plexus, it is mainly concerned with controlling function within the inner wall of each minute segment of the intestine. By receiving sensory signals from the mucosal epithelium and from stretch receptors in the wall of alimentary canal it helps to control local intestinal secretion, local absorption and local contraction of the submucosal muscle.

The Auerbach's and Meissner's plexuses are interconnected with each other and are under the control of parasympathetic and sympathetic components of extrinsic nervous system. In both the plexuses, the axons branch profusely, so that stimulation of one region produces a widespread response in the GIT.

2. Extrinsic innervation

The extrinsic system of nerves supplying the gut consists of the parasympathetic and sympathetic components of autonomic nervous system.

Parasympathetic innervation. The parasympathetic supply to the gut is made up of cranial and sacral divisions:

- *Cranial parasympathetic fibres* originate in medulla, come through vagus and supply the oesophagus, stomach, small intestine, pancreas and first half of the large intestine. They also make synaptic connections with intramural plexuses.
- *Sacral parasympathetic fibres* originate in sacral spinal cord, pass through pelvic nerves to hypogastric (pelvic) ganglion as a postganglionic fibre, supply the distal half of large intestine and rectum. The sigmoid, rectal and anal regions have an especially rich supply of parasympathetic fibres that function in the defaecation reflex.

Functions of parasympathetics. Parasympathetic stimulation causes excitation of all the musculature of gut except the sphincters to which it inhibits. There occurs an increase in the gastrointestinal motility and secretory activity.

Sympathetic innervation. The sympathetic fibres to gut arise from eighth thoracic (T_8) to second lumbar (L_2) spinal segments. The sympathetics innervate all portions of the GIT

rather than being more extensively supplied to portions near the oral cavity and anus, as is true for parasympathetics.

Functions of sympathetic innervation. Sympathetic stimulation causes:

- Vasoconstriction,
- Excitation of ileocaecal and internal anal sphincters, and smooth muscles of muscularis mucosa throughout (to increase number of folds),
- Inhibition of motility in the gut.

Thus most of the effects of sympathetic stimulation are opposite to that of the parasympathetic stimulation.

GASTROINTESTINAL BLOOD FLOW

The blood supply of the GIT forms part of splanchnic circulation, which has been described in Chapter 4.6 on 'Regional Circulation' (see page 277). The main characteristic feature of the gastrointestinal blood flow is that it is *usually proportional to the level of local activity.*

GENERAL PRINCIPLES OF GASTROINTESTINAL FUNCTIONS

The main activities involved in the functioning of GIT are GIT motility and GIT secretion. The general principles governing these activities are discussed here.

GENERAL PRINCIPLES OF GASTROINTESTINAL MOTILITY

Characteristics of gastrointestinal smooth muscle functioning

The motor functions of the gut are performed by the different layers of smooth muscles in its wall. The *gastrointestinal smooth muscle functions as a syncytium,* i.e. when an action potential is elicited within the muscle mass, it travels in all directions in the muscle and it contracts as a whole mass. The distances that it travels depend on the excitability of the muscle. This occurs because of the fact that the smooth muscle fibres in the longitudinal and circular muscle layers are electrically connected through the gap junctions that allow the ions to move from one cell to the next.

Electrical activity of gastrointestinal smooth muscle

Resting membrane potential (RMP) of gut smooth muscle fluctuates between –50 and –60 mV and thus shows undulating changes in the form of slow waves. The cause of these waves is not exactly known, probably might be due to slow undulation of the activity of sodium–potassium pump. These waves determine the rhythm of most gastrointestinal contractions.

Factors affecting RMP of gastrointestinal smooth muscles. The basic level of RMP of gastrointestinal smooth muscle can be increased or decreased.

Factors that depolarize the membrane include:

- Stretching of the muscle,
- Stimulation by acetylcholine,
- Stimulation by parasympathetic nerves that secrete acetylcholine at their endings and
- Stimulation by gastrointestinal hormones.

Factors that hyperpolarize the membrane are:

- Effect of norepinephrine or epinephrine on the muscle membrane,
- Stimulation by sympathetic nerves that secrete norepinephrine at their endings.

Action potentials that cause muscle contraction occur in the form of spike potentials. They occur when the RMP becomes more positive than about –40 mV. The channels responsible for the action potentials are called *calcium–sodium channels.* These channels allow particularly large number of calcium ions to enter along with a smaller number of sodium ions.

Functional types of gastrointestinal movements

Peristalsis refers to the movement of gut. Functionally, two types of peristalsis are recognized: propulsive movements and mixing movements. The details of these movements are described in the physiological activity by different parts of GIT.

GASTROINTESTINAL HORMONES

Gastrointestinal hormones: an overview

- Gastrointestinal hormones regulate the secretions and even to some extent the motility of GIT.
- The *glandular cells* secreting gastrointestinal hormones are individually scattered in the epithelium of the stomach and small intestine and not in the form of clusters of cells as in the endocrine glands.
- The luminal surface of glandular cells when stimulated by various chemicals present in the chyme release hormone from the opposite surface into blood capillaries of portal circulation.
- Through portal circulation, the released hormone reaches the target tissue situated in the nearby region of GIT and exhibits physiological actions on the target cells with specific receptors for the hormone. For example, the hormone gastrin released by G-cells present in mucosa of pyloric part of stomach in response to presence of peptides in the chyme reaches the body of stomach via portal circulation and increases the acid secretion as well as motility of the stomach.

- The effects of gastrointestinal hormones persist even after nervous connections between the site of release and the site of action have been severed.
- Gastrointestinal hormones are characterized by two specific features:
  - Each hormone (even at physiological concentration) may affect more than one target tissue. For instance, the secretin increases the secretion of not only pancreatic juice, but also of bile.
  - Each target tissue usually responds to more than one gastrointestinal hormone. For example, acid secreting cells of gastric glands are stimulated by gastrin but inhibited by secretin.

Classification of gastrointestinal hormones

The gastrointestinal hormones, based on their physioanatomical similarities, can be broadly classified into three types:

1. Gastrin family of hormones includes:
- Gastrin (for details see page 468), and
- Cholecystokinin PZ or CCK-PZ (see page 485).

2. Secretin family of hormones includes:
- Secretin (for details see page 485),
- Gastric inhibitory polypeptide or GIP,
- Vasoactive intestinal peptide or VIP (see page 499),
- Glucagon (for details see page 606) and
- Glucagon-like immune reactivity or GLI or glicentin.

3. Other gastrointestinal hormones include:
- Motilin,
- Neurotensin,
- Substance P,
- Gastrin releasing peptide or GRP, and
- Somatostatin (see page 607).

Actions of gastrointestinal hormones

The details about the action of various gastrointestinal hormones are described somewhere else (see pages given in parentheses above). However, the outlines of the action of each gastrointestinal hormone as well as the stimulus for secretion and site of secretion are depicted in Table 7.1-1.

Table 7.1-1 Stimuli for secretion, site of action and actions of gastrointestinal hormones

Hormone	Stimuli for secretion	Site of secretion	Actions						
			Gastric secretion	Gastric motility	Pancreatic secretion	Bile secretion	Gall bladder contraction	Small intestine secretion	Small intestine motility
• Gastrin	Small peptides, amino acids, gastric distention, vagal stimulation	G cells of gastric antrum	+	+	+	0	0	0	0
• Cholecystokinin (CCK)	Small peptides, amino acids, fatty acids	Type I cells of duodenum and Jejunum	0	–	+	0	+	0	+
• Secretin	Acid, fatty acids	S cells of duodenum	–	0	+	+	0	0	0
• Gastric inhibitory polypeptide (GIP)	Fatty acids, amino acids, oral glucose	Duodenum and Jejunum	–	–	0	0	0	+	–
• Vasoactive intestinal polypeptide (VIP)	Fatty acids	Jejunum	–	–	0	0	0	0	0
• Somatostatin	Acid in stomach	δ cells of islets of Langerhans	–	–	–	0	–	0	0

0 = No effect; + = stimulatory effect; – = Inhibitory effect.

Chapter 7.2

Physiological Activities in Mouth, Pharynx and Oesophagus

INTRODUCTION

The functioning of digestive system starts from the mouth (oral cavity) and ends at the anus. **Ingestion** of food involves following processes:

- Placing of food into the mouth,
- Mastication, i.e. chewing the food into smaller pieces,
- Lubrication of the food with saliva and
- Swallowing, i.e. deglutition.

The above mentioned physiological activities which take place in the mouth, pharynx and oesophagus are discussed in this chapter.

MASTICATION

Mastication or chewing refers to the process by which the food placed in the mouth is cut and grounded into smaller pieces. It involves:

- Movements of the jaws,
- Action of teeth—the incisors provide a strong cutting action, whereas the molars have a grinding action and
- Co-ordinated movements of the tongue and muscles of the oral cavity.

Chewing reflex

Mastication or chewing, though a voluntary act, is co-ordinated by a chewing reflex that facilitates the opening and closing of the jaw. The chewing reflex operates as:

When the mouth is opened to place the food inside it, the muscles of jaw are stretched which leads to their contraction due to stretch reflex, thereby raising the jaw to cause closure of the mouth.

- When the mouth is closed, the food comes in contact with buccal receptors which cause reflex inhibition of the muscles of mastication and also initiate a reflex contraction of the digastric and lateral pterygoid muscles, causing the mouth to open.
- This cycle of opening and closing the jaw leads to mastication. The tongue contributes to the grinding process by positioning the food between the upper and lower teeth.

Muscles of mastication

- *Masseter* raises the mandible, *clenches* the teeth and helps to protract the mandible.
- *Temporalis* raises the mandible and helps to retract the mandible after protraction.
- *Internal and external pterygoids* protrude the mandible, depress the chin and, therefore, help in opening the mouth.

Grinding movements are produced by these when right and left muscles are acting alternatively.

- *Buccinator* is an accessory muscle of mastication which prevents accumulation of food between the cheek and teeth.

Functions of mastication

1. Breaking of food into smaller pieces increases the total surface area. As the digestive enzymes act mainly on the surface of food particles, so digestion rate is increased.
2. Undigestive cellulose membrane present around the nutrition portion of most fruits and raw vegetables is broken, making it easier for them to be digested.
3. Mixing of food with saliva initiates the process of starch digestion by salivary amylase and lipid digestion by lingual lipase.
4. Swallowing becomes easy because of breaking of food into smaller pieces, and lubrication and softening of the food bolus by saliva.
5. Chewing brings food into contact with taste receptors and releases odour that stimulates the olfactory receptors. Stimulation of taste receptors and olfactory receptors increase the pleasure of eating and stimulate gastric secretions.

Net effect of mastication. The bolus of food becomes a homogenized mixture of small food particles, saliva and mucus, which is easy to swallow and digest.

LUBRICATION OF FOOD BY SALIVA

SALIVARY GLANDS

In addition to the chewing, another important physiological activity which takes place in the mouth is lubrication of food by saliva. The saliva is secreted by three pairs of major salivary glands:

1. Parotid glands

Location. Parotid glands are the largest salivary glands (each weighing 20–30 g), located near the angles of jaw.

Acini. The parotid glands are purely serous glands (Fig. 7.2-1) which secrete watery saliva containing more than 90% water. Parotid glands secrete 25% of the total salivary secretion (which is about 1500 mL/day).

Ducts. Ducts of the parotid glands open on the inner side of the right and left cheek and pour their secretions in the vestibule.

2. Sublingual glands

Location. The sublingual gland is the smallest of the three main salivary glands. It lies just below the mucosa on the floor of mouth. Each gland raises a ridge of mucosa which starts at the sublingual papilla and runs laterally and backwards. The ridge is called the sublingual fold.

Acini. The sublingual gland contains both serous and mucous acini (Fig. 7.2-1), the latter predominating.

Ducts. Ducts of sublingual gland are 8–20 in number. Most open into the mouth on the summit of the sublingual fold but a few may open into the submandibular duct.

3. Submandibular glands

Location. The submandibular glands are the large salivary glands which lie (one on each side) partly under cover of the body of the mandible.

Acini. The submandibular gland is composed of a mixture of serous and mucous acini, the former predominating (Fig. 7.2-1).

Ducts. S-shaped duct of each submandibular gland opens on the sublingual papilla located just lateral to the frenulum lingua.

Note. The sublingual and the submandibular glands secrete a fluid that contains a higher concentration of proteins and so is more viscus as compared to the watery secretion of parotid glands.

Smaller salivary glands

In addition to the three pairs of salivary glands described above, several smaller glands are located throughout the oral cavity. Those in the tongue secrete lingual lipase.

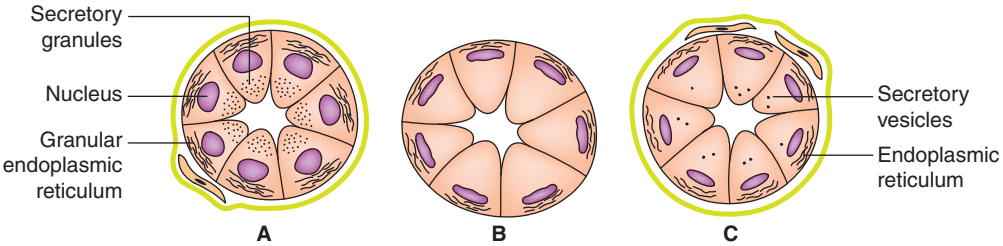


Fig. 7.2-1 Different types of acini in salivary glands: A, serous; B, mucous and C, seromucous.

SALIVA

SECRETION AND COMPOSITION

Amount. Under normal circumstances, the salivary glands secrete about 500–1500 mL of saliva every day. pH of saliva varies from 6 to 7.4.

Composition. Saliva is composed of *water* 99%, and *solids* 1%, which include:

- *Organic substances*, such as L-amylase (ptyalin), lingual lipase, kallikrein, lysozyme, small amounts of urea, uric acid, cholesterol and mucin.
- *Inorganic substances*, mainly are Na^+, Cl^-, K^+ and HCO_3^- whereas, Ca^{2+}, PO_4^{3-} and Br^- are in traces.

Note. Composition of saliva varies with the salivary flow rate.

Mechanism of formation of saliva

Mechanism of formation of saliva involves two processes:

1. Primary secretion of saliva. The acinar cells of salivary glands secrete the initial saliva into the salivary ducts. The initial saliva is isotonic, i.e. has the same Na^+, Cl^-, K^+ and HCO_3^- concentrations as plasma (Fig. 7.2-2). However, the initial saliva is soon modified by the salivary ducts.

2. Modification of saliva. The ductal cells that line the tubular portions of the salivary ducts change the composition of initial saliva by following processes (Fig. 7.2-2):

- *Reabsorption of Na^+ and Cl^-* occurs in the ductal cells, therefore, the concentration of these ions is lower than their plasma concentration.
- *Secretion of K^+ and HCO_3^-* is caused by the ductal cells, therefore, the concentrations of these ions are higher than their plasma concentrations.

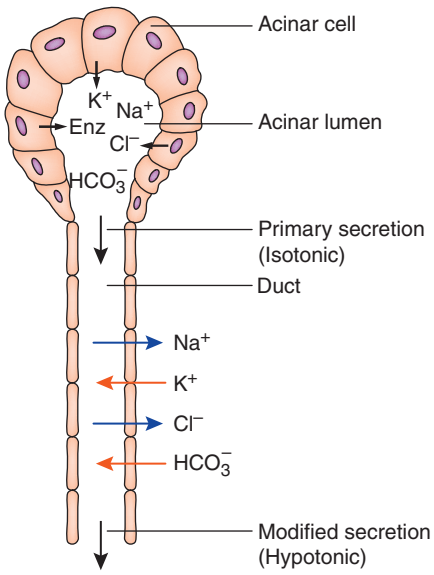


Fig. 7.2-2 Mechanism of formation of saliva.

Modified saliva becomes hypotonic in the ducts because the ducts are relatively impermeable to water.

Note. Aldosterone acts on the ductal cells to increase the reabsorption of Na^+ and Cl^- from the salivary ducts (analogous to its actions on renal tubule). Thus a high Na^+/Cl^- ratio is seen when aldosterone is deficient in Addison's disease (see page 595) and in presence of excess aldosterone, the concentration of sodium chloride in saliva falls almost to zero and increases K^+ concentration.

Effects of flow rate on the composition of saliva

1. At high flow rates, there is less time for reabsorption and secretion, and therefore the saliva is most like the initial secretion by the acinar cells. Thus, with the increase in flow rate the concentration of ions changes (Fig. 7.2-3):

- Sodium ion (Na^+) concentration increases progressively to a plateau value of 80–90 mEq/L.
- Chloride ions (Cl^-) concentration increase to about 50 mEq/L.

  Note. Na^+ and Cl^- concentrations of saliva are always lower than that in the plasma.

- Potassium ion (K^+) concentration decreases to 15–20 mEq/L.
- Bicarbonate ion (HCO_3^-) concentration increases when salivary flow rate increases (up to 50–70 mEq/L).

2. At low flow rates, there is more time for reabsorption and secretion, therefore, the modified saliva under resting conditions contains:

- Low concentration of Na^+ (about 15–20 mEq/L)
- Low concentration of Cl^- (15–20 mEq/L)
- Low concentration of HCO_3^- (10–15 mEq/L)
- High concentration of K^+ (25–30 mEq/L)

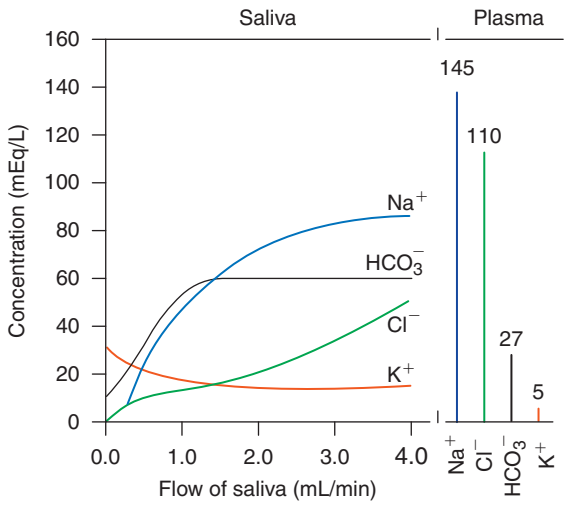


Fig. 7.2-3 Effect of flow rate on composition of saliva.

PHASES OF SALIVARY SECRETION

1. Cephalic phase refers to the secretion of saliva before entering of food into the mouth. It is caused by a conditioned reflex initiated by the mere sight or smell of food.

2. Buccal phase refers to the secretion of saliva caused by stimulation of buccal receptors by the presence of food in the mouth. It is an unconditioned reflex, partially regulated by the appetite area of the brain.

3. Oesophageal phase occurs due to the stimulation of salivary glands to a slight degree by the food passing through oesophagus.

4. Gastric phase refers to the secretion of saliva by the presence of food in the stomach. It specially occurs when irritant food is present in the stomach (e.g. increased salivation before vomiting).

5. Intestinal phase refers to a salivary secretion caused by the presence of irritant food in the upper intestine.

CONTROL OF SALIVARY SECRETION

- Salivary secretion is controlled entirely by the autonomic nervous system reflexes.
- Salivary secretion production is increased by both parasympathetic and sympathetic activity; however, the activity of former is more important.

I. Parasympathetic control

Parasympathetic nerve supply (Fig. 7.2-4)

Parotid glands are supplied by the parasympathetic fibres (preganglionic), which arise from the inferior salivary nucleus (dorsal nucleus of IXth nerve) of medulla.

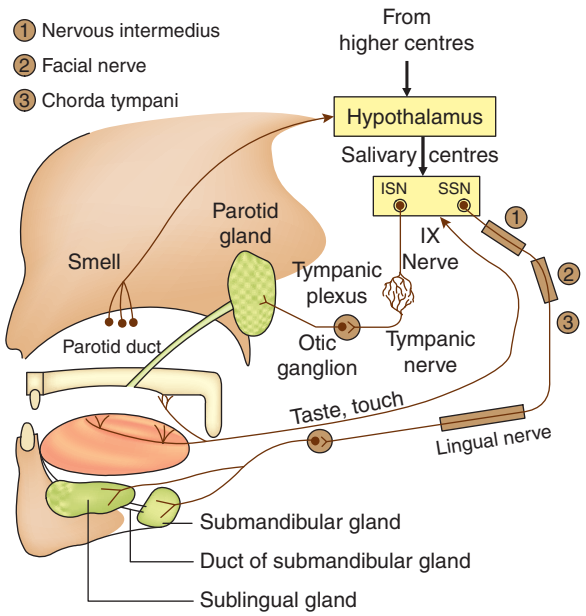


Fig. 7.2-4 Parasympathetic nerve supply to salivary glands.

- Pre-ganglionic fibres run via tympanic nerve and small superficial petrosal nerve to otic ganglion.
- Post-ganglionic fibres from the otic ganglion join auriculotemporal nerve to reach parotid gland where fibres are supplied along with blood vessels of gland.

Submandibular and sublingual glands are supplied by the parasympathetic fibres originating from superior salivary nucleus (dorsal nucleus of VIIth nerve).

- Pre-ganglionic fibres run in the nervous intermedius (sensory division of VIIth nerve), join the facial nerve and leave by its chorda tympani branch to join lingual nerve. They synapse in the ganglia present near the glands.
- Post-ganglionic fibres arising from the ganglia present near the glands are supplied to the glands along with the blood vessels.

Parasympathetic reflexes

Parasympathetic nerves are secretomotor to the salivary glands and control their secretion via following reflexes:

1. Conditioned reflexes. Sight, smell or even thought of palatable food increase the salivary secretion by the conditioned reflexes. In conditioned reflexes, the parasympathetics supplying the salivary glands are stimulated by impulses coming from higher centres of brain.

2. Unconditioned reflexes are initiated by the stimulation of receptors in the buccal cavity. Receptors and afferents, and efferents of unconditioned reflexes are:

Receptors and afferents

- *Mechanoreceptors* which are excited by tactile stimulation from the tongue, mouth and pharynx. The tactile stimuli occur due to the presence of food in the buccal cavity, chewing movements and irritation of buccal mucosa.
- *Afferent* run in trigeminal nerve branches (such as lingual, buccal and palatine nerves), pharyngeal branches of vagus and glossopharyngeal nerve:
- *Chemoreceptors*, i.e. taste buds are stimulated by the sensation of taste and chemicals in the food. Afferents for taste sensation from:
  - Posterior 1/3rd of tongue pass via glossopharyngeal nerve to end in inferior salivary nucleus (dorsal nucleus of IXth nerve), and
  - From anterior 2/3rd of tongue pass via nervous intermedius (branch of VIIth nerve) to end in superior salivary nucleus (dorsal nucleus of VIIth nerve).
- *Salivary centre* is thus constituted by superior and inferior salivary nuclei.
- *Efferents* from superior salivary nucleus stimulate the submandibular and sublingual salivary glands, while those from the inferior salivary nucleus stimulate the parotid glands.

Effects of parasympathetic stimulation

Parasympathetic nerve stimulation causes the salivary gland cells to secrete a large volume of watery fluid that is high in electrolytes but low in proteins.

II. Sympathetic control

Sympathetic nerve supply

- *Pre-ganglionic fibres* originate from the lateral horn cells of T_1 and T_2 segments of spinal cord and enter paravertebral sympathetic chain via ventral roots to synapse with the cells in *superior cervical ganglion*.
- *Post-ganglionic fibres* run along the carotid artery branches and are supplied to the three pairs of salivary glands along with their blood supply.

Effects of sympathetic stimulation

Stimulation of sympathetic fibres causes: vasoconstriction in the salivary glands and transient secretion of a very small amount of thick viscid saliva, rich in mucus and other organic constituents.

Paralytic secretion

Claude Bernard observed that cutting the chorda tympani nerve (parasympathetic) in dog or cat produces scanty secretion of thin turbid saliva, which increases to peak on seventh day and diminishes in 3 weeks. He called it as paralytic secretion because it was caused by cutting the nerve supply. However, later on, it was shown that the increased secretion is due to an increased sensitivity of the gland to adrenaline after cutting the chorda tympani nerve.

FUNCTIONS OF SALIVA

1. Protective function

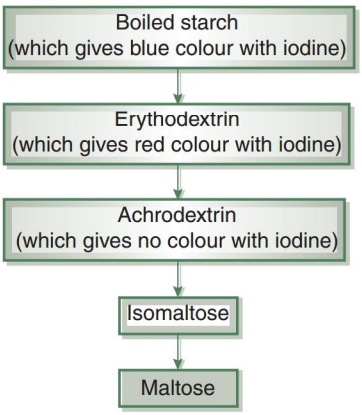
- *Dilutes hot and irritant* food substances thus preventing injury to buccal mucosa.
- *Washes away food particles* that remain in the oral cavity at the end of meal and thus cleans the oral cavity, i.e. helps in maintaining oral hygiene. In this way, growth of several harmful bacteria in the oral cavity is prevented.
- *Destroys harmful bacteria* in the mouth and thus minimizes the risk of buccal infection and dental caries because of its following constituents:
  - *Lysozymes*, which have bactericidal action,
  - *IgA*, which provides immunological defence against bacteria and viruses,
  - *Lactoferrin*, which has bacteriostatic action, i.e. prevents multiplication of bacteria.
- *Dilutes any hydrochloric acid (HCl) and bile*, which regurgitate into oesophagus and mouth.

2. Role in mastication and deglutition

- Salivary mucus lubricates the food and buccal mucosa and thus aids in mastication and swallowing.
- Helps bolus formation by acting as a glue.

3. Digestive functions

- *Initial starch* digestion starts by α-amylase (ptyalin) present in the saliva. Salivary amylase is an alpha amylase, therefore acts on alpha 1–4 linkage (but not on alpha 1–6 linkage) and digests starch to maltose in the following way:



- However, the role of salivary amylase in the digestion of polysaccharides is limited by the short duration of salivary action. When the bolus of food reaches the stomach and mixes with the gastric juice, the gastric acidity (pH1) stops the action of salivary amylase (which acts at an optimum pH of 6.5–7).
- *Initial triglyceride digestion* is caused by the lingual lipase present in the saliva.

4. Role in taste sensation

Saliva acts as a solvent for various foodstuffs. As taste is a chemical sense, the taste receptors respond only to the dissolved substances.

5. Role in speech

Salivary mucus lubricates the oral mucosa and thus aids speech by the facilitating movements of lips and tongue.

6. Excretory function

Saliva acts as a vehicle for excretion of certain heavy metals, thiocyanate ions, alcohol and morphine.

7. Role in temperature regulation

- During state of dehydration, the salivary secretion is reduced which induces thirst.
- Panting mechanism. In dogs, saliva is evaporated from the surface of tongue to cause evaporative heat loss.

DEGLUTITION (SWALLOWING)

PHASES OF SWALLOWING

Deglutition or swallowing refers to the passage of food from the oral cavity into the stomach. It comprises three phases:

- Oral phase (voluntary),
- Pharyngeal phase (reflex or involuntary) and
- Oesophageal phase (reflex or involuntary).

ORAL PHASE

- Oral phase or the first stage of swallowing is a voluntary phase.
- During this phase, the bolus of food formed after mastication is put over the dorsum of tongue. The tongue forces the bolus into the oropharynx by pushing up and back against the hard palate (Fig. 7.2-5A).

PHARYNGEAL PHASE

Pharyngeal phase or second stage of swallowing is an involuntary phase caused by a swallowing reflex.

Components of swallowing reflex (Fig. 7.2-6)

- *Receptors* present around the opening of pharynx (especially over tonsillar pillars) are stimulated when bolus moves from the mouth into the pharynx and initiate the reflex activity.
- *Afferent* arc that carries impulses from the receptors to the deglutition centre comprises the trigeminal, glossopharyngeal and vagus nerve.
- *Deglutition centre* co-ordinating the reflex activity is located in the medulla oblongata and lower pons (i.e. in the nucleus of the tractus solitarius and the nucleus ambiguus).
- *Efferent arc*, which initiates a series of muscular contractions, reaches the pharyngeal musculature and tongue through the 5th, 9th, 10th and 12th cranial nerves.

Events during pharyngeal phase

Events which take place during movement of bolus from the pharynx into the oesophagus occur in following sequence (Fig. 7.2-5B, C and D):

- *Oral cavity is shut off* from the pharynx by the approximation of posterior pillars of the fauces.
- *Nasopharynx* is closed by the upward movement of soft palate, preventing regurgitation of food into the nasal cavities.
- *Palatopharyngeal folds are pulled medially*, to make a slit-like opening for food, allowing only properly masticated food to pass through (selective action).
- *Vocal cords* strongly approximate stopping the breathing temporarily (*deglutition apnoea*), *larynx* is pulled upward and anteriorly by neck muscles enlarging the opening of oesophagus, which is normally a slit and *epiglottis* swings backwards to close laryngeal opening. All this guides the food towards the oesophagus and prevent its entry into the trachea.
- *Upper oesophageal sphincter (UES)* which normally remains contracted tonically opens up and allows the bolus of food to be pushed into the upper part of oesophagus by the rapid peristaltic contraction wave of pharynx which also continues in the oesophagus.
- Once the bolus of food has passed into the oesophagus, cricopharyngeus contracts, vocal cords open up allowing normal breathing to be resumed and the UES once again goes into tonic contraction.

The entire process of pharyngeal phase is completed in 1–2 seconds.

OESOPHAGEAL PHASE

During oesophageal phase, the food bolus is propelled from the upper part of oesophagus to the stomach by the oesophageal peristalsis and aided by gravity. Before describing the features of oesophageal peristalsis, it will be worthwhile to discuss briefly the applied anatomy of oesophagus (Fig. 7.2-7).

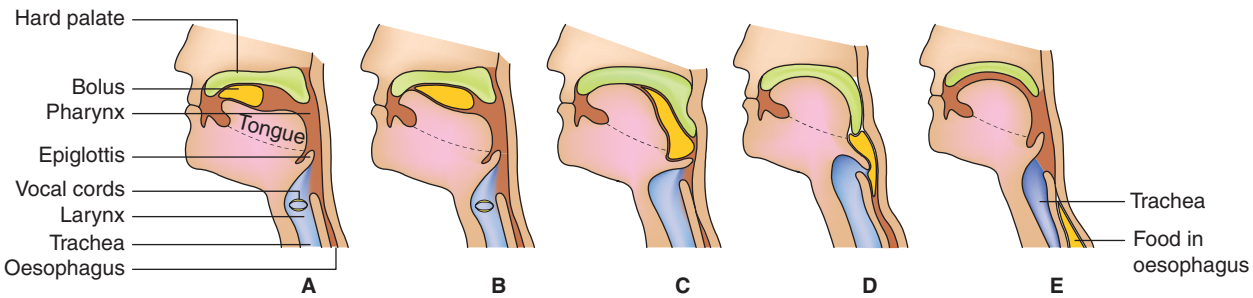


Fig. 7.2-5 Phases of swallowing: A, oral phase; B, C, D, early, middle and late pharyngeal phase; and E, oesophageal phase.

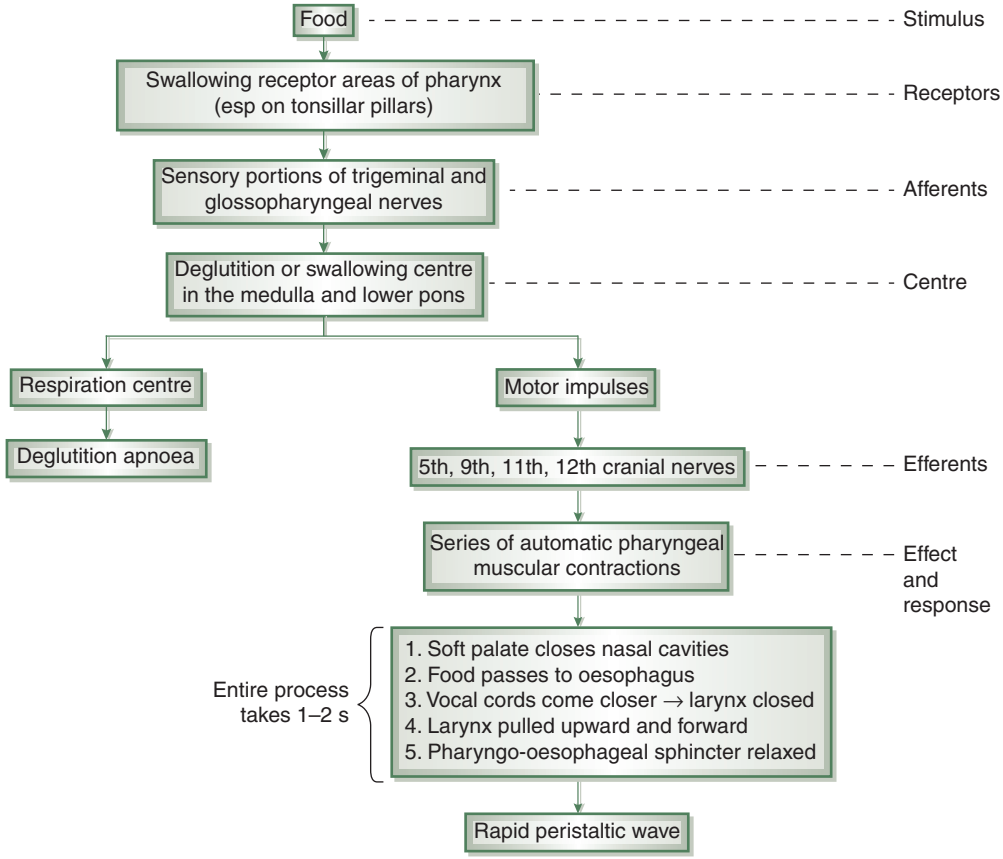


Fig. 7.2-6 Summary of swallowing reflex.

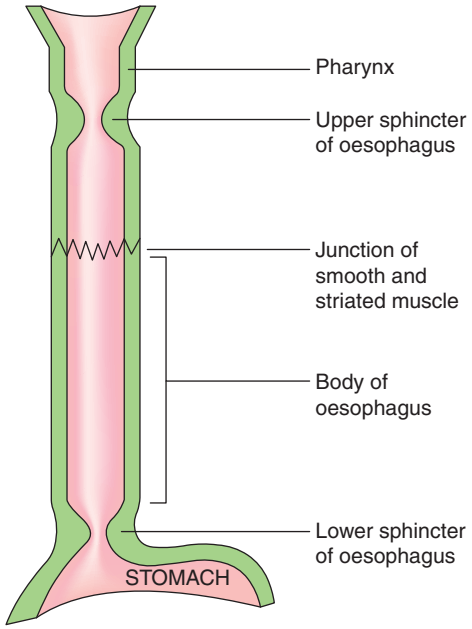


Fig. 7.2-7 Schematic structure of oesophagus.

Applied anatomy of oesophagus

Oesophagus is a fibromuscular tube about 25 cm long. It is separated from the pharynx by the UES and from the stomach by the lower oesophageal sphincter (LES).

Musculature of oesophagus

- *Upper one-third of oesophagus*, including the upper oesophageal sphincter is made up of striated muscle that is under the control of vagal fibres emerging from the nucleus ambiguus.
- *Lower two-thirds of oesophagus*, including the lower oesophageal sphincter is composed of smooth muscle. Its activity is regulated by vagal fibres originating within the *dorsal motor nucleus*. These fibres innervate intrinsic neurons within the muscle layers of oesophagus that release an inhibitory neurotransmitter (either *vasoactive intestinal peptide*—VIP, or nitric oxide).

Upper oesophageal sphincter. UES is a true sphincter formed by the cricopharyngeal muscle. The UES is normally contracted tonically and serves to prevent the entry of air into the oesophagus during normal respiration. Its tone is maintained by the continual firing of vagal fibres originating from the nucleus ambiguus. The neurotransmitter released by these fibres is acetylcholine (ACh).

The UES opens during swallowing when a rapid peristaltic wave starting in the pharyngeal muscles passes on to oesophagus.

Lower oesophageal sphincter. Lower oesophageal sphincter also known as cardiac sphincter refers to distal 2 cm of

oesophagus. Its contractile characteristics are quite different from the rest of oesophageal smooth muscle (that is why it is called physiological sphincter).

The principal function of LES is to prevent regurgitation of gastric contents (food, gastric juice and air) into the oesophagus. When the intragastric pressure is markedly raised (e.g. after a heavy meal or ingestion of carbonated drinks), the resistance of LES is overcome and air escapes into the mouth (belching). The local hormone, gastrin, increases the tone of LES and helps to keep the sphincter more tightly closed during digestion.

Oesophageal peristalsis

- The oesophageal phase of deglutition (Fig. 7.2-5E) is completed by two types of oesophageal peristalsis, primary and secondary.

Primary oesophageal peristalsis

- Primary oesophageal peristalsis is initiated by swallowing, i.e. is a part of swallowing and is thus co-ordinated by vagal fibres emerging from swallowing centre.
- As soon as the food bolus enters the oesophagus from pharynx, the UES contracts to prevent regurgitation of food into the mouth, and primary oesophageal peristalsis begins which propel the food downwards.
- The LES (which normally remains tonically contracted) relaxes as the peristaltic wave approaches the sphincter and allows the bolus of food to enter the stomach without causing any resistance.

Secondary oesophageal peristalsis

- When the primary oesophageal peristalsis is not able to push a bolus of solid food all the way down the oesophagus, the food remaining in the oesophagus stretches mechanical receptors and initiates another peristaltic wave called the secondary oesophageal peristalsis.
- Secondary oesophageal peristalsis is co-ordinated by the *intrinsic nervous system* of the oesophagus.

Note. After the food enters the stomach, the LES contracts to prevent regurgitation of food into the oesophagus. With this the oesophageal phase of deglutition is completed.

DISORDERS OF SWALLOWING

1. Abolition of deglutition reflex

Abolition of deglutition reflex causes regurgitation of food into the nose or aspiration into the larynx and trachea. It may occur:

- When IXth or Xth nerve is paralysed in lesions of medulla and
- When pharynx is anaesthetized with cocaine (deglutition reflex is abolished temporarily).

2. Aerophagia

- Aerophagia refers to the unavoidable swallowing of air along with the swallowing of food bolus and liquids.
- It usually occurs in nervous individuals having low tone of the UES.
- Some of the gases present in the air swallowed are absorbed, partly the air is regurgitated into the oral cavity and out in the atmosphere (belching), and majority of it passes on the colon and is then expelled as flatus through the anus.

3. Dysphagia

Dysphagia is a term used to denote difficulty in swallowing due to any cause.

4. Cardiac achalasia

- Cardiac achalasia is a neuromuscular disorder of the lower two-thirds of oesophagus, characterized by absence of oesophageal peristalsis and failure of the LES to relax during swallowing.
- Because of this, food transmission to stomach is impeded. In severe cases, oesophagus fails to empty the swallowed food into stomach for several hours.
- Over months and years, oesophagus becomes enlarged and infected due to long standing stasis of food.

5. Gastroesophageal reflux disease

Gastroesophageal reflux disease refers to a condition in which incompetence of the LES causes reflux of acidic gastric contents into the oesophagus. Reflux of stomach acid causes oesophageal pain (heartburn) and may lead to irritation of oesophagus or bronchioles (due to aspiration).

Chapter 7.3

Physiological Activities in Stomach

FUNCTIONAL ANATOMY
- Gross anatomy
- Structural characteristics
- Innervation of stomach

PHYSIOLOGY OF GASTRIC SECRETION
- **Gastric juice**
  - Composition of gastric juice
  - Secretion of gastric juice
- **Regulation of gastric secretion**
  - Neural control
  - Chemical control
- **Phases of gastric secretion and their regulation**
- **Experimental demonstration of regulation of gastric secretion**

PHYSIOLOGY OF GASTRIC MOTILITY
- **General considerations**
- **Initiation of gastric motility**
- **Types of gastric motility**
  - Motility of empty stomach
  - Gastric motility related to meals
    - Receptive relaxation and accomodation
    - Mixing peristaltic waves
    - Gastric emptying

FUNCTIONS OF STOMACH
- Mechanical functions
- Digestive functions
- Absorptive function
- Excretory function
- Stimulating functions
- Reflex functions
- Antiseptic function

APPLIED ASPECTS
- **Gastric mucosal barrier and pathophysiology of peptic ulcer**
- **Physiology of vomiting**
- **Total gastrectomy**
- **Gastric function tests**

FUNCTIONAL ANATOMY

GROSS ANATOMY

General features

- Stomach is a J-shaped hollow muscular bag connected to the oesophagus at its upper end and to the duodenum at the lower end.
- The *volume* of stomach is 1200–1500 mL, but its *capacity* is greater than 3000 mL.
- The stomach has two curvatures. The concavity of the right inner curve is called *lesser curvature* and the convexity of the left outer curve is the *greater curvature.* An angle along the lesser curvature is called the *incisura angularis.*

Parts of stomach

The stomach can be divided into five anatomic regions (Fig. 7.3-1):

- Cardia is the narrow conical portion of the stomach immediately distal to the gastroesophageal junction.

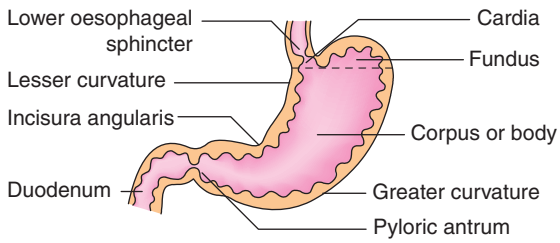


Fig. 7.3-1 Gross anatomy of stomach.

- *Fundus* is the dome-shaped proximal portion of the stomach.
- *Body or corpus* is the main part of the stomach that extends up to the incisura angularis.
- *Pyloric antrum* extends from the incisura angularis to the pyloric canal.
- *Pyloric canal or pylorus* is the distal most 1 in long tubular part of stomach.

Note. Anatomically, the antrum and pylorus are continuous and respond to nervous control as a unit. Functionally, first part of duodenum is associated with the pyloric part of stomach.

STRUCTURAL CHARACTERISTICS

The gastric wall consists of mucosa, submucosa, muscular coat and serosa (serous layer).

Gastric mucosa

Gross features

- The inner surface of the stomach exhibits coarse *rugae.* These infoldings of mucosa and submucosa are most prominent in the proximal stomach.
- The delicate texture of the mucosa is punctured by millions of gastric foveolae or pits, leading to the mucosal glands.

Histological features

Gastric mucosa comprises (Fig. 7.3-2):

Surface foveolar cells are tall columnar mucin secreting cells, which line the entire gastric mucosa as well as the gastric pits. These cells have basal nuclei and mucin-containing granules in the supranuclear region.

Mucous neck cells are present deeper in the gastric pits. These cells are thought to be the progenitors of both, the surface epithelium and the cells of gastric glands.

Glandular cells form the gastric glands. There are three types of gastric glands, main gastric glands, cardiac tubular glands and pyloric (antral) glands.

1. Main gastric glands are found in the body and fundus of stomach. These are simple tubular glands (Fig. 7.3-2). The alveoli of main gastric glands contain two types of cells:

- *Chief cells,* also known as peptic or zymogen cells, are basophilic. These cells are concentrated at the base of main gastric glands and secrete proteolytic proenzymes, pepsinogen I and II.
- *Parietal cells* (oxyntic cells) are acidophilic. These cells line predominantly the upper half of glands and have an extensive intracellular canalicular system. These secrete *hydrochloric acid* (HCl) and the *intrinsic factor.*

2. Cardiac tubular glands are found in the mucosa of cardia (a small conical part of the stomach), just around the distal end of oesophagus. These secrete soluble mucus.

3. Pyloric (antral) glands are found in the antrum and pylorus region of the stomach. These glands contain two types of cells:

- *Mucus cells,* which secrete *soluble mucus* and
- *G-Cells* are responsible for the release of the hormone gastrin.

Submucosa

The submucus coat consists of loose areolar connective tissue connecting the muscular and mucus coat of the stomach.

Musculature of stomach

Characteristic features of gastric musculature are:

- The muscle coat of stomach has three layers: an outer longitudinal, middle circular and an inner oblique (Fig. 7.3-3).
- The stomach and duodenum are divided by a thickened circular smooth muscle layer called *pyloric sphincter.*

Serosa

The serous coat of the stomach is part of peritoneum which covers the organ.

INNERVATION OF STOMACH

Innervation of stomach, as elsewhere in the gut (page 452), includes an intrinsic and an extrinsic system.

1. Intrinsic innervation comprises two interconnected plexuses:

- *Myenteric (Auerbach's) plexus* located between the layers of circular and longitudinal muscles of the stomach and
- *Submucosal (Meissner's) plexus* located in the submucosal layer.

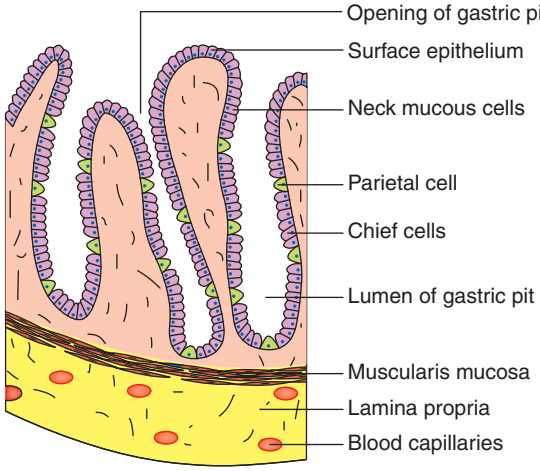


Fig. 7.3-2 Histological features of gastric mucosa.

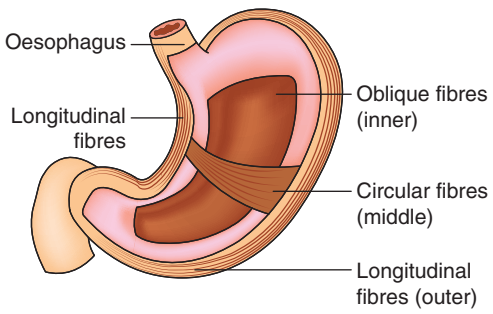


Fig. 7.3-3 Three layers of gastric musculature.

The intrinsic innervation is directly responsible for peristalsis and other contractions. Because this system is continuous between the stomach and duodenum, peristalsis in the antrum influences the duodenal bulb.

2. Extrinsic innervation modifies the co-ordinated motor activity that arises independently in the intrinsic nervous system. It consists of the two components of the autonomic nervous system:

- *Sympathetic innervation* comes via the coeliac plexus and *inhibits motility* and
- *Parasympathetic innervation* comes via the vagus nerve and *stimulates motility.*

PHYSIOLOGY OF GASTRIC SECRETION

The gastric secretions include:

- *Exocrine secretions,* i.e. gastric juice and
- *Endocrine secretions,* i.e. gastrin hormone (see regulation of gastric secretion, page 468).

GASTRIC JUICE

COMPOSITION OF GASTRIC JUICE

Gastric glands secrete about 2–2.5 L of gastric juice in the lumen of stomach per day. It is acidic with a pH varying from 1 to 2. Important constituents of gastric juice are:

Water – 99.45%
Solids – 0.55%, which include:

Electrolytes, such as Na^+, K^+, Mg^{2+}, Cl^-, HCO_3^-, HPO_4^- and SO_4^-. The electrolyte content of gastric juice varies with the rate of secretions. At low secretory rates, Na^+ concentration is high and H^+ concentration is low, but as acid secretion increases Na^+ concentration falls.

Enzymes present in the gastric juice are:

- *Pepsin* is a proteolytic enzyme, which is secreted by the chief cells of gastric glands in an inactive form pepsinogen.
- *Gastric lipase* is a weak fat splitting enzyme. It is of little importance in fat digestion except in pancreatic insufficiency.
- *Gastric gelatinase* liquefies gelatin, a protein contained in the connective tissue.
- *Gastric amylase* is present in small amounts.
- *Lysozyme* is bactericidal.
- *Carbonic anhydrase* is present in small amounts.
- *Urease* hydrolyses urea to produce ammonia.

Mucin or mucus is of two types:

- *Soluble mucus* secreted by the mucus cells of pyloric and cardiac glands, and
- *Insoluble mucus* secreted by the surface foveolar cells (tall columnar mucin secreting cells) lining the entire gastric mucosa.

Intrinsic factor is secreted by the parietal cells of gastric glands.

SECRETION OF GASTRIC JUICE

Secretion of HCl

General consideration

- Hydrochloric acid (HCl) is secreted by the *parietal cells* (also called oxyntic cells). These cells show, under electron microscope, a complex network of *intracellular canaliculi* into which HCl is secreted.
- Gastric glands secrete about 2.5 L of HCl in a day having a pH of approximately 1.0.
- At high rates of secretion, H^+ concentration may be as high as 155 mEq/L, i.e. about three million times greater than its concentration in the blood.

Mechanism of HCl secretion

Various theories have been put forward to explain the origin of H^+ of HCl. The hypothesis more widely accepted is shown in Fig. 7.3-4. Hydrochloric acid is made up of hydrogen (H^+) and chloride ions (Cl^-), therefore its secretion can be described in two steps, i.e. secretion of H^+ and secretion of Cl^-.

Secretion of H^+

- The H^+ ions are believed to be generated inside the parietal cell from metabolic CO_2 and H_2O present in the cell. The enzyme carbonic anhydrase present in abundance in the parietal cells is essential for the secretion.

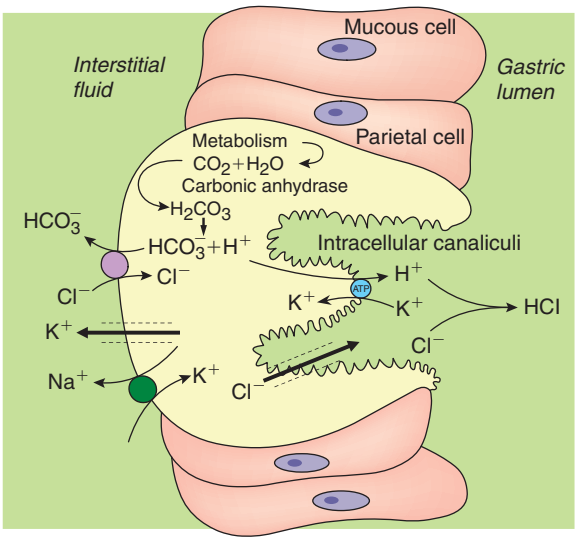


Fig. 7.3-4 Mechanism of HCl secretion in the parietal cells of stomach.

It accelerates the formation of $H_2CO_3^-$ which dissociates to release H^+ and HCO_3^- as

$$CO_2 + H_2O \xrightarrow{\text{Carbonic anhydrase}} H_2CO_3 \rightarrow H^+ + HCO_3^-$$

- The H^+ ions generated by the above reaction are then secreted into the lumen of the canaliculi in exchange for K^+ by a primary active transport mediated by a H^+–K^+–ATPase pump (Fig. 7.3-4).

  Note. The drug omeprazole used to decrease HCl formation inhibits the H^+–K^+–ATPase and blocks H^+ secretion.

- The HCO_3^- ions produced in the parietal cell are transported by an antiport in the serosal (basolateral) membrane into the blood in exchange of Cl^- by an active transport.

The HCO_3^- released into the blood is responsible for the post-prandial alkaline tide associated with an increased gastric acid secretion after meals, which is characterized by alkaline urine, slightly depressed breathing and raised alveolar pCO_2.

Secretion of Cl^-. Because of the high intracellular negativity, the Cl^- present in the parietal cell is forced out into the lumen of gland through the Cl^- channels located on the apical membrane of the cell. These Cl^- channels are activated by cAMP. The high intracellular negativity is the result of following (Fig. 7.3-4):

- The Na^+–K^+ pump located on the basolateral membrane of parietal cell pumps out three Na^+ for every two K^+ pumped in, thereby creating intracellular negativity.
- The K^+ pumped in diffuses out through the K^+ channels present on the basolateral as well as on the apical membranes. This diffusion further increases intracellular negativity of parietal cells.

Note. It is important to note that the active transport processes involved in the generation of HCl require a large amount of adenosine triphosphate (ATP). The ATP is generated by mitochondria found in very high concentration (40% of cell volume) within the parietal cell (Fig. 7.3-5).

Factors affecting HCl secretion

Factors stimulating HCl secretion are:

- Vagal stimulation
- Gastrin and
- Histamine.

Factors that inhibit HCl secretion are:

- Low pH in stomach (< 3) by negative feedback mechanism,
- Intestinal influences,
- Somatostatin and
- Prostaglandins (PGE and PGI), epidermal growth factor and transforming growth factor.

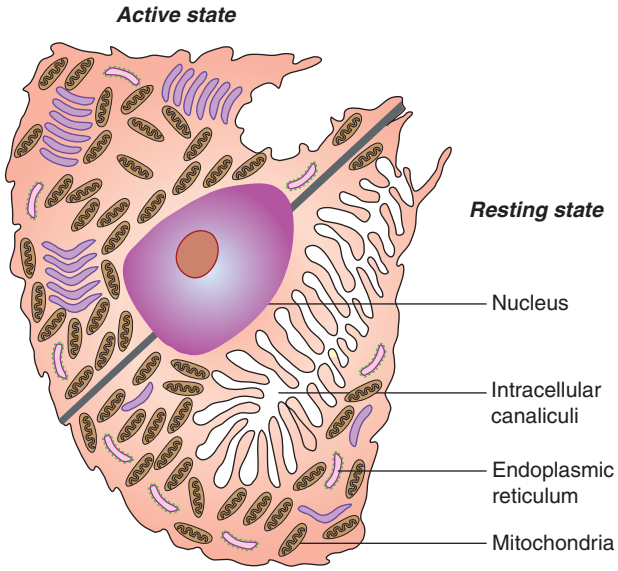


Fig. 7.3-5 Partial cell showing secretion of H^+ and Cl^- in the intracellular canaliculi which pour HCl into stomach. Note the presence of numerous mitochondria which provide energy for the active transport process.

Functions of HCl

- HCl participates in the breakdown of protein.
- It provides an optimal pH for the action of pepsin.
- It hinders the growth of pathogenic bacteria.

Pepsinogen secretion

- Pepsinogen is an inactive precursor (proenzyme) of pepsin. It is mainly secreted by the *chief cells* of the main gastric glands. A small amount of pepsinogen is also secreted by the pyloric glands. The pepsinogen secreted by the chief cells is called *pepsinogen I* and that secreted by the pyloric glands is called *pepsinogen II.*
- Pepsinogen is synthesized and stored as zymogen granules in the apical region of the chief cells.
- Pepsinogen secretion is stimulated by vagal stimulation, gastrin and histamine.
- Pepsinogen is converted to pepsin (the active form) by the action of HCl or preformed pepsin.

$$\text{Pepsinogen} \xrightarrow{\text{HCl}} \text{Pepsin}$$

$$\text{Pepsinogen} \xrightarrow{\text{Pepsin}} \text{Pepsin}$$

Function of pepsinogen

Pepsin, the active form of pepsinogen, is a proteolytic enzyme that begins the process of protein digestion. It splits protein into proteoses, peptones and polypeptides. It is important to note that the optimum pH for the action of pepsin is 2.0, therefore, acid secretion by the stomach is as essential as pepsinogen secretion for the digestion of proteins.

Pepsin acts on water-soluble caseinogens (milk protein) to form casein, which combines with calcium to form insoluble

calcium caseinate (curdling of milk). In other mammals, this function is carried out by the renin present in the gastric juice.

Secretion of mucus

Mucus is of two types, insoluble and soluble.

Insoluble mucus is secreted by the mucous secreting cells lining the entire gastric mucosa. The insoluble mucus is such a viscid that it forms a gel-like coat over the mucosa. These cells also secrete bicarbonate ions which make the mucus with alkaline pH of 7 that forms an extremely important protective layer saving the stomach from the destruction of HCl.

Soluble mucus is secreted by the mucus cells of pylorus and cardiac glands.

Mucus secretion is increased by direct stimulation of mucosa by the rough food. Neural or hormonal control over secretion of mucus, if any, is not known.

Secretion of intrinsic factor

Intrinsic factor (IF), a glycoprotein, is secreted by the parietal cells of gastric mucosa, chiefly by those in the fundus.

Functions. The intrinsic factor is essential for the absorption of Vitamin B_{12}. It forms a complex with B_{12} which is carried to the terminal ileum where the vitamin is absorbed.

Deficiency of intrinsic factor in some patients with idiopathic atrophy of gastric mucosa may cause a serious disorder called pernicious anaemia (see page 119).

REGULATION OF GASTRIC SECRETION

Mechanisms regulating the gastric secretion include neural control and chemical control.

NEURAL CONTROL

Neural control over the gastric glands is exerted by a local enteric plexus involving cholinergic neurons and impulses from the central nervous system via vagal (extrinsic) innervation.

Vagal stimulation increases the secretion of HCl by the parietal cells and pepsin by the chief cells. Vagal stimulation increases H^+ secretion by a direct path and an indirect path (Fig. 7.3-6):

- In the *direct path,* the vagus nerve fibres innervating the parietal cells stimulate H^+ secretion by releasing neurotransmitter acetylcholine (ACh), which acts on the *muscarinic receptors* on the parietal cells. In addition, ACh also potentiates the effects of histamine on H_2 receptors of parietal cells.

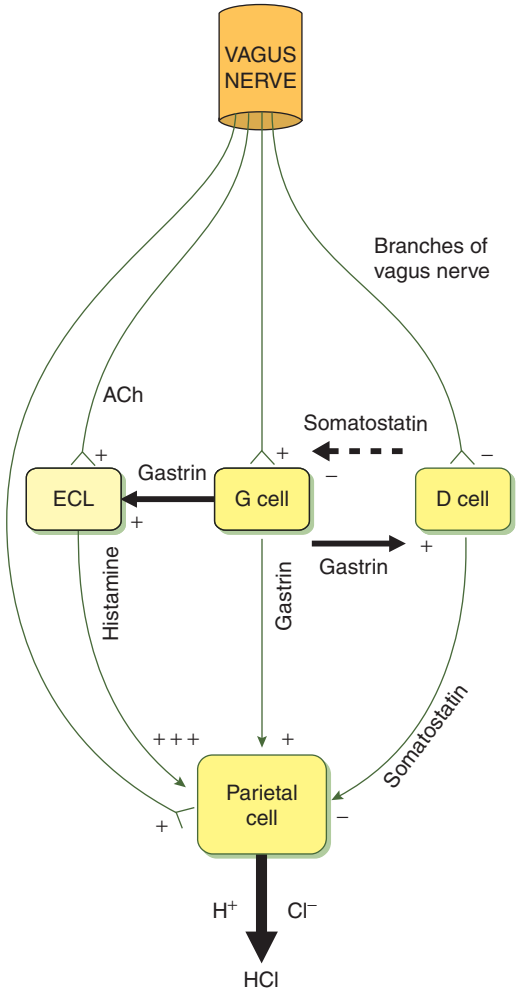


Fig. 7.3-6 Mechanisms by which vagal stimulation increase H^+ secretion in the parietal cell.

- In the *indirect path,* the vagus nerve innervates G cells and stimulates the release of gastrin into circulation through gastrin-releasing peptide (GRP). The gastrin in turn stimulates H^+ secretion.
- Further, vagal stimulation also inhibits the release of *somatostatin* and thus indirectly stimulates H^+ secretion by removing the inhibitory effect of somatostatin on the parietal cells.
- ACh released on the vagal stimulation also acts on the enterochromaffin-like (ECL) cells, which release histamine. Histamine increases H^+ secretion by acting on H_2 receptors on the parietal cells.

CHEMICAL CONTROL

Chemical control on the gastric glands is exerted mainly through:

1. Role of gastrin

Gastrin, a hormone, is secreted by the G cells into the blood circulation (and not into gastric juice). It reaches the

stomach through the arterial circulation and stimulates secretory activity of the parietal cells and chief cells.

G cells, or the gastrin-secreting cells, located at the base of gastric glands are specially abundant in the pyloric glands. The G cells are also called APUD cells, as these are also responsible for the amine precursor uptake and decarboxylation. G cells are flask shaped with a broad base and contain many gastrin granules near the base. The gastrin granules disappear after feeding, leaving the empty vacuoles.

Types of gastrin. Gastrin is secreted in an inactive form, the progastrin, which gets converted to active form (gastrin) by HCl or product of digestion. There are three types of gastrin, namely G-34, G-17 and G-14 (depending upon the number of amino acids). G-17, containing 17 amino acids, is the principal product concerned with gastric acid secretion. It has a half-life of 2–3 min in the circulation and is inactivated mainly in the kidney and small intestine.

Functions of gastrin

- The main function of gastrin is to stimulate HCl secretion by binding with the cholecystokinin (CCK)-β gastrin receptors on the parietal cell. It acts to open Ca^{2+} channels and to release Ca^{2+} from the intracellular stores. The Ca^{2+} along with cyclic AMP act via protein kinase to increase the transport of H^+ into the gastric lumen by H^+–K^+–ATPase (Fig. 7.3-7).
- Gastrin also increases the HCl secretion by stimulating secretion of histamine from the ECL cells present in the body of the stomach.

Other functions are:

- Increases gastric and intestinal motility.
- Increases pancreatic secretion of insulin and glucagon.
- Trophic action, i.e. it is necessary for the proper growth of gastrointestinal mucosa.

Factors affecting gastrin secretion

Factors that stimulate gastrin secretion

- *Vagal stimulation.* Vagal stimulation increases gastrin release through GRP and not through release of neurotransmitter ACh. Because of this reason, atropine (which blocks muscarinic receptors) does not affect release of gastrin from G cells.
- *Distension of the pyloric antrum* increases gastrin secretion through the intrinsic innervation. This fact can be proved experimentally by the Heidenhain pouch (see page 472).
- *Products of protein digestion* (e.g. peptides and amino acids), alcohol and coffee also increase gastrin secretion.
- *Calcium and epinephrine* are also reported to increase gastrin secretion.

Factors that inhibit gastrin secretion

- *Low pH of gastric juice* (<3) inhibits gastrin secretion and forms the basis of negative feedback mechanism controlling HCl release (see page 470).
- *Somatostatin,* released by D cells (located adjacent to G-cells or parietal cells in gastric mucosa) inhibits gastrin secretion by G-cells.

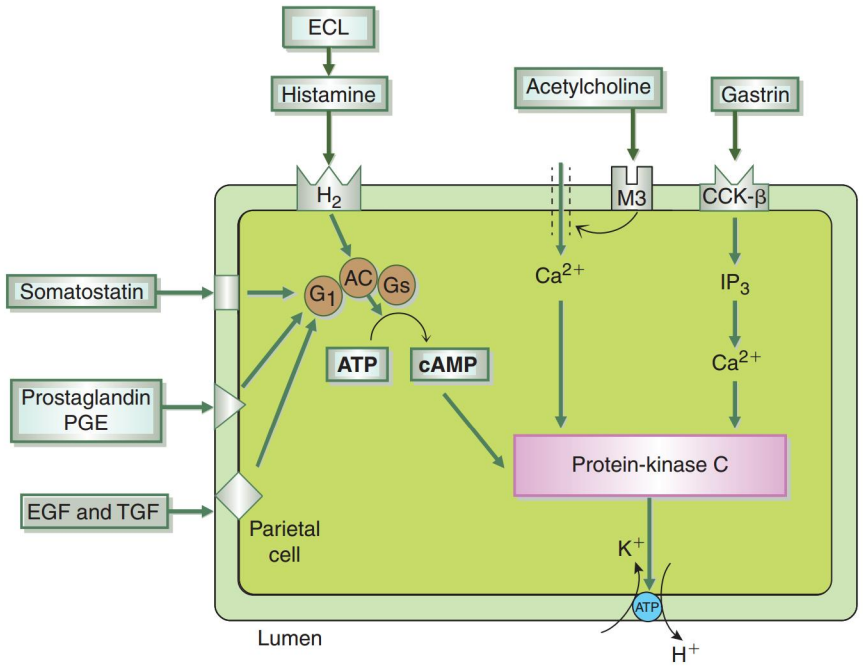


Fig. 7.3-7 Control of HCl secretion at the parietal cell level. Acetylcholine (ACh), histamine and gastrin act agonists of HCl secretion by distinct mechanism (for details see text). Somatostatin, prostaglandins (PGE), epidermal growth factor (EGF) and transforming growth factor (TGF) act as endogenous antagonists of HCl secretion by inhibiting adenylyl cylase.

- *Other substances* that inhibit gastrin secretion are secretin, gastric inhibitory peptide (GIP), vasoactive intestinal peptide (VIP), glucagon, and calcitonin.

2. Role of histamine

Histamine is released from the ECL cells found in the base of the gastric gland. ECL cells bear both gastrin receptors and ACh receptors. They release histamine in response to both circulating gastrin as well as ACh released by vagal fibres. The histamine released stimulates *HCl secretion* from the parietal cells by acting on H_2 receptors. The H_2 receptors increase intracellular cAMP via Gs. The cAMP acts as a second messenger to activate cAMP-dependent protein kinase. H_2 receptor-blocking drugs, such as *cimetidine* and *ranitidine*, inhibit H^+ secretion by blocking the stimulatory effect of histamine.

3. Role of somatostatin

- Somatostatin is secreted by D cells located adjacent to the G cells or the parietal cells in the gastric glands.
- Somatostatin inhibits HCl secretion in two ways:
  - Directly by its action on parietal cells,
  - Indirectly by inhibiting gastrin secretion by G cells.
- Somatostatin, prostaglandins, epidermal growth factor and transforming growth factor act on the parietal cells to inhibit HCl secretion by inhibiting adenylyl cyclase (Fig. 7.3-7).

4. Role of low pH (<3) in stomach

Low pH (<3) in the stomach inhibits the secretion of H^+ by the parietal cells by a negative feedback mechanism.

- When the pH of stomach contents is < 3.0, *gastrin secretion is inhibited*, which in turn inhibits H^+ secretion. This forms the so-called *negative feedback mechanism*.
- On the other hand, if the pH of gastric contents rises above 3.5 (due to buffering action of food), the release of gastrin is stimulated. In this way, the negative feedback control over gastrin release maintains the pH of gastric contents near 3 (Fig. 7.3-8).

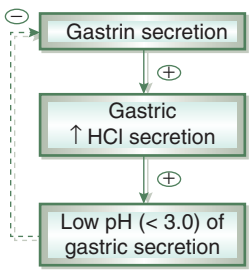


Fig. 7.3-8 Negative feedback mechanism to control gastrin secretion.

5. Intestinal influences

Chyme containing acid, fats and products of protein digestion when reaches the duodenum causes the release of several intestinal hormones like secretin, CCK and GIP.

PHASES OF GASTRIC SECRETION AND THEIR REGULATION

Meal-related gastric secretion can be divided into three phases:

- Cephalic phase,
- Gastric phase and
- Intestinal phase.

1. Cephalic phase

- Cephalic phase of the gastric secretion occurs before the entry of food into the stomach.
- The secretion is *initiated* by the thought, sight, smell or taste of food. Neurogenic signals originate in the cerebral cortex and appetite centres of amygdala or hypothalamus. The impulses are transmitted to dorsal vagal nuclei and from there through vagii to the stomach (Fig. 7.3-9).
- Emotions also influence this vagally mediated gastric secretion. *Anger and hostility* are associated with an increased gastric secretion and motility. *Fear and depression* decrease the gastric secretion and motility.
- Rate of gastric juice secretion during this phase is high, about 500 mL/h, but this phase lasts for a short time and accounts for about 45% of total gastric juice secretion during a meal.

2. Gastric phase

- Gastric phase of the gastric secretion occurs when food enters the stomach.
- Rate of gastric juice secretion during this phase is less as compared to that in the cephalic phase. But this phase lasts for a long time (as long as food remains in the stomach) and so accounts for about 50% the total gastric secretion.
- The presence of food in the stomach induces gastric secretion by following *mechanisms:*
  - *Distension of the body of stomach* acting through local myenteric and vagovagal reflexes, results in an increase in HCl secretion.
  - *Distension of the antrum* initiates vagally mediated and local reflexes that result in gastrin release from the antral G-cells. Gastrin release is inhibited when pH becomes low (<3).
  - *Products of partial protein digestion* also stimulate gastrin secretion and this increases mainly secretion of gastric acid.

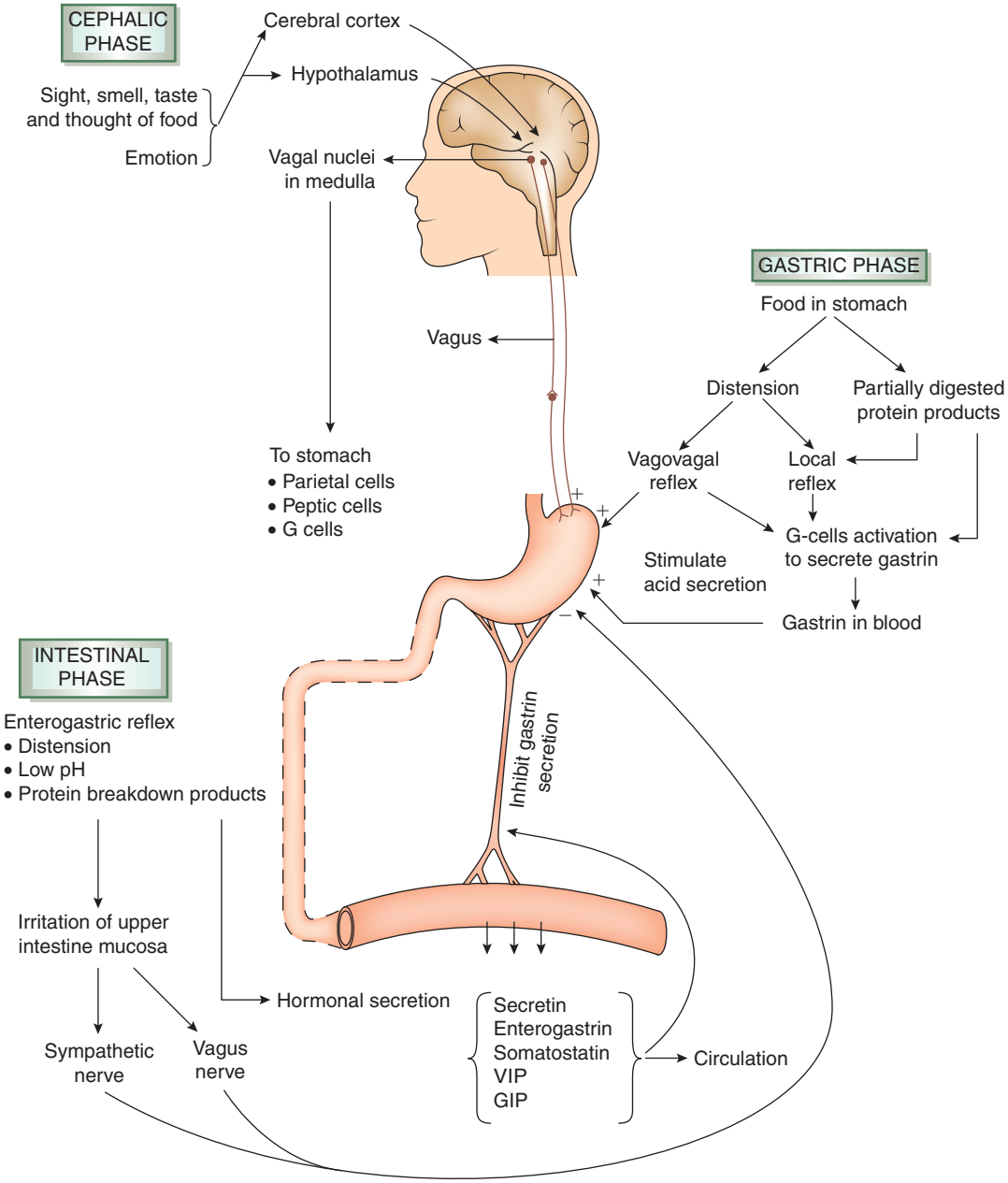


Fig. 7.3-9 Phases and regulation of gastric secretion.

- *Low pH* causes increased pepsinogen secretion through local reflexes.

3. Intestinal phase

- Intestinal phase of gastric secretion begins as the chyme begins to empty from the stomach into the duodenum.
- In contrast to the excitatory cephalic and gastric influences, the intestinal influence on the gastric secretion is chiefly inhibitory in nature. Intestinal factor inhibits gastric secretion by following mechanisms:
  - *Enterogastric reflex* is initiated by the distension of small intestine, presence of acid or protein breakdown products in the upper intestine and irritation of mucosa.
  - *Hormonal mechanism.* Presence of acid, fat, hyper- or hypotonic solution and irritating factors in the upper small intestine release several hormones, such as secretin, CCK, GIP, VIP and somatostatin which inhibit gastric secretion.
- The inhibitory influences discussed above help to terminate the gastric secretion when all the food has left stomach.

EXPERIMENTAL DEMONSTRATION OF REGULATION OF GASTRIC SECRETION

Phases and regulation of gastric secretion has been studied by certain experiments which are described briefly.

1. Sham feeding: an experiment to demonstrate cephalic phase of gastric secretion. Cephalic phase of gastric secretion, i.e. secretion of gastric juice before the entry of food can be demonstrated by the sham feeding experiment. For this, oesophagus of a dog is exposed and divided in the middle of neck and two cut ends are brought to the surface (Fig. 7.3-10). When dog swallows, the food comes out through the upper cut end of oesophagus and does not enter the stomach. Gastric secretion, which occurs before the entry of food into stomach (caused by sight, smell and taste of food), represents the *cephalic phase* and is collected for study by passing a tube in the stomach through the lower cut end of oesophagus.

Cephalic phase of the so-called appetite juice begins after a latency of 5–7 min.

2. Pavlov's pouch experiment to demonstrate that vagus is secretomotor nerve to stomach. For this, under general anaesthesia, a pouch of stomach with intact nerve and blood supply is separated from the body of stomach by incising the mucosa and keeping the muscle layer intact. The intactness of the larger main part of stomach is restored by applying sutures. An outlet is made in the smaller part (pouch) so separated and is brought out through the abdominal wall to provide drainage for the pouch secretion. The vagus nerve is exposed and divided in the neck and the animal is allowed to recover (Fig. 7.3-11).

After some days, the peripheral cut end of the vagus is stimulated in the unanaesthetized dog. Flow of gastric juice rich in HCl and pepsin after a short latent period demonstrates that vagus nerve is secretomotor to stomach.

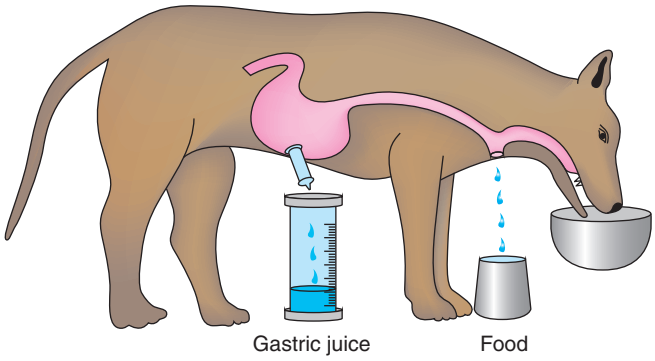


Fig. 7.3-10 Sham feeding experiment to demonstrate initiation of cephalic phase of gastric secretion.

3. Heidenhain's pouch experiment to demonstrate existence of some blood-borne mechanism regulating gastric secretion.

- Heidenhain's pouch is modified Pavlov's pouch which is separated in such a way from the antral part of stomach that it is denervated but with intact blood supply (Fig. 7.3-12).
- Distension of the denervated pouch (with intact blood supply) induces gastric secretion. Occurrence of gastric secretion in a denervated pouch of stomach demonstrates that there exists some blood-borne mechanism which also regulates gastric secretion.
- Intravenous injection of gastrin is followed after 5 min by secretion of gastric juice from the denervated Heidenhain's pouch. This demonstrates that blood-borne mechanism is mediated by a gastrin hormone released from the antral mucosa.

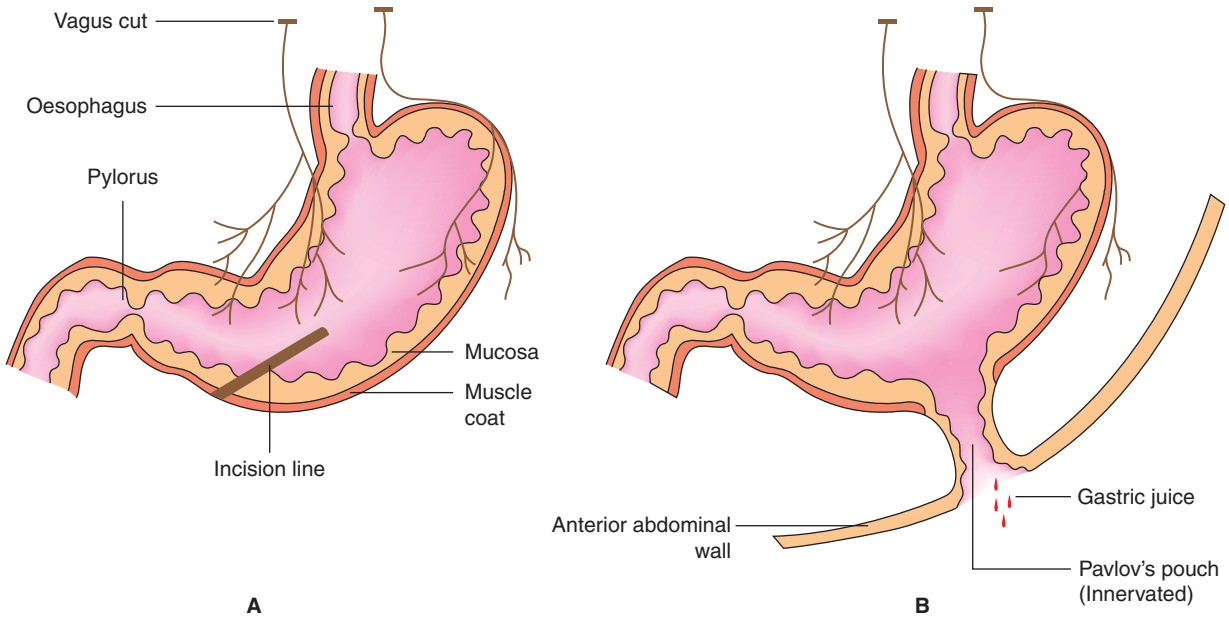


Fig. 7.3-11 Preparation of Pavlov's pouch: A, showing the site of incision and B, Pavlov's pouch opening outside through anterior abdominal wall.

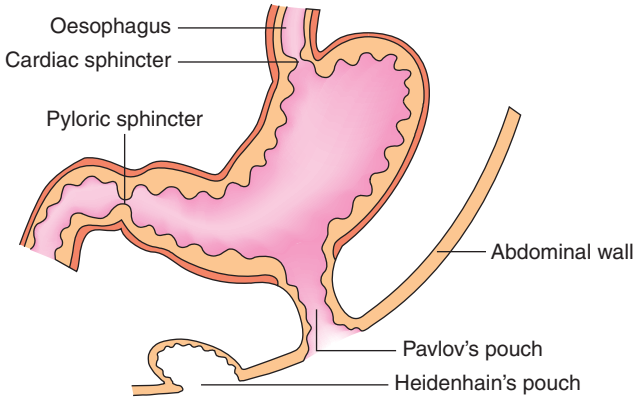


Fig. 7.3-12 Heidenhain's pouch.

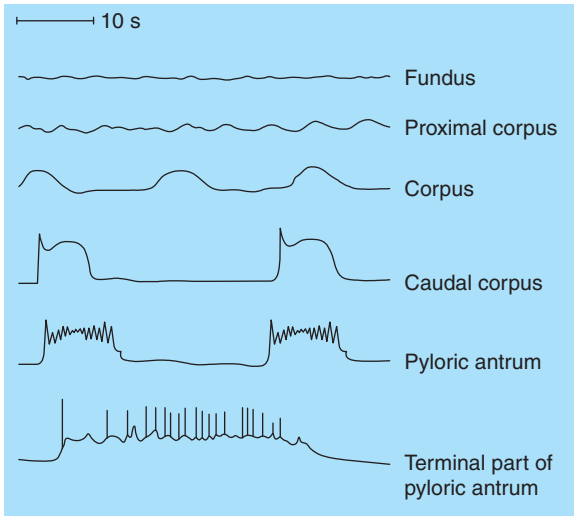


Fig. 7.3-13 Basic electrical rhythm (BER) recorded from different parts of the stomach.

PHYSIOLOGY OF GASTRIC MOTILITY

GENERAL CONSIDERATIONS

- Gastric motility is the function of gastric musculature which consists of three layers of smooth muscle fibres: an outer longitudinal layer, middle circular layer and an inner oblique layer.
- From the viewpoint of gastric contractions, the stomach can be divided into two regions:
  - *Oral region* of the stomach includes the fundus and proximal body. This region is responsible for receiving the ingested food.
  - *Caudal region* of the stomach includes the antrum and the distal part of body of stomach. This region is responsible for the contractions that mix food and propel it into the duodenum.

Motor functions of stomach observed by the gastric motility are:

- Storage of food,
- Mixing of food and
- Slow emptying of food.

INITIATION OF GASTRIC MOTILITY

BASAL ELECTRICAL RHYTHM

- The musculature of stomach, being a single unit smooth muscle, has its only rhythmic contractile myogenic tone due to the *basic electrical rhythm* (BER) or *gastric slow waves.* Thus, the BER represents a wave of depolarization of smooth muscle cells proceeding from the circular muscles of the fundus of stomach to the pyloric sphincter.
- The gastric slow waves are initiated *by the pacemaker cells* located near the fundus on the greater curvature of the stomach.
- Gastric slow waves consist of an *upstroke* and an *plateau* phase (Fig. 7.3-13) and occur at a rate of approximately 3–4 waves/min.
- Electrophysiological basis of BER is not entirely known, however, it is assumed that the upstroke is due to flow of Na^+ and Ca^{2+} into the cell and that the plateau is dependent primarily on the flow of Ca^{2+} into the cell.
- In the stomach, ACh increases contractile activity (produces peristalsis). Other agents that initiate contraction of smooth muscles of the stomach are gastrin, histamine, nicotine, barium and K^+. Agents that inhibit the activity are enterogastrone, epinephrine, norepinephrine, atropine and Ca^{2+}.

TYPES OF GASTRIC MOTILITY

The peristaltic activity of the gastric musculature has been given various names depending upon its features and motor function subserved by it. Gastric motility can be described as:

- Motility of the empty stomach, which includes:
  - Migrating motor complex and
  - Hunger contractions
- Gastric motility related to meal, includes:
  - Receptive relaxation,
  - Mixing peristaltic waves and
  - Gastric emptying.

MOTILITY OF EMPTY STOMACH

Migrating motor complexes

- Migrating motor complex (MMC) is the name given to the peristaltic wave that begins in the oesophagus and travels through the entire gastrointestinal tract (migratory motor activity) during interdigestive period.

- The MMCs remove any food remaining in the stomach and intestines during interdigestive period in the preparation for the next meal because of this they have been called the *interdigestive housekeepers.*
- The MMC wave travels at a regular rate (5 cm/min) and occurs every 60–90 min during the interdigestive period.
- There is a close correlation of the BER and MMC. When there are no MMC, the BER consists of rhythmic oscillation of the RMP between about –65 and –45 mV. During the MMC, the electrical oscillations are superimposed with spikes (Fig. 7.3-14).
- The hormone motilin, which is released from the endocrine cells within the epithelium of small intestine, increases the strength of MMC.
- The MMC are abolished immediately after the entry of food in the stomach.

Hunger contractions

Mild peristaltic contractions occur in the empty stomach, which over a period of hours increase in intensity and are called hunger contractions. Migrating motor complexes are probably responsible for hunger contractions. When they become extremely strong they fuse to cause *tetanic contraction* lasting for 2–3 min which can be felt and may even be painful. These are associated with sensation of hunger.

GASTRIC MOTILITY RELATED TO MEALS

Receptive relaxation and accommodation

- Storage function of stomach is accomplished by the receptive relaxation and accommodation (Fig. 7.3-15).

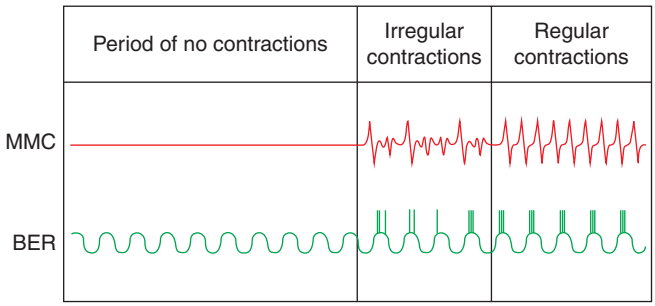


Fig. 7.3-14 Relation of basic electrical rhythm (BER) with migratory motor complexes (MMCs).

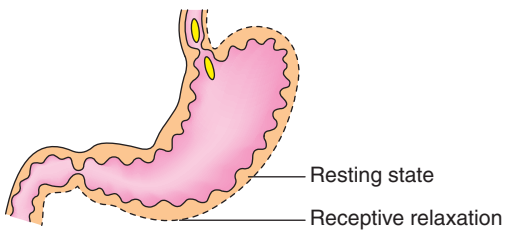


Fig. 7.3-15 Receptive relaxation of stomach.

- The passage of each bolus of food stimulates the stretch *receptors of oral region* and produces relaxation. By the end of meal about, 1–2 L of food can be accommodated.
- Receptive relaxation is a *vagovagal reflex* initiated by distension of stomach and is synchronized with the primary peristaltic waves in the oesophagus.
- Cholecystokinin participates in a receptive relaxation by increasing the distensibility of the oral stomach.
- The inhibitory neurotransmitter responsible for the receptive relaxation and accommodation is either VIP or NO.
- Vagotomy abolishes receptive relaxation.

Mixing peristaltic waves

The presence of food in the caudal region (distal body and antral part) of stomach increases the contractile activity of this part of stomach. This enhanced contractile activity (a combination of peristalsis and retropulsion) is called mixing waves, which mix the food with stomach acid and enzymes and break it into smaller and smaller pieces. When the food is mixed into a pasty consistency it is called *chyme.*

Initiation and production of peristalsis

Peristalsis is co-ordinated pattern of smooth muscle contraction and relaxation where wave of relaxation precedes wave of contraction. Peristaltic contractions are produced by the periodic changes in membrane potential (basal electrical rhythm, described earlier). The rhythmicity of gastric peristalsis is determined by the BER, which has a frequency of 3–4/min (slow waves). The number of spikes fired in a slow wave determines the force of each peristaltic contraction (Fig. 7.3-16).

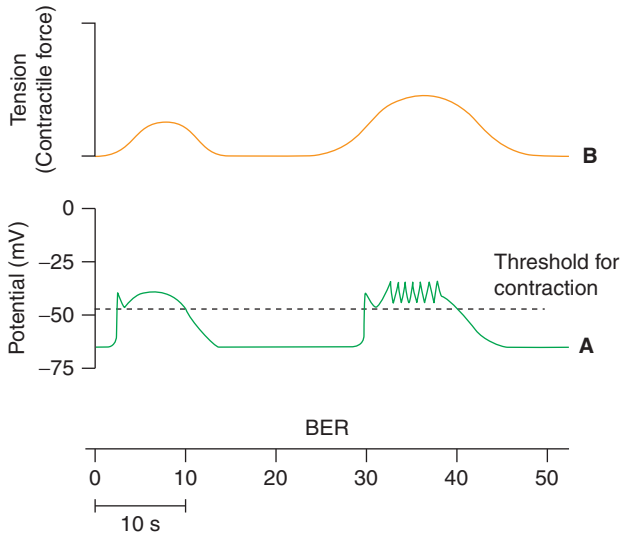


Fig. 7.3-16 Membrane potentials of smooth muscle of stomach (A) and their relation to mechanical response (B).

Mixing mechanism of peristalsis and retropulsion

- Peristaltic contractions begin in the mid stomach (Fig. 7.3-17A) and proceed caudally As the wave proceeds towards the pylorus it deepens. Thus, the peristaltic waves are most marked in the distal half of stomach (called antral systole).
- The food particles also move towards pylorus (Fig. 7.3-17B) along with the deep wave of contraction, but the wave of contraction reaches pyloric sphincter and causes its contraction before the food reaches there.
- When the food reaches the pylorus, it strikes against the closed pyloric sphincter with a force. As a result, most of the antral contents are forced back into the body of stomach and only a small amount of chyme passes into the duodenum (Fig. 7.3-17C). The backward movement of the food is called *retropulsion.*
- The forward and backward movements (caused by forceful propulsion and retropulsion) of the gastric contents help to break the food particles into smaller pieces and mixes it with gastric secretion converting it into a semiliquid paste called chyme.

Gastric emptying

- Gastric emptying occurs when the chyme is decomposed into enough small pieces (typically less than 1 mm^3) to fit through the pyloric sphincter.
- Gastric emptying results from a progressive wave of forceful contraction, which sequentially involves antrum, pylorus (pyloric sphincter) and proximal duodenum, thus all the three function as a unit.
- Each time the chyme is pushed against the pyloric sphincter, contraction ahead of advancing gastric contents prevent bigger food particles from entering the duodenum. Therefore, chyme is pumped in a bit (2–7 mL) at a time into the small intestine.

Factors regulating the gastric emptying

After a normal meal, the emptying time is 2–3 h. The gastric emptying is regulated by various factors:

1. Fluidity of the chyme. The rate of gastric emptying of solids depends on the rate at which the chyme is broken down into smaller particles. Liquids empty much faster than solids.

2. Gastric factors, which affect emptying, are:

- *Volume of food in the stomach.* Greater the volume of food in the stomach, greater is the stretching of stomach wall leading to strong peristaltic waves and an increased rate of gastric emptying.
- *Gastrin hormone.* Gastrin enhances the activity of pyloric pump and therefore promotes gastric emptying.
- *Type of food ingested* (present in the stomach) affects the gastric emptying as:
  - Carbohydrate-rich food causes rapid gastric emptying,
  - Protein-rich food causes slow gastric emptying and
  - Fat-rich food causes slowest gastric emptying.

3. Duodenal factors, which inhibit gastric emptying are:

- *Enterogastric reflex*. It is a neural-mediated reflex. It is initiated by the stimulation of receptors in the duodenal mucosa. The important stimuli are: distension of duodenum, acidity of the contents (pH < 4), high or low osmolarity of chyme, presence of fat and protein digestion products in the chyme.

  The enterogastric reflex is initiated in the duodenum and passes to the stomach through the myenteric plexus and also extrinsic nerves to inhibit or even stop emptying by inhibiting antral propulsive contractions and increasing slightly the tone of pyloric sphincter.

  Size of duodenal osmoreceptors affects the rate of gastric emptying. High osmolality of the chyme causes shrinkage and low osmolality increases the size of osmoreceptors. In both the conditions the rate of gastric

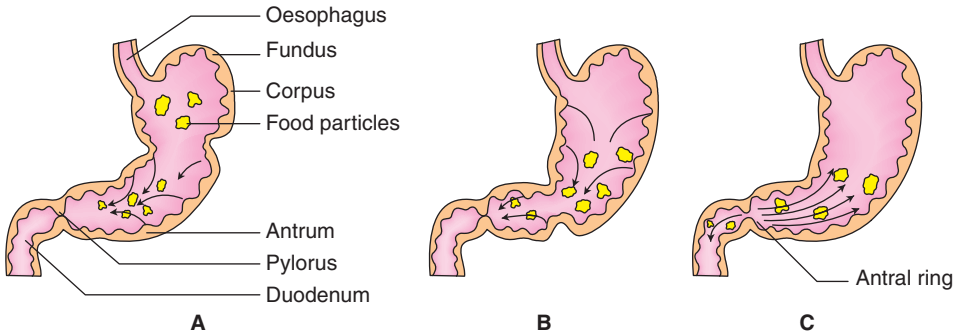


Fig. 7.3-17 Mixing peristaltic waves of stomach: A, peristaltic contractions begin in the mid stomach and pushes the food towards pylorus; B, when food reaches the pylorus it strikes against closed pyloric sphincter and C, antral contents forced back (retropulsion).

emptying decreases, but it is markedly inhibited when osmolality is high.

- *Enterogastric hormones.* A variety of intestinal hormones, collectively called enterogastrones, inhibit gastric contractions. Some of the hormones which have been identified are:
  - *Cholecystokinin,*
  - *Secretin* and
  - *Gastric inhibitory peptide.*

  Purpose of duodenal inhibitory effect on gastric emptying. The duodenal inhibitory effects (exerted through enterogastric reflex and enterogastrones) prevent the flow of chyme from exceeding the ability of intestine to handle it (especially longer time is required for fat digestion). It does not allow disturbance in electrolyte balance even if hypo or hypertonic solutions are drunk.

4. Other factors affecting gastric emptying:

- *Emotions e.g.* anger and aggression increase gastric motility whereas depression and fear decrease it.
- *Vagotomy and peptide Y* slow the gastric emptying.

FUNCTIONS OF STOMACH

After studying the physiological activities of stomach, its functions can be summarized as:

1. Mechanical or motor functions include:

- Storage of food
- Mixing of food
- Slow emptying of food into the duodenum occurs to provide proper time for digestion and absorption by the small intestine.

2. Digestive functions. Only small amounts of foods are digested in stomach as:

- *Carbohydrate digestion* in the stomach depends on the action of salivary amylase, which remains active until halted by the low pH of stomach.
- *Protein digestion.* About 10% of ingested protein is broken down completely in the stomach. *Gastric pepsin* facilitates later digestion of protein by breaking protein into peptone.
- *Fat digestion* in stomach is minimal due to the restriction of gastric lipase activity to triglycerides containing short-chain (>10 carbon) fatty acids.

3. Absorptive function. Stomach contributes little in absorptive function:

- *Absorption of nutrients.* Very little absorption of nutrients takes place in the stomach.
- *Ethanol* is absorbed rapidly in proportion to its concentration.
- *Water absorption.* Water-soluble substances, including Na^+, K^+, glucose and amino acids, are absorbed in insignificant amounts.
- *Intrinsic factor* released from the gastric glands helps in absorption of vitamin B_{12} from the small intestine.

4. Excretory function. Stomach excretes following substances:

- Certain toxins, as in case of uraemia and
- Certain alkaloids, such as morphine.

5. Stimulating functions. Stomach performs stimulatory function for the release of:

- Gastrin
- Enterogastrin
- Intrinsic factor of Castle

6. Reflex functions. Various reflexes initiated from the stomach are:

- Gastrosalivary reflex,
- Gastroileal reflex,
- Gastrocolic reflex and
- Presence of food in the stomach reflexly stimulates secretion of pancreatic juice and expulsion of bile.

7. Antiseptic action. HCl present in the gastric juice kills the bacteria and other harmful substances.

APPLIED ASPECTS

Important applied aspects of stomach which need special attention are:

- Gastric mucosal barrier and pathophysiology of peptic ulcer
- Physiology of vomiting

GASTRIC MUCOSAL BARRIER AND PATHOPHYSIOLOGY OF PEPTIC ULCER

GASTRIC MUCOSAL BARRIER

The gastric mucosal barrier protects the gastric mucosa from damage by the intraluminal HCl, i.e. autodigestion. It is created by the following:

- *Mucin secretion.* The thin layer of surface mucus in the stomach and duodenum prevents the direct contact of acid and pepsin-containing fluid with surface epithelial cells.
- *Bicarbonate secretion.* Surface epithelial cells in both the stomach and the duodenum secrete bicarbonate which create an essentially pH neutral microenvironment immediately adjacent to the cell surface.
- *Epithelial barrier.* Intercellular tight junctions provide a barrier to the back-diffusion of H^+. Any damaged cells

are quickly replaced, as the turnover rate of gastric mucosa is very high. Approximately, 5×10^5 mucosal cells are shed each minute, replacing the entire mucosa in 1–3 days.

- *High mucosal blood flow.* It rapidly carries away any acid that penetrates the cellular lining and also provides oxygen, bicarbonate and nutrients to the epithelial cells.
- *Prostaglandins* are responsible for maintaining the gastric mucosal barrier.

PATHOPHYSIOLOGY OF PEPTIC ULCER

Peptic ulcer refers to an excavation of mucosa of duodenum or pyloric part of stomach caused by the digestive action of gastric juice. Peptic ulcer can be caused by either of two ways:

1. Diminished ability of the gastroduodenal mucosal barrier to protect against the digestive properties of the acid–pepsin complex. Factors that disturb mucosal barrier include:

- *Bacterial infection by Helicobacter pylori.* At least 75% patients with peptic ulcer have recently been found to have chronic infection by *H. pylori.* The bacterium releases digestive enzymes that liquefy the barrier, which allows gastric secretion to digest the epithelial cells leading to peptic ulceration.
- *Other factors,* which can disrupt the mucosal barrier are ethyl alcohol, vinegar, bile salts, cigarette smoking and non-steroidal anti-inflammatory drugs (NSAIDs), such as aspirin and ibuprofen.

2. Excessive secretion of gastric acid

Hyperacidity leads to ulcer formation in the duodenum and pyloric part of stomach. Hyperacidity may occur due to:

- Increased parietal cell mass,
- Increased sensitivity for secretory stimuli,
- Excess gastrin secretion, as seen in the Zollinger–Ellison syndrome in which patients are having *gastrinomas* (tumours that secrete gastrin).

Physiologic basis of management of peptic ulcer

Commonly employed measures for treatment of peptic ulcer are:

Antacids. These form a gel that coats the mucosa and neutralizes the acid.

Drugs such as:

- *H_2-receptor blocking drugs,* such as cimetidine decreases HCl secretion by blocking the effect of histamine on H_2-receptors of parietal cells.
- *M_1-muscarinic receptor blocking drugs,* such as atropine decreases H^+ secretion by blocking the effect of ACh on M_1-muscarinic receptors of the parietal cells.
- *Gastric H^+–K^+–ATPase inhibiting drugs,* such as omeprazole, obviously decreases H^+ secretion by blocking the action of gastric H^+–K^+–ATPase in the parietal cells.

Bilateral vagotomy combined with resection of gastrin-producing pyloric part of stomach is performed in very severe cases of duodenal and pyloric ulcers.

PHYSIOLOGY OF VOMITING

Vomiting refers to the forceful expulsion of contents from stomach and intestine.

Initiation of vomiting

Vomiting may be initiated by either activation of vomiting centre or by an activation of the chemoreceptor trigger zone.

1. Activation of vomiting centre. Vomiting centre is situated in the reticular formation of medulla oblongata near the vagal nucleus. It may be activated directly or through afferents:

(i) Direct activation of the vomiting centre occurs due to an injury to the area or by raised intracranial pressure.

(ii) Afferent impulses activating vomiting centre includes (Fig. 7.3-18):

- *Visceral afferent pathway* in the sympathetic and vagi relay impulses arising due to the irritation of mucosa of upper GIT. There are 5HT receptors in the stomach and small intestine, and 5HT (serotonin) released from the enterochromaffin cells appears to initiate impulses in the afferents that trigger nausea and vomiting. These receptors are stimulated by local irritants. The receptors are stimulated by local irritants, such as drugs, viruses, radiations, bacteria, $CuSO_4$ and cytotoxic agents.
- *Afferent impulses from the vestibular nuclei* mediate nausea and vomiting of motion sickness.
- *Afferents from higher centres* (diencephalon and limbic system) mediate emetic response to stimuli, such as nauseating smell, sickening sights and memory, etc.

2. Activation of chemoreceptor trigger zone. Chemoreceptor trigger zone (CTZ) is located in the area postrema, a V-shaped band of tissue on the lateral walls of fourth ventricle near the obex which is outside the blood–brain barrier and is thus more permeable to many substances than the underlying medulla (Fig. 7.3-18). Chemoreceptor cells present in the CTZ initiate vomiting when they are stimulated by:

- Circulating emetic substances in patients with uraemia and radiation sickness.
- Circulating emetic agents, such as apomorphine, emetin, digitalis and glucosides.

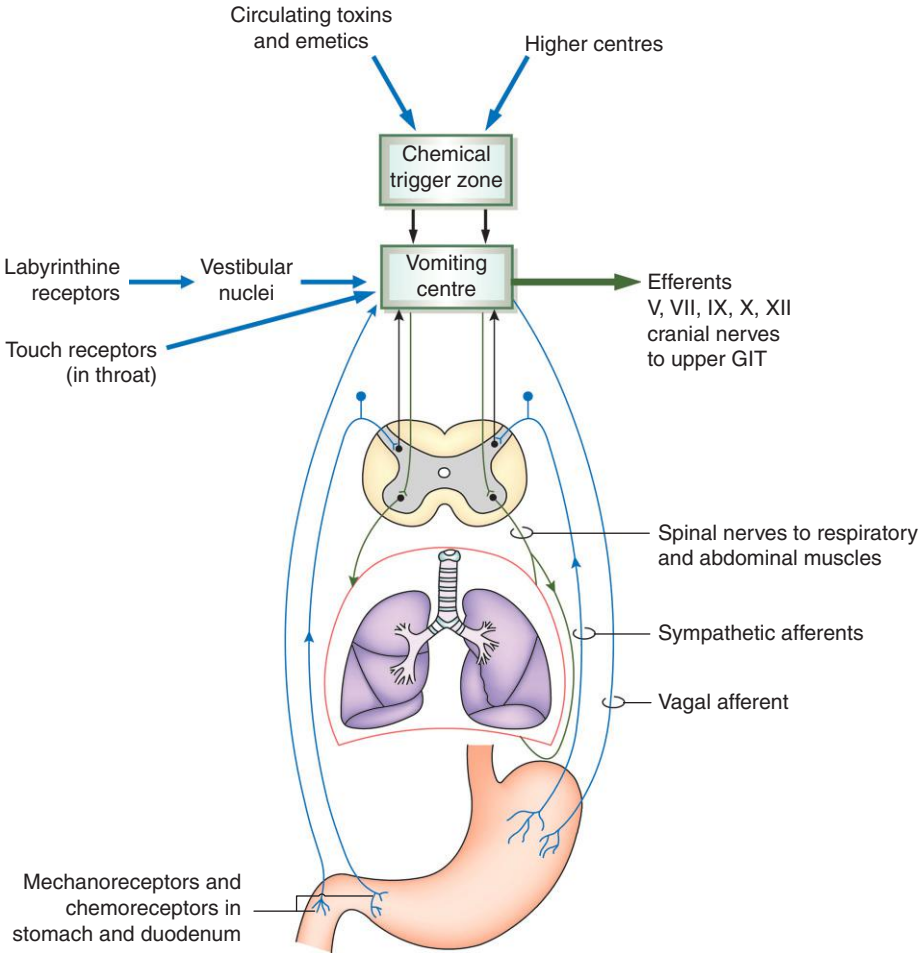


Fig. 7.3-18 Pathway for vomiting reflex.

Efferent impulses from vomiting centre and CTZ

Efferent impulses from the vomiting centre and CTZ, which give effect to act of vomiting, are transmitted via V, VII, IX, X and XII cranial nerves to upper GIT and through spinal nerves to the muscles of respiration.

Sequence of mechanical events of vomiting

Vomiting is a complex process and consists of following phases:

1. Pre-ejection phase. During this phase, there occurs gastric relaxation and retroperistalsis. It is characterized by the feeling of nausea, excessive salivation and deep, rapid and irregular breathing.

2. Retching phase may precede in many cases. It is characterized by:

- *Closure of glottis*, which remains so till the end of the act of vomiting. It increases intrapulmonary pressure causing compression of the oesophagus. It also prevents aspiration of vomitus in the trachea.
- *Rhythmic action of respiratory muscles* preceding vomiting and consisting of contraction of abdominal, intercostal and diaphragmatic muscles against a closed glottis.

3. Ejection phase during which the GIT contents are actually expelled out consists of events which occur in following sequence:

- Closure of glottis is continued.
- Pyloric part of stomach contracts firmly and its contents are transferred to the flaccid body of stomach.
- Simultaneous intense contraction of abdominal muscles and the descent of diaphragm raise the intra-abdominal pressure for such an extent that all the contents are squeezed out of the stomach, into the oesophagus as reflex relaxation of its cardiac sphincter also occurs at this moment.
- From the oesophagus, the contents are expelled into the mouth due to the effect of positive intrapulmonary pressure and antiperistaltic waves in the oesophagus. At this juncture, soft palate is raised and shuts off the nasal cavity from the throat.

- Towards the end of act of vomiting, diaphragm relaxes (i.e. ascends) and expiratory muscles and abdominal wall contracts.

TOTAL GASTRECTOMY

Nutritional disturbances after gastrectomy

Total gastrectomy, i.e. total removal of stomach, may produce following effects:

1. Effect on carbohydrate metabolism. The carbohydrates directly enter the duodenum and are digested and absorbed rapidly resulting in hyperglycaemia. As a result of *hyperglycaemia*, there occurs abrupt rise in insulin secretion which leads to hypoglycaemia after about 2 h of meals. Thus there occur sharp oscillations between hypoglycaemia and hyperglycaemia.

2. Effect on protein metabolism. The stomach plays an important role in the digestion of protein, but the near normal digestion of protein can occur in the absence of pepsin and nutrition can be maintained.

3. Effect on fat digestion. Almost no effect on fat digestion is seen except on butter fat, for which gastric juice has enzyme tributarase.

4. Effect on absorption of Vitamin B_{12}. Due to the deficiency of intrinsic factor, absorption of Vitamin B_{12} is affected markedly and there may occur *pernicious anaemia*.

5. Effect on iron absorption. Conversion of iron from ferric (Fe^{3+}) to ferrous (Fe^{2+}) form requires HCl. Therefore, iron absorption which occurs in ferrous form is affected, predisposing the individual to iron deficiency anaemia. However, only traces of iron are absorbed in the stomach.

Dumping syndrome

Dumping syndrome refers to a condition characterized by the development of weakness, dizziness and sweating after meals. This is seen in the cases of partial or total gastrectomy, where oesophagus is directly anastomosed with duodenum.

Causes of symptoms of dumping syndrome

1. Main cause of symptoms in dumping syndrome is sharp oscillations between hyperglycaemia and hypoglycaemia.
2. Another cause of symptoms is rapid entry of hypertonic meals into the intestine. It provokes the movement of so much water into the gut that the resultant reduction of plasma volume is great enough leading to a significant decrease in the cardiac output.

Treatment

- Meals should be small in bulk and dry.
- Milk and carbohydrate meals should be avoided and meals should have some dietary fibres.
- Daily and regular supplement of iron and Vitamin B complex is necessary to prevent the development of anaemia.

GASTRIC FUNCTION TESTS

Gastric function tests are employed to establish the presence of *hyperchlorhydria* (associated with peptic ulcer) or achlorhydria (complete absence of acid secretion) associated with pernicious anaemia. Gastric function tests include:

1. Fractional test meal test

Fractional test meal (FTM) test, previously used to be employed commonly for analysis of gastric juice. Presently, it is no more used because of its relative insensitivity and inconvenience.

Procedure, in brief, is described:

- After an overnight fast, the gastric fluid present in the stomach is aspirated with the help of the Ryles' tube passed into the stomach through nose or throat.
- A test meal is then given to stimulate gastric juice. Any one of the standard test meals can be given: 300 mL of oat meal gruel or dry toast and cup of tea or wheat biscuits and 300 mL of water.
- After 15 min of giving test meal, a sample of gastric content (10 mL of fluid) is aspirated. The procedure is repeated every 15 min for 3 h. Thus, including the fasting sample a total of 13 samples are collected in separate containers.
- Each sample is analyzed for free acidity, combined acidity (Fig. 7.3-19), starch and sugar, bile, total chlorides, blood lactic acid and mucus.

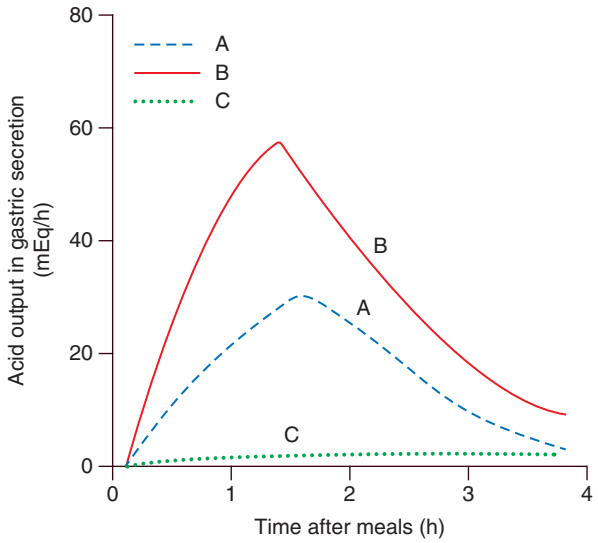


Fig. 7.3-19 Human gastric acid secretion in response to fractional test meal: A, normal response; B, hyperchlorhydria in duodenal ulcer and C, hyposecretion or achlorhydria.

2. Histamine test

Histamine test is comparatively more sensitive test than the fractional test meal test for studying the gastric acid secretion. It is because:

- First, the histamine is a powerful stimulator of acid secretion and
- Secondly, due to histamine injection, acid level increases very rapidly and therefore there is no time for neutralization of acid.

Procedure. After overnight fast, in the morning stomach is aspirated and washed with distilled water. Then, 0.5 mg histamine is injected subcutaneously and the gastric samples are aspirated and analysed as described in FTM test.

If there is no free acid present in any of the samples, then it is called true achlorhydria or histamine fast achlorhydria.

3. Augmented histamine test

- This test is performed with progressively increasing dose of histamine in a stepwise manner till a maximal secretory response is obtained.
- The maximal secretory response correlates well with the total number of the parietal cells in the gastric mucosa.

4. Pentagastrin test

- This test is also performed to assess the gastric acid status. It is performed similar to the histamine test except that in this, instead of histamine, 6 mg of pentagastrin (a synthetic gastrin) is given as subcutaneous injection.

5. Insulin test

- This test is based on the fact that hypoglycaemia (blood sugar below 45 mg%) produces vagal stimulation through hypothalamus and that vagal stimulation causes secretion of acid from the stomach.
- In this test, seven units of insulin are given intravenously and the gastric samples are tested for the presence of free acid, as done in histamine test.
- Acid secretion occurs after insulin injection (positive test) only if vagus is intact. Therefore, insulin test performed after vagotomy operation (for gastric ulcer) to know whether all the fibres of nerve supplying stomach are cut or not. If vagotomy is done properly, insulin test is negative.

6. Barium meal study

Barium meal study is a radiographic evaluation of the status of mucosa and lumen of the upper intestinal tract. In this test, patient swallows a suspension of radiopaque barium sulphate, while its passage through the GIT is observed by radiograph on fluorescent screen and films are taken to provide a permanent record. Diagnosis of gastric ulcer, duodenal ulcer or other abnormalities in the lumen of GIT is made from the typical finding.

7. Endoscopic examination and biopsy

These days condition of oesophagus, gastric and duodenal mucosa can be directly visualized by the endoscopic examination. This is more reliable than the conventional barium meal studies. The endoscopes carry a channel through which biopsy forceps or a brush can be introduced to obtain specimen for histological and cytological examination.

Chapter 7.4

Pancreas, Liver and Gall Bladder

PANCREAS

FUNCTIONAL ANATOMY

General considerations

- The pancreas—an elongated, accessory digestive gland—lies retroperitoneally and transversely across the posterior abdominal wall, posterior to the stomach between the duodenum on the right and spleen on the left (Fig. 7.4-1).
- *Anatomically,* pancreas is divided into four parts: head, neck, body and tail.
- *Physiologically,* on the basis of functions performed, the pancreas consists of two parts:
  - ***Exocrine part,*** which produces a secretion called pancreatic juice.
  - ***Endocrine part*** of the pancreas, the islets of Langerhans, produces the hormones insulin and glucagon. The endocrine part is discussed elsewhere (page 601).

Structural characteristics of exocrine part of pancreas

The exocrine part of the pancreas consists of serous, compound tubuloalveolar glands, very similar to the parotid gland in general structure (Fig. 7.4-2).

Acinar cells lining the alveoli appear triangular in section. Numerous secretory (or zymogen) granules can be demonstrated in the cytoplasm, especially in the apical part of cells.

The acinar cells produce thick secretion containing numerous enzymes (listed in composition of pancreatic juice).

Centroacinar cells. The centroacinar cells that are so called because they appear to be located near the centre of the acinus (alveolus). These cells really belong to the intercalated ducts, which are invaginated into the acinus (Fig. 7.4-2).

Pancreatic ducts

The intercalated ducts, which receive secretions produced by the acini, pass it on to the interlobular ducts. Ultimately, the pancreatic secretion passes into the duodenum through the main pancreatic duct and accessory pancreatic duct.

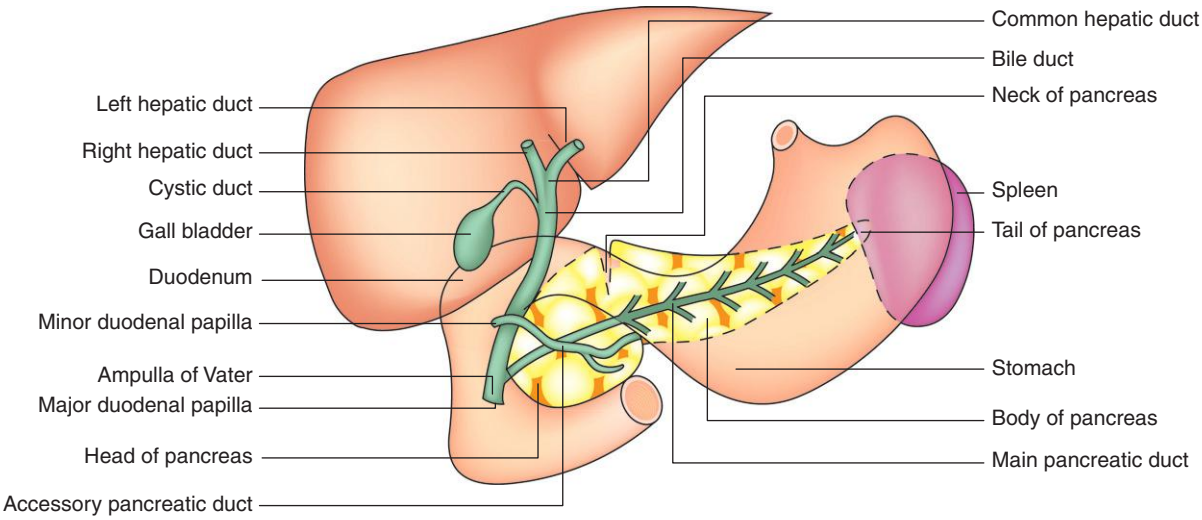


Fig. 7.4-1 Anatomical relations of pancreas, pancreatic duct and extrahepatic biliary system.

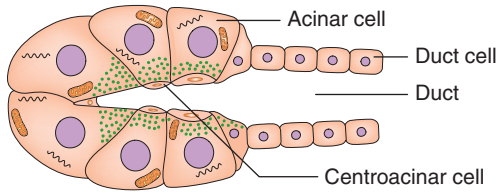


Fig. 7.4-2 Histology of functional unit of pancreas.

Main pancreatic duct, also known as a duct of *Wirsung*, begins in the tail and runs the length of the gland, receiving numerous tributaries on the way. It joins the common bile duct to form the ampulla of Vater, which opens into the second part of the duodenum at about its middle on the major duodenal papilla (Fig. 7.4-1). Ampulla of Vater is guarded by the sphincter of Oddi.

Accessory pancreatic duct, also called a duct of *Santorini*, when present, drains the upper part of the head and then opens into the duodenum about 2 mm above the main duct on the minor duodenal papilla.

Vessels and nerves of pancreas

Arterial supply to the pancreas comes from the splenic and superior as well as inferior pancreaticoduodenal arteries.

Veins, corresponding to arteries, drain into the portal system.

Lymphatics drain into the lymph nodes situated along the arteries that supply the gland. The efferent vessels ultimately drain into the coeliac and superior mesenteric lymph nodes.

Nerve supply comes from both sympathetic and parasympathetic (vagi) nerves. Pre-ganglionic vagal fibres synapse with the ganglionic cells embedded in the pancreatic tissue; the post-ganglionic fibres innervate both the acinar cells and the smooth muscles of the ducts. Vagal stimulation increases pancreatic juice secretion.

PANCREATIC JUICE

PROPERTIES

- Pancreatic juice is a transparent colourless fluid isotonic with plasma.
- About 1200–1500 mL of pancreatic juice is secreted per day.
- Its specific gravity varies from 1.010 to 1.018.
- Pancreatic juice is markedly alkaline (pH 7.8–8.4), due to very high concentration of HCO_3^- (about 4–5 times that of plasma).

COMPOSITION

Pancreatic juice is composed of 99.5% water and 0.5% solids, which include organic and inorganic substances.

Organic constituents of pancreatic juice are certain enzymes amylase, lipase, protease and trypsin inhibitor and other organic substances present in traces are albumin and globulin.

Inorganic substances present in the pancreatic juice are cations like Na^+, K^+, Ca^{2+}, Mg^{2+} and Zn^{2+}; and anions such as HCO_3^-, Cl^- and traces of SO_4^{2-} and HPO_4^{2-}. Electrolyte composition varies with rate of secretion.

Pancreatic enzymes

Pancreatic acini secrete four major types of enzymes:

1. Pancreatic α amylase. It is secreted in its active form. Its action on the carbohydrates is like that of the salivary amylase. It hydrolyses glycogen, starch and most other complex carbohydrates except cellulose to form disaccharides.

2. Pancreatic lipases or lipolytic enzymes include pancreatic lipase, cholesterol ester hydrolase and phospholipase A_2.

- *Pancreatic lipase.* It is a powerful lipolytic enzyme. It hydrolyses neutral fats to glycerol esters and fatty acids.
- *Cholesterol ester hydrolase* converts cholesterol esters to cholesterol.
- *Phospholipase A_2.* It is secreted in an inactive form, the pro-phospholipase A_2, and gets converted to an active form, the phospholipase A_2, by the action of trypsin. It acts on lysolecithin and lysocephalin and converts them into phosphoryl choline.

3. Pancreatic proteases or proteolytic enzymes include three endopeptidases (trypsin, chymotrypsin and elastase) and two exopeptidases (carboxypeptidase A and B).

- *Trypsin.* It is the most powerful proteolytic enzyme of the pancreatic juice. It is secreted in an inactive form of trypsinogen, which is activated by the enzyme enterokinase (enteropeptidase) secreted by duodenal mucosa. Once formed, trypsin also activates trypsinogen—an autocatalytic reaction.

$$\text{Trypsinogen} \xrightarrow{\text{Enterokinase}} \text{Trypsin}$$
$$\text{Trypsinogen} \xrightarrow{\text{Trypsin}} \text{Trypsin}$$

  - Trypsin hydrolyses proteins into proteoses and to polypeptides.
  - It activates trypsinogen and other pancreatic enzymes.
- *Chymotrypsin.* It is also secreted in an inactive form chymotrypsinogen and is activated by trypsin. It hydrolyses the proteins into small polypeptides.
- *Elastase.* It is secreted as pro-elastase, which is activated by trypsin. It digests elastin.
- *Carboxypeptidase A and B.* These are secreted as pro-carboxypeptidase A and B and are activated by enterokinase and trypsin.
  - Carboxypeptidase A cleaves the carboxyl-terminal amino acids that have aromatic or branched aliphatic side chains.
  - Carboxypeptidase B cleaves the carboxyl-terminal amino acids that have basic side chains.
- *Nucleases* (ribonuclease and deoxyribonuclease). They split nucleic acids of ribose and deoxyribose type into nucleotides.
- *Collagenase.* It is also activated by trypsin and digests collagen.

4. Trypsin inhibitor. If even a small amount of trypsin is released into the pancreas, the resulting chain reaction would produce active enzymes that could digest the pancreas. It is therefore, not surprising that the pancreas normally contains a trypsin inhibitor which is secreted by the same cells and at the same time as the pancreatic proenzymes. Trypsin inhibitor protects the pancreas from autodigestion.

FUNCTIONS OF PANCREATIC JUICE

1. Digestive functions. The pancreatic juice is the major source of digestive enzymes that digest all components of the food—proteins, carbohydrates, fats and nucleic acids.

2. Neutralizing function. Pancreatic juice is highly alkaline due to high concentration of HCO_3^- and neutralizes the gastric HCl in the chyme that enters the duodenum.

MECHANISM OF PANCREATIC SECRETION

Secretion of pancreatic enzymes

The acinar cells of the exocrine part of pancreas produce the pancreatic enzymes, which are synthesized within the ribosomes of rough endoplasmic reticulum amino acids derived from the Blood.

Formation of aqueous component of pancreatic secretion

The aqueous component of the pancreatic juice is produced principally by the columnar epithelial cells, which line the pancreatic ducts. The characteristics of aqueous component of pancreatic juice secreted by the acinar cells, intralobular ductal cells and extralobular ductal cells (Fig. 7.4-3) are as:

Secretion by acinar cells. The acinar fluid is isotonic and resembles plasma in its concentrations of Na^+, K^+, Cl^- and HCO_3^-.

Secretion by intralobular ductal cells. The spontaneous secretion that is produced by intralobular ductal cell has higher concentration of K^+ and HCO_3^- than does plasma.

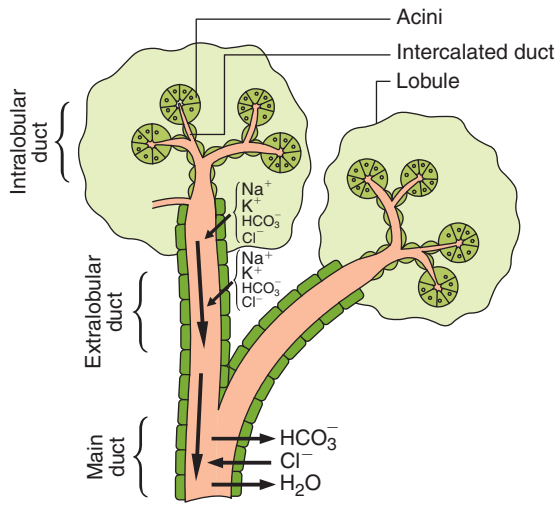


Fig. 7.4-3 Formation of the aqueous component of pancreatic juice at the level of acinar cell, intralobular ductal cells and extralobular ductal cells, and modification at the level of the main collecting duct.

Secretion by extralobular ductal cells is stimulated by the hormone secretin. The secretin stimulated secretion by the extralobular ductal cells is still richer in HCO_3^- than the spontaneous secretion.

Modification in the main collecting ducts. As the secretion flows through the main ducts, water moves into the duct across the epithelium (because the pancreatic duct cells are permeable to water) and makes the pancreatic juice isotonic to plasma. In addition some HCO_3^- move out of the ducts in exchange for Cl^- (Fig. 7.4-3).

Effect of flow rate on composition of aqueous component of pancreatic juice

The electrolyte composition of pancreatic juice varies with the secretion rate (Fig. 7.4-4).

- *Bicarbonate ion* (HCO_3^-) concentration of pancreatic juice at *low secretory* rate is as high as 80 mEq/L (much more than that of plasma, 26 mEq/L) and increases up to 120 mEq/L at *high flow rates.*
- *Chloride ion (Cl^-)* concentration decreases, as the flow rate of pancreatic juice increases; in other words, the HCO_3^- concentration rises. Thus, the total concentration of HCO_3^- and Cl^- remains constant and there exists a reciprocal relationship between their concentrations (Fig. 7.4-4).
- *Sodium (Na^+) and potassium (K^+)* concentrations in the pancreatic juice, unlike the saliva, are similar to those in the plasma and do not vary with the rate of secretion.

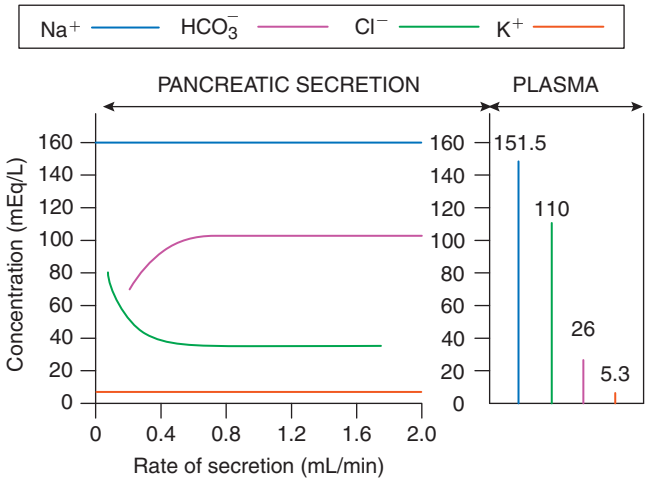


Fig. 7.4-4 Effect of rate of secretion of pancreatic juice on its electrolyte composition.

REGULATION OF PANCREATIC SECRETION

Both, neural and hormonal mechanisms are involved in the regulation of pancreatic secretion, with the later playing the predominant role. *Neural regulation* is through vagal efferents supplying the exocrine gland of pancreas and hormonal regulation is through secretin, cholecystokinin (CCK), gastrin and somatostatin. The exact role of these regulatory mechanisms in regulation of different phases of pancreatic secretion viz. cephalic phase, gastric phase and intestinal phase is summarized in Table 7.4-1.

1. Regulation of cephalic phase

Cephalic phase of pancreatic secretion like that of gastric secretion occurs before the entry of food into the stomach.

Regulation of this phase is mainly through the *reflex vagal stimulation* which occurs:

- *By conditioned reflexes,* initiated by sight, smell and thought of food, and
- *Unconditioned reflexes* initiated by stimulation of taste buds by the food in the mouth cavity, the act of chewing and swallowing.

Table 7.4-1 Summary of regulation of cephalic, gastric and intestinal phases of pancreatic secretion

Phase	Stimulus	Mediator	Pancreatic response
Cephalic	Conditioned reflex initiated by: • Taste, • Smell, and • Thought of food Unconditioned reflex initiated by taste of food in mouth	Vagus	Little secretion of pancreatic enzymes and HCO_3^-
Gastric	• Distension of stomach by food • Amino acids and peptides • Low pH of chyme in duodenum	Vagus Gastric Secretin	Low volume of pancreatic HCO_3^- and enzymes Low volume high enzyme secretion Large amount of aqueous secretion with high HCO_3^- concentration

2. Regulation of gastric phase

Gastric phase of pancreatic secretion occurs when stomach is distended by the food. This phase is regulated by *neural control* exerted through vagus and *hormonal control* executed through the hormone gastrin.

3. Regulation of intestinal phase

The intestinal phase of pancreatic secretion begins when the chyme enters the duodenum and jejunum. It is characterized by a marked increase in the secretion of both enzymes and aqueous component of the pancreatic juice. This phase is regulated by the hormones secretin and CCK.

(i) Roles of secretin

Secretin was the first hormone ever discovered by Bayliss and Starling in 1902. It is a polypeptide with 27 amino acids.

Source of secretin is endocrinal S-cells located among the epithelial cells of mucous membrane of the duodenum and jejunum.

Stimulant for the release of secretin is *low pH* (<4.5) of chyme caused by the presence of gastric HCl.

Actions. Secretin enters the blood circulation and after reaching the pancreas it *acts on the duct* cells and produces a large amount of watery juice with high concentration of HCO_3^-.

Other actions of secretin are:
- Also stimulates bile secretion,
- Potentiates the effect of CCK on pancreas and
- Along with CCK causes contraction of pyloric sphincter delaying gastric emptying and thus preventing the reflux of the duodenal contents into the stomach.

Regulation of secretin occurs through a negative feedback mechanism (Fig. 7.4-5).

(ii) Role of cholecystokinin

Cholecystokinin is polypeptide containing 33 amino acids.

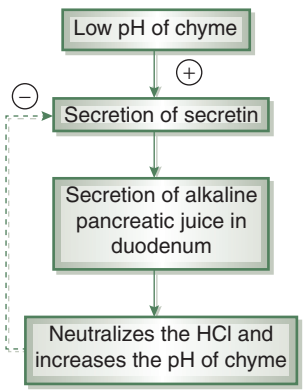


Fig. 7.4-5 Regulation of secretin secretion.

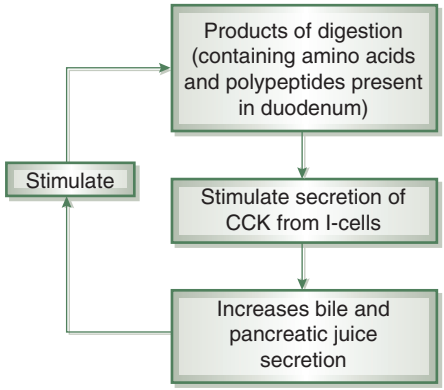


Fig. 7.4-6 Regulation of cholecystokinin secretion.

Source of secretion of CCK is endocrinal I-cells located among the epithelial cells of mucosa of the duodenum and jejunum. Stimulants for the release of CCK are amino acids, fatty acids and monoglycerides present in the chyme.

Actions. CCK passes via blood to the pancreas and causes *secretion of pancreatic juice rich in enzymes.*

Other actions of CCK are:
- Contraction of gall bladder to release bile.
- Potentiates the effect of secretin to produce more alkaline pancreatic juice.
- Increases the secretion of enterokinase (enteropeptidase) from the duodenum.
- Inhibits the gastric motility.
- Increases the motility of the small and large gut.
- Increases the pancreatic growth (trophic effect).
- It is also found in the neurons in the brain (especially in the cerebral cortex), where it is involved in the regulation of food intake and is related to the production of anxiety and analgesia.

Regulation of CCK secretion occurs through a positive feedback mechanism (Fig. 7.4-6):

Interaction of nervous and humoral regulation. A vagovagal reflex is initiated during the intestinal phase of digestion, which greatly potentiates the effects of secretin and CCK through the acetylcholine. Thus, vagus stimulation is much more potent in stimulating pancreatic secretions when CCK and secretin are present in the plasma.

APPLIED ASPECTS

DISORDERS OF PANCREAS

Common disorders of pancreas are:

Acute pancreatitis. It is an acute inflammatory disease of pancreas, thought to result from autodigestion of pancreatic tissue by the proteolytic enzymes, which leak out of the acini and are activated within the pancreas.

Chronic pancreatitis is a chronic inflammation of pancreas, which results in a slow destruction of the tissue resulting in the deficiency of pancreatic secretions. Patients with an extensive destruction of pancreas may develop:

- *Diabetes mellitus* due to pancreatic endocrine deficiency of insulin. For details see page 609.
- *Digestive disturbances* due to the deficiency of pancreatic enzymes mainly affect the fat metabolism resulting in *steatorrhoea,* which is characterized by bulky, foul smelling, pale and greasy stools (due to an increase in faecal fat content).

Cystic fibrosis is a disorder of pancreatic secretion. It is caused by a mutation in the cystic fibrosis transmembrane conductance regulator gene. It is associated with a deficiency of pancreatic enzymes resulting in steatorrhoea.

Pancreatectomy, i.e. surgical removal of pancreas, is usually performed in the carcinoma of the pancreas. It results in a deficiency of pancreatic enzymes characterized by the same features as described in the chronic pancreatitis.

PANCREATIC FUNCTION TESTS

Pancreatic function tests are performed to evaluate the normal functioning of the pancreas and to detect abnormality if any. The function tests to evaluate the functioning of the exocrine part of pancreas can be divided as follows:

- Analysis of pancreatic juice,
- Analysis of products of digestion and
- Estimation of serum amylase levels.

I. Analysis of pancreatic juice

Collection of pancreatic juice

A double lumen radiopaque tube (D veiling tube) is inserted through nose or mouth, till the tip of tube reaches the duodenum near the ampulla of Vater. The tube has weighted bulbous end and contains two sets of holes, one for the duodenal and other for the gastric aspiration. In this way, uncontaminated pancreatic juice can be collected from the duodenum.

The recent advanced method of collecting pure pancreatic juice involves use of *fibreoptic catheter* introduced under direct vision into the pancreatic duct.

Analysis of pancreatic juice collected after direct stimulation of pancreas

1. Secretin test. Secretin, which stimulates ductal cells, is used to measure the secretory capacity of these cells:

- After overnight fasting duodenal and gastric contents are aspirated in the morning.
- Intravenous infusion of secretin (12.5 units/kg body weight) is given and duodenal aspirate is collected at 10 min interval over the next 80 min.
- Aspirated contents are examined for volume, pH, HCO_3^- concentration and HCO_3^- output.
- Normal values are:
  - Volume output: > 2.0 mL/kg in 80 min.
  - HCO_3^- concentration: > 80 mEq/L.
  - HCO_3^- output: > 10 mEq/L in 30 min.
- Secretory activity of the ductal cells is decreased in chronic pancreatitis.

2. Combined secretin and CCK test. Combined secretin and CCK test is employed to evaluate the secretory capacity of both ductal cells and acinar cells. The test is performed as:

- First, secretin test is performed as described above.
- Then, CCK is given intravenously and the whole process is repeated.
- Curves for normal values of volume of pancreatic juice, HCO_3^- concentration and enzyme levels obtained by this test are shown in Fig. 7.4-7.
- Abnormalities can be detected from the results. With mild pancreatic damage, there is dissociation between the bicarbonate and enzyme output, i.e. only former is affected; with advanced damage, both are affected.
- This test helps to differentiate patients with steatorrhoea due to the intestinal malabsorption (in which test will be normal) from that due to the chronic pancreatitis (in which there will be decreased secretion of enzymes).

II. Analysis of products of digestion

1. Faecal fat excretion test. For this test, subject is placed on a diet containing 100 g of fat per day. The stools are collected over 3–5 days and tested for fat content by the Van de Kramer method and results are interpreted as:

- Normally, fats are digested by lipase (mainly from pancreas) and about 5–6 g/day are excreted in stools.
- In patients with exocrine pancreatic insufficiency it may increase to 40–50 g/day.

2. Tripeptide hydrolysis test. In this test, patient is given a synthetic peptide—B_2–T_4-PABA. Normally, B_2–T_4-PABA is cleaved by the chymotrypsin into B_2–T_4 and PABA. PABA is rapidly absorbed and excreted in urine. In exocrine pancreatic insufficiency, cleavage of B_2–T_4-PABA is decreased leading to decreased excretion of PABA in the urine. Thus, from the values of PABA in urine, activity of pancreatic chymotrypsin can be studied.

III. Estimation of serum amylase levels

This test is particularly useful to rule out acute pancreatitis in patients presenting with acute pain in the upper abdomen. Normal values of serum amylase are 50–120 units/L. The levels of serum amylase are markedly raised in the patients with an acute pancreatitis.

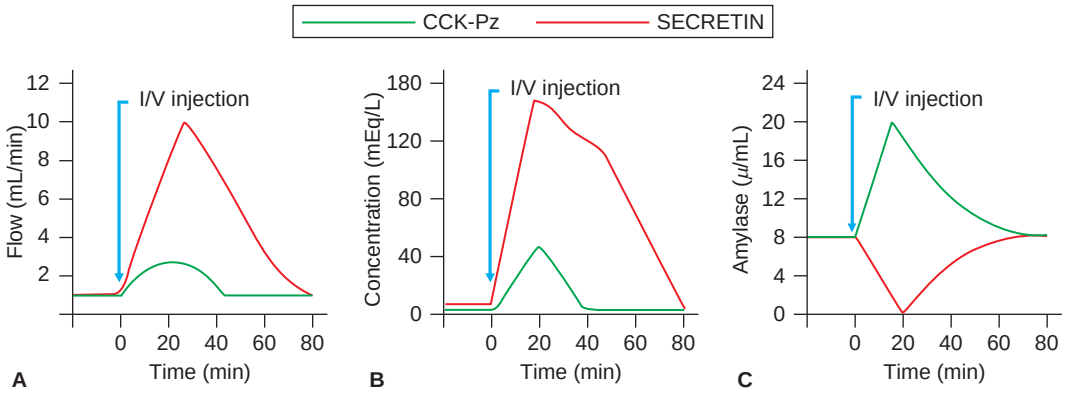


Fig. 7.4-7 Normal curves for combined secretin-cholecystokinin test: A, volume of pancreatic juice; B, HCO_3^- concentration of

LIVER AND GALL BLADDER

LIVER: PHYSIOLOGICAL ANATOMY

GENERAL CONSIDERATIONS

- Liver, the largest gland in the body, weighs approximately 1500 g.
- Anatomically, the liver has been divided into *right and left lobes.* Right lobe is much larger and includes caudate lobe and quadrate lobe. Left lobe is much smaller.
- In current terminology, the liver consists of right and left functionally independent parts called the *portal lobes* that are approximately equal in size (Fig. 7.4-8).
- The right and left functional parts of the liver have their own blood supply from the hepatic artery and portal vein and their own venous and biliary drainage.

IMPORTANT NOTE

Liver has got considerable physiological reserve. Even after removal of 80% of liver tissues, all *physiological* functions of liver can be accomplished normally.

The liver possesses considerable *regeneration power.* Original liver mass is restored within 6–8 weeks of removal of up to 3/4th of liver. This occurs due to an active mitotic division of the cells.

STRUCTURAL CHARACTERISTICS

The liver tissue comprises about one lac hexagonal areas that constitute the *hepatic lobules* (Fig. 7.4-9A).

- Each hepatic lobule is made of ramifying columns of hepatic cells *(hepatocytes)* that are arranged in the form of one cell thick plates. In between the cells are present bile canaliculi. These hepatic cell plates are tunnelled by a communicating system of lacunae called *blood sinusoids.* The sinusoids open into a central vein present in the centre of each lobule (Fig. 7.4-9B).

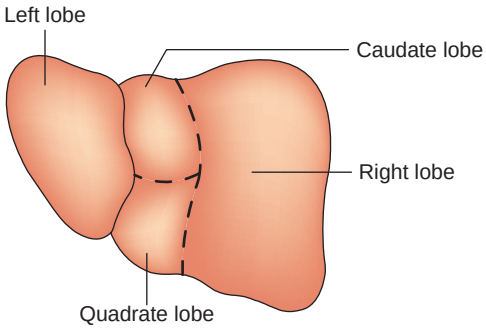


Fig. 7.4-8 Gross anatomy of the liver viewed from the back.

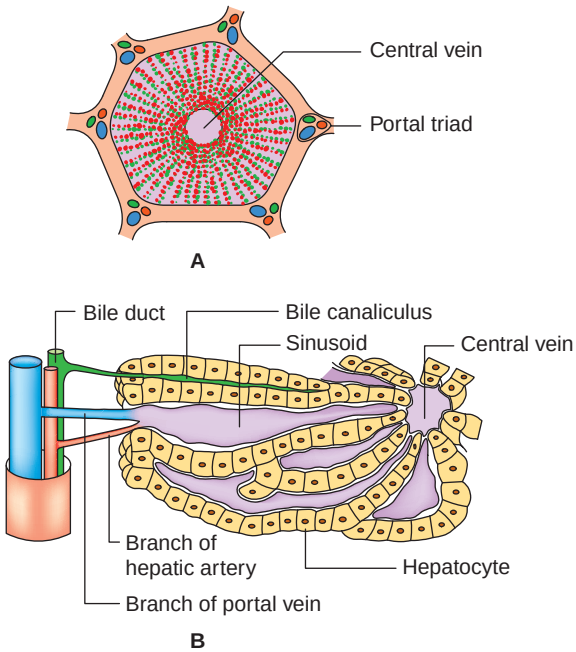


Fig. 7.4-9 Histological characteristics of the liver: A, hexagonal lobule with portal triad and B, hepatocyte, sinusoid and biliary canaliculi seen under high magnification.

- Blood sinusoids are lined by the endothelial cells. Few tissue macrophages called *Kupffer* cells are found at regular intervals in between the endothelial cells.
- Along the periphery of each lobule are present *portal triads* consisting of a branch of portal vein, branch of hepatic artery and an interlobular bile duct. Blood from the branch of portal vein and hepatic artery enters the sinusoids, which drain into central vein.
- *Concept of portal lobule,* instead of hepatic lobule, has been suggested by some workers. It has been described to consist of adjoining part of three hepatic lobules centred on a portal triad.
- Presently, acinus is considered the functional unit of liver (Fig. 4.6-14). Each acinus has been considered to have three zones: 1, 2 and 3.

Zone 1 refers to the central portion of the acinus immediately surrounding the terminal hepatic arteriole and terminal portal venule. This zone is well oxygenated.

Zone 2, i.e. the intermediate zone, which is present in between zone 1 and 3 is moderately well oxygenated.

Zone 3 refers to the peripheral most part of the acinus. It is least oxygenated and most susceptible to an anoxic injury.

HEPATIC CIRCULATION

Liver receives about 1500 mL blood/min from two sources:

Hepatic artery, which is a branch of coeliac trunk, supplies about 20–25% (300–400 mL/min) of total blood which caters to metabolic requirement of the liver tissue.

Portal vein, which collects blood from the mesenteric and splenic vascular bed, supplies about 75–80% (1100–1200 mL/min) of the total blood.

Hepatic vein. The hepatic and portal streams of blood meet in the sinusoids. The various substances produced by the liver cells, the waste products and CO_2 are discharged into the sinusoids. The sinusoids drain into the central vein of the lobule. The central veins from different lobules unite to form bigger veins. These veins ultimately form the right and left hepatic veins, which open into the inferior vena cava.

For details about the hepatic circulation see page 278.

HEPATIC BILIARY SYSTEM

INTRAHEPATIC BILIARY SYSTEM

The bile is secreted by the liver cells into bile canaliculi. These canaliculi have no walls of their own. In fact, the bile canaliculi are the spaces bounded by the canalicular surfaces of adjacent hepatic cells. These canaliculi form hexagonal network around the liver cells. At the periphery of a lobule, the canaliculi become continuous with delicate *intralobular ductules,* which in turn become continuous with larger interlobular ductules of portal triads. The *interlobular ductules* are lined by cuboidal epithelium. Some smooth muscle is present in the wall of larger ducts. Ultimately, the larger ducts join to form the *right and left hepatic ducts,* which leave the right and left parts of the liver and form a part of extrahepatic biliary system.

EXTRAHEPATIC BILIARY APPARATUS

The extrahepatic biliary apparatus consists of gall bladder and the extrahepatic bile ducts (Fig. 7.4-1).

Gall bladder

The gall bladder is a pear-shaped sac lying on the undersurface of liver. It has a capacity of about 30–50 mL and stores bile, which it concentrates by absorbing water.

For descriptive purposes, the gall bladder is divided into fundus, body and neck. The neck becomes continuous with the *cystic duct.*

Extrahepatic ducts

Hepatic ducts. The right and left hepatic ducts emerge from the right and left lobes of the liver and after a short course, join to form *common hepatic duct* which is about 4 cm long.

Cystic duct. It is also about 4 cm long and connects the neck of gall bladder to the common hepatic duct to form the common bile duct.

Common bile duct (CBD) is about 8 cm long. It joins the pancreatic duct to form the common hepatopancreatic duct which is otherwise called the ampulla of Vater. Ampulla of Vater opens into the duodenum at major duodenal papilla. The terminal parts of bile ducts and ampulla of Vater are surrounded by circular muscle fibres known as sphincter of Oddi, which plays an important role in the storage and release of bile from the gall bladder.

FUNCTIONS OF LIVER

The fact that mitochondria are maximally present in the liver emphasizes that liver is involved in many biochemical functions. Although details of the various functions performed by liver are discussed under their respective places, they are summarized here briefly.

I. Secretory functions

Liver cells act as an exocrine gland and continuously secrete bile, which is important for digestion and absorption of fats. Various aspects of the bile juice are discussed in this chapter.

II. Metabolic functions

Liver is the key organ and the principal site where the metabolism of carbohydrates, lipids and proteins takes place.

Liver is also involved in the metabolism of vitamins and minerals to certain extent.

1. Role in carbohydrate metabolism includes:

(i) Liver acts as a glucostat in three ways:
- *Glycogenesis,* i.e. glycogen is formed from glucose and stored in liver.
- *Glycogenolysis,* i.e. breaking down of liver glycogen to glucose.
- *Glucogenesis,* i.e. formation of glucose from non-carbohydrate sources, such as non-nitrogenous residues of amino acids.

(ii) Liver is the main site of alcohol metabolism for which liver cells contain the enzyme *alcohol dehydrogenase.*

(iii) The interconversion of three monosaccharides, such as glucose, galactose and fructose, also occurs in liver.

2. Role in fat metabolism. Both degradation and synthesis of fats take place in the liver.

Degradation of fat. Liver contains enzyme lipoprotein lipase which hydrolyses triglycerides, cholesterol and phospholipids into fatty acids.

- β-oxidation, i.e. a process which oxidises the fatty acids into acetoacetic acid occurs within the mitochondria.

Synthesis of fat also takes place in liver.

- Liver synthesizes triglycerides from carbohydrates.
- Cholesterol and phospholipids are synthesized from unused free fatty acids.
- Saturated fatty acids are synthesized from the active acetate via Krebs' cycle within the mitochondria.
- Lipoproteins, such as HDL, LDL, VLDL and chylomicrons are also synthesized in liver.

3. Role in protein metabolism. In man, the protein turnover involves breakdown and resynthesis of 80–100 g of tissue protein per day, and its 50% part (i.e. 40–50 g) occurs in the liver. Important activities are:

- Liver brings about *deamination* of amino acids and this is essential for energy production, and their conversion into carbohydrates or fats.
- Liver is the main site of *urea formation.*
- Liver is the main site for formation of all non-essential amino acids by the transamination of ketoacids.
- Albumin is solely resynthesized in liver and also to some extent α- and β-globulins.

III. Detoxicating and protective functions

- Kupffer cells efficiently remove bacteria and other foreign bodies from the portal circulation. This is the blood cleansing action of liver.
- Liver detoxifies certain drugs by either oxidation, or hydrolysis, or reduction or conjugation and excretes out through bile.

IV. Storage functions

Liver stores glucose (in the form of glycogen), vitamin B_{12} and vitamin A.

- Liver acts as a blood iron buffer and iron storage medium. It stores 60% of excess of iron mainly in the form of ferritin and partly as haemosiderin.

V. Excretory functions

Certain exogenous dyes like bromsulphthalein (BSP) and rose bengal dye are exclusively excreted through the liver cells.

VI. Synthesis functions

Liver is the site for synthesis of:

- *Plasma proteins,* especially albumin and to some extent α- and β-globulins.
- *Some blood coagulation factors.* Liver cells are responsible for the conversion of pre-prothrombin (inactive) to active prothrombin in the presence of Vitamin K. It also produces other clotting factors, such as fibrinogen (I), factors V, VII, IX and X.
- *Enzymes,* such as alkaline phosphatase, serum glutamic oxaloacetic transaminase (SGOT), serum glutamic pyruvic transaminase (SGPT), serum isocitrate dehydrogenase.
- *Urea.* Liver removes ammonia from the body to synthesize urea.
- *Cholesterol.* It is synthesized from the active acetate.

VII. Miscellaneous functions

- *Reservoir of blood.* Liver acts as a reservoir of blood and it stores about 650 mL of blood. Also helps in the regulation of blood volume.
- *Erythropoiesis.* Liver is an important site of erythropoiesis in the fetal life.
- *Hormone metabolism.* Liver causes:
  - *Inactivation* of some hormones, such as insulin, glucagon and vasopressin.
  - *Reduction and conjugation* of adrenal and gonadal steroid hormones, such as cortisol, aldosterone, oestrogen and testosterone.
- *Destruction of RBCs* also occurs in the liver.
- *Thermal regulation.* Liver also helps in thermoregulation, as it produces a large amount of heat.

BILE AND GALL BLADDER

GENERAL CONSIDERATIONS

- Bile is a digestive juice, formed continuously in the liver.
- It is poured into the bile canaliculi from where it ultimately goes to a common hepatic duct which joins with cystic duct to form common bile duct. During the interdigestive period when the sphincter of Oddi is closed, the

bile is directed via cystic duct to the gall bladder, where it is stored and concentrated.

- During meals, the sphincter of Oddi is relaxed and when food reaches the duodenum, there occurs release of CCK which causes contraction of the gall bladder. Then the bile is released into the duodenum along with the pancreatic juice through the common opening ampulla of Vater.

FORMATION AND COMPOSITION OF BILE

The bile is formed by the hepatocytes and ductal cells lining the hepatic ducts. The hepatocytes, one surface of which is adjacent to the blood sinusoids and other to the biliary canaliculi, pick up some constituents of bile from the blood (e.g. *bile pigments*), synthesize some constituents (e.g. *bile salts*) and secrete a mixture into the biliary canaliculi. Ductular cells contribute HCO_3^- and Cl^- to the mixture giving rise to the *hepatic bile* (Fig. 7.4-10).

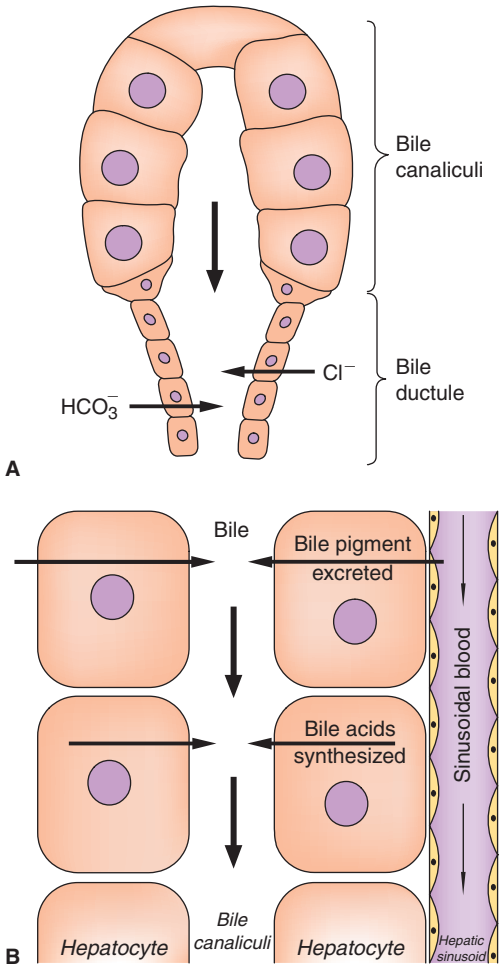


Fig. 7.4-10 Mechanism of bile formation: A, secretion by hepatocytes and ductal cells and B, bile pigments picked up from the blood sinusoids are excreted while bile salts are synthesized and secreted by the hepatocytes.

The bile so formed is an alkaline juice comprised of:

- Water and solids
- Solids include organic and inorganic substances
- Organic substances are bile salts, bile pigments, cholesterol, lecithin, fatty acids and enzyme alkaline phosphatase
- Inorganic substances are Na^+, K^+, Ca^{2+}, HCO_3^- and Cl^-.

Since the bile is concentrated in the gall bladder, so the concentration of its ingredients in the liver bile and gall bladder bile is bound to differ as shown in Table 7.4-2.

Salient features of the some of the ingredients of bile are described here:

1. Bile salts

Formation of bile salts

Bile salts are sodium and potassium salts of bile acids conjugated with either taurine or glycine. Bile acids are of two types: primary and secondary. Steps in the formation of bile salts (Fig. 7.4-11) are:

- *Primary bile acids* are cholic acid and chenodeoxycholic acid. These are synthesized by the hepatocytes from cholesterol.
- *Secondary bile acids* are deoxycholic acid and lithocholic acid. These are formed from the primary bile acids in the colon by the action of intestinal bacteria.
- *The conjugation of bile acids.* In the liver, the bile acids are conjugated with either glycine (an amino acid) or taurine forming the conjugated bile acids.
- The conjugated bile acids, namely glycocholic acid and taurocholic acid form bile salts in combination with sodium or potassium.

Table 7.4-2 Liver bile versus gall bladder bile

Properties and composition	Liver bile	Gall bladder bile
• pH	7.8–8.6	7–7.6
• Colour	Light golden yellow	Blackish
• Consistency	Watery	Thicker
• Water	97.5%	87.5%
• Solids	2.5%	12.5%
Organic substances		
• Bile salts	1.10 g/dL	8.0 g/dL
• Bile pigments	0.20 g/dL	1.0 g/dL
• Cholesterol	0.10 g/dL	0.5 g/dL
• Fatty acid	0.15 g/dL	0.5 g/dL
• Fat	0.10 g/dL	0 g/dL
• Lecithin	0.1 g/dL	0.8 g/dL
• Mucin	Absent	Present
Inorganic substances	0.75 g/dL	8.7 g/dL

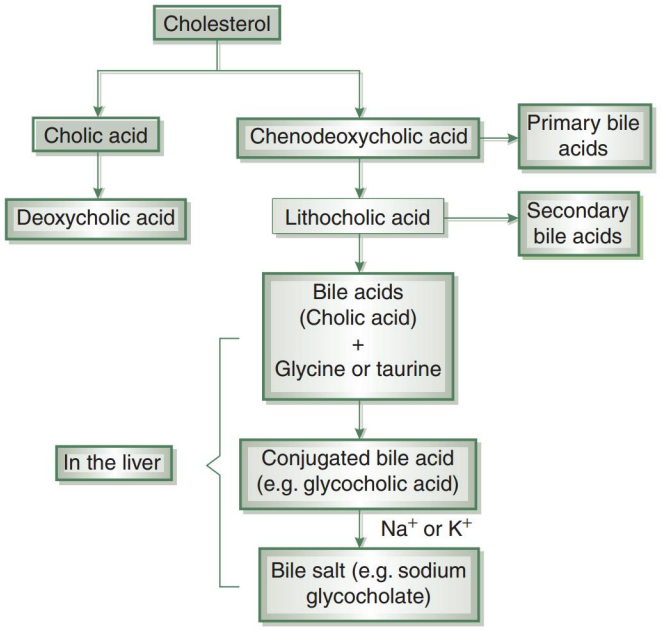


Fig. 7.4-11 Formation of bile salts from the bile acids.

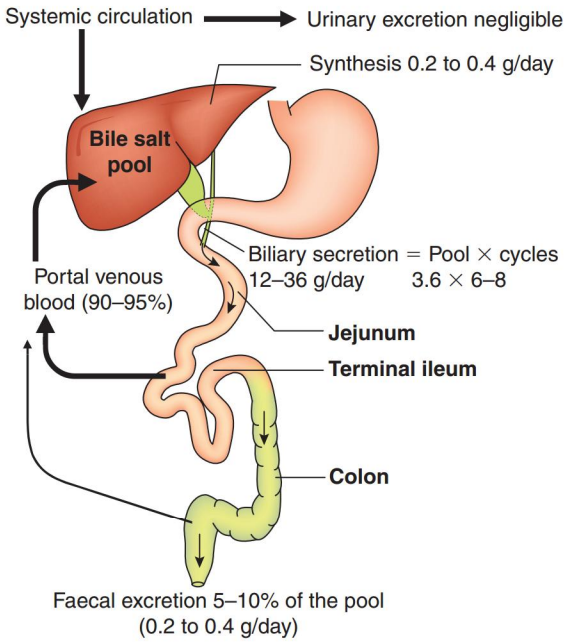


Fig. 7.4-12 Enterohepatic circulation of bile salts.

Functions of bile salts

Bile salts help in digestion as well as absorption of fat by their following actions:

- *Emulsification of fat,* i.e. breaking of large fat drops into smaller droplets and their stabilization is caused by the bile salts because of their power of lowering surface tension.
- *Acceleration of action of pancreatic lipase* occurs in the presence of bile salts due to binding of colipase for the lipase.
- *Micelle formation.* The bile salts combine with the products of hydrolysis of triglycerides to form *small water-soluble cylindrical disc-shaped particles called* micelle, which are transported to the brush border of the epithelial cells for absorption.
- *Absorption of fat-soluble vitamins* (A, D, E and K) is aided by the bile salts by forming complexes more soluble in water (hydrotrophic action).
- *Choleretic action,* i.e. they stimulate liver to secrete bile and then make more bile salts available for fat digestion.
- *Cholesterol is kept in soluble form* in the gall bladder bile by the bile salts. This property of bile salts prevents formation of gall stone.
- *Intestinal motility* is stimulated by the bile salts. This action of bile salts help in defaecation (laxative action).

Enterohepatic circulation of bile salts

Enterohepatic circulation is the recirculation of bile salts from the liver to small intestine and back again.

Path of circulation (Fig. 7.4-12). Bile salts travel from the liver to duodenum via the common bile duct.

- When the bile salts reach *terminal ileum,* 90–95% of bile salts are reabsorbed into the portal circulation. It is important to note that no absorption of bile salts occurs in the duodenum and jejunum. The liver then extracts the bile salts from the portal blood and secretes them once again into the bile.
- The remaining 5–10% of bile salts are excreted into the faeces.
- *Circulating pool.* The total circulating pool of bile salts is approximately 3.6 g. About 4–8 g of bile salts are required (more if the meal is high in fat) for digestion of fats during each meal; thus the total content of bile salts in the body (3.6 g), most circulate twice during the digestion of each meal. Consequently, the bile salts usually circulate 6–8 times daily.

2. Bile pigments

- The two principal bile pigments, *bilirubin* and *biliverdin* are the other major constituents of bile which have no digestive function.
- Bile pigments are metabolites of haemoglobin formed in the liver. The hepatic cells extract bilirubin and biliverdin from the blood, conjugate them with the glucuronic acid. They are responsible for the golden yellow colour of bile.
- Intestinal bacteria metabolize bilirubin further to urobilinogen, which is responsible for the brown colour of stools.

Formation and circulation of bile pigments and jaundice are described on page 114.

3. Phospholipids

The phospholipids (primarily lecithins) are, after bile salts, the most abundant organic compound in bile.

4. Cholesterol

Cholesterol is another important constituent of bile that does not have digestive function. Its presence in the bile seems to be a byproduct of bile salt synthesis in the hepatic cells. Normal biliary content of cholesterol is about 100 mg/dL (60–170 mg/dL) as compared to 150–240 mg/dL in the blood.

- Biliary secretion of cholesterol is important because it is one of the few ways in which cholesterol stores can be regulated.
- Biliary cholesterol forms an important component of gall stones (large sand) like particles found in the gall bladder of some patients.

5. Electrolytes

- Biliary content of inorganic substances is about 0.75 g/dL.
- The cations Na^+, K^+ and Ca^{2+} are all present in concentration about 20% greater than in the plasma.
- Two major anions are Cl^- and HCO_3^-, Cl^- is present in concentrations lesser than in the plasma while HCO_3^- is far greater than in the plasma, which makes the bile juice considerably alkaline. Further, HCO_3^- concentration increases with an increased rate of bile secretion.

FUNCTIONS OF GALL BLADDER

Gall bladder subserves following functions:

1. Storage of bile. The bile secreted during interdigestive period is stored in the gall bladder. The gall bladder, typically stores 30–50 mL of bile. During meals, the gall bladder contracts and releases its contents into the duodenum.

2. Concentration of bile. The mucosa of gall bladder is extensively folded and can actively absorb fluid and electrolytes. In this way, the gall bladder bile in comparison to liver bile:

- Becomes thicker, viscous and darker in colour.
- Water content is decreased (from 97% to 87.5%).
- All organic constituents, which are not absorbed become 5–6 times concentrated.
- Cl^- and HCO_3^- ions decrease by 5–6 times (due to active absorption).
- Ca^{2+} and K^+ which are not absorbed increased by two times.

3. Effect on the pH of bile. In the gall bladder, due to rapid absorption of HCO_3^- (mainly), Na^+ and Cl^-, the pH of bile is decreased from 8–8.6 to 7–7.6.

4. Secretion of mucous. Gall bladder secretes mucin, which is added to the bile stored in it. The mucin acts as a lubricant in the intestine for the chyme.

5. Regulates equalization of pressure in biliary system. Due to continuous absorption of water from the stored bile, the gall bladder regulates equalization of pressure in the biliary system. This fact can be understood by following observations:

- When both the bile duct and cystic duct are clamped, the pressure in the biliary system rises to above 30 cm of bile in 30 min and bile secretion is stopped.
- When the bile duct is clamped alone, water is continuously reabsorbed in the gall bladder, and the pressure in the biliary system rises to only 10 cm of bile in several hours. Thus, gall bladder prevents the rise of pressure in biliary system.

FUNCTIONS OF BILE

Functions subserved by the bile poured into the duodenum are because of its constituents (mainly bile salts), which have already been discussed. However, they are compiled and summarized once again:

1. Digestive function. Bile salts help in the digestion of fats by *emulsifying fat drops* (see page 515).

2. Absorptive functions. Bile salts help in the absorption of fats (by micelle formation) and fat-soluble vitamins (see page 515).

3. Excretory function. Bile pigments are the major excretory products of the bile. The other substances excreted in bile are heavy metals (e.g. copper and iron), some toxins, some bacteria (e.g. typhoid bacteria), cholesterol, lecithin and alkaline phosphatase.

4. Laxative action. Bile salts increase the gastrointestinal motility and act as a laxative.

5. Protective action. Bile is a natural detergent. So, it inhibits the growth of certain bacteria in the lumen of intestine.

6. Choleretic action, i.e. bile salts stimulate the liver to secrete bile.

7. Maintenance of pH of gastrointestinal tract. Being highly alkaline, the bile juice neutralizes the gastric HCI present in the chyme entering the small intestine. Thus, an optimum pH is maintained for the action of digestive enzymes.

8. Prevention of gall stone formation. Bile salts keep the cholesterol and lecithin in solution and thus prevent the formation of gall stones. In the absence of bile salts, the cholesterol precipitates along with lecithin and may form gall stones.

9. Lubricating function. The mucin secreted by the gall bladder mucosa into the bile lubricates the chyme in the intestine.

10. Cholagogue function. Cholagogue is an agent, which increases the release of bile from gall bladder into the intestine. The bile salts perform this function indirectly. The bile salts stimulate the secretion of hormone CCK, which has got cholagogue action.

REGULATION OF BILE

The regulation of bile juice released into the duodenum after the meals is performed at two levels:

- Regulation of biliary secretion and
- Regulation of release of bile from the gall bladder.

A. Regulation of biliary secretion

The secretion of aqueous component (water and electrolytes) and the bile (containing bile salts and other organic substances), though occur together but is controlled separately by following mechanisms (Fig. 7.4-13A):

- Regulation of bile-independent fraction of biliary secretion and
- Regulation of bile-dependent fraction of biliary secretion.

I. Regulation of bile-independent fraction of biliary secretion. The bile-independent fraction of biliary secretion refers to the amount *of fluid-containing water and electrolytes.* Secretion of this fraction of the bile juice is controlled by secretin and vagal stimulation.

- Secretin. It acts on the ductal cells of hepatic ducts (Fig. 7.4-13A) via cyclic AMP, second messenger and produces a large amount of watery fluid with high concentration of HCO_3^-.
- Vagovagal reflex initiated during the intestinal phase of digestion also affects the ductal secretion by potentiating the effects of secretion through the acetylcholine.

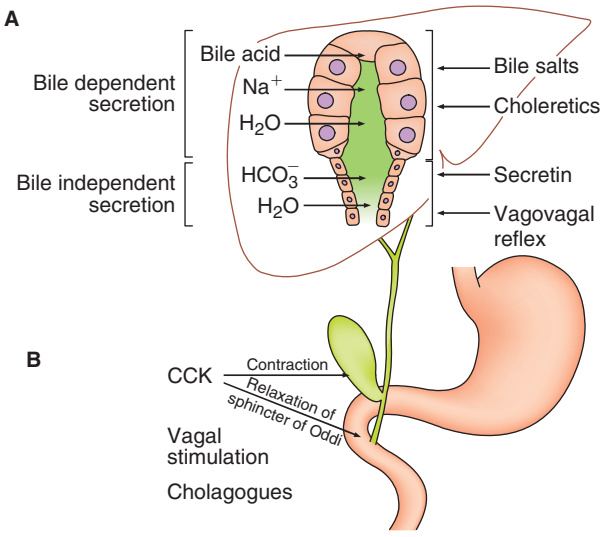


Fig. 7.4-13 A, Regulation of bile secretion and B, release from gall bladder.

Note. The agents (e.g. secretin and acetylcholine) which cause secretion of bile from liver with more amount of water and less amount of solids are called hydrocholeretics.

II. Regulation of bile-dependent fraction of biliary secretion. The bile-dependent fraction of biliary secretion refers to the quantity of bile salts secreted by the liver. It depends upon the following factors (Fig. 7.4-13):

- The amount of bile salts secreted by the hepatocytes is directly proportional to the *amount of bile salts reabsorbed* by them from the portal circulation. As the bile salts are recycled in the enterohepatic circulation, they maintain high level of bile secretion during digestive period.
- Bile salts and bile acids are the major agents which enhance synthesis of bile salts. Substances that enhance the secretion of bile salts by the hepatocytes are called choleretics.

Note. Synthesis of bile salts by liver is not controlled by any hormonal or nervous factor.

B. Regulation of release of bile juice from gall bladder

Filling of gall bladder by the bile is simply controlled by the pressure gradient. During the interdigestive period the sphincter of Oddi remains closed. As bile is secreted continuously, it gets accumulated in the CBD. When pressure of bile in the CBD rises, it forces its way through cystic duct into the gall bladder. During interdigestive period, the pressure in CBD and gall bladder reaches to 7 cm of water.

Emptying of gall bladder. When the chyme enters the duodenum, the gall bladder is contracted along with relaxation of sphincter of Oddi, raising the pressure to about 20 cm of water. Because of the increase in pressure, the bile from the gall bladder enters the duodenum. The contraction of gall bladder and relaxation of sphincter of Oddi are regulated by following factors (Fig. 7.4-13B):

Hormonal control. The hormone CCK is the major stimulus for the gall bladder contraction and sphincter of Oddi relaxation. When chyme enters the small intestine, *fat and protein digestion products* directly stimulate the secretion of CCK (For details about CCK see page 485).

Neural control. Vagal stimulation also causes contraction of the gall bladder and relaxation of sphincter of Oddi. Vagal stimulation occurs directly during the cephalic phase of digestion and indirectly via a vagovagal reflex during the gastric phase of digestion.

Note. Substances that cause contraction of gall bladder are called *cholagogues.* Thus CCK is well known cholagogue and acetylcholine released on the vagal stimulation also acts as a cholagogue.

APPLIED ASPECTS

DISORDERS OF LIVER AND GALL BLADDER

Jaundice or icterus

Jaundice or icterus refers to the yellow discolouration of skin and mucous membrane due to raised levels of bilirubin in the blood.

- *Normal values* of serum bilirubin range between 0.3 and 1 mg/dL,
- *Jaundice* manifests when serum bilirubin becomes more than 2 mg/dL.

Types of jaundice. Jaundice is of three types:
- *Haemolytic jaundice* pre-hepatic jaundice.
- *Hepatocellular jaundice* occurs due to damage to hepatocytes, as seen in viral hepatitis, cirrhosis and drug-induced hepatitis.
- *Obstructive jaundice,* also called post-hepatic jaundice, occurs due to blockage in bile duct either due to stone or any growth (e.g. in carcinoma of head of pancreas).

For further details and differences in the three types of jaundice, see page 116.

Cirrhosis of liver

Cirrhosis of liver refers to an irreversible chronic damage to liver with extensive fibrosis and regenerative nodule formation.

Viral hepatitis

Aetiology. Viral hepatitis (inflammation of liver) is caused by hepatitis virus A, B, C, D, E, F or G.

Hepatitis-B is more popular.

The ***clinical features*** depend on the loss of liver tissue.

Signs of liver insufficiency or damage are:

1. *Oedema* occurs due to hypoproteinaemia.
2. *Haemorrhagic disorders* occur due to lack of clotting factors synthesis.
3. *Muscle weakness, tremors and convulsions* (features of hepatic encephalopathy) occur due to fall in blood glucose level because of less production of glucose.
4. Liver enlargement and portal hypertension lead on ascites.
5. *Steatorrhoea* occurs due to defective fat digestion and absorption owing to low bile salts concentration in the bile. For details see page 495.
6. Anaemia and jaundice.
7. *Levels of alkaline phosphatase, SGOT and SGPT* rise because these enzymes are released from damaged hepatic cells.
8. *Blood urea* level decreases and a*mmonia* level in the blood and urine rises (as liver removes ammonia from the body by synthesising urea).

Cholecystitis

Cholecystitis refers to the inflammation of gall bladder.

GALL STONES

Gall stone formation (cholelithiasis) is not an uncommon problem.

Two types of gall stones are known:
- *Cholesterol stones* account for 80–85% of the cases. These are formed due to precipitation of cholesterol. Normally, cholesterol is present in soluble form due to a proper ratio of cholesterol and bile salts (1:20–1:30). When, this ratio falls below 1:13, the cholesterol is precipitated forming many small crystals. This stimulates further formation of crystals, so that the crystals grow larger and larger. In these crystals, bile pigments and calcium also get inspirated forming gall stones. These stones are *radiolucent*, i.e. cannot be visualized on radiograph.
- *Pigment stones*, also called *calcium bilirubinate stones* account for 15–20 cases of gall stones. These stones are formed when the conjugated bilirubin in the bile is disconjugated by the action of β-glucuronidase found in certain bacteria. The free bilirubin combines with calcium to form calcium bilirubinate, which is highly insoluble in bile. These stones are radiopaque.

LIVER FUNCTION TESTS

The liver function tests (LFTs) are the investigations to assess the capacity of the liver to carry any of the functions it performs. Thus, LFTs help in:

- Assessing the extent of functional damage to the liver,
- Diagnosing the cause of hepatic insufficiency and
- Assessing the progress/regress of the disease.

I. Tests to assess secretory functions of liver

1. Serum bilirubin. The *normal values* are:
- Total serum bilirubin: 0.3–1.0 mg/dL,
- Conjugated bilirubin: 0.1–0.3 mg/dL and
- Unconjugated bilirubin: 0.2–0.7 mg/dL.

Van den Bergh test is specific to identify the increase in serum bilirubin (above reference level). For procedure, see page 115. The results of test are interpreted as:

- Normal serum: Negative reaction,
- Haemolytic jaundice: Indirect positive reaction,
- Obstructive jaundice: Direct positive reaction and
- Hepatic jaundice: Biphasic reaction.

2. Urine bilirubin and bile salts. Normally, urine does not contain bilirubin and bile salts. In liver insufficiency, when the serum bilirubin levels increase above 2 mg/dL the bilirubin is excreted in the urine *(bilirubinuria).*

3. Urine urobilinogen. In normal individuals with urine flow of about 1 mL/min, less than 4 mg of urobilinogen is excreted in the urine. In liver insufficiency, initially there occurs mild increase in the daily excretion of urine urobilinogen. However, in later stages urobilinogen is absent in the urine. This occurs because of the fact that the swollen liver cells block bile canaliculi and prevent excretion of the conjugated bilirubin in the bile.

4. Faecal stercobilinogen

- Normal levels 20–25 mg/dL.
- Increased initially in liver insufficiency producing dark brown stools.
- Decreased in later stages of liver insufficiency producing pale coloured stool.

5. Faecal fat levels

- Normally 5–6% of total fat intake per day in excreted in the faeces.
- In liver insufficiency, fat excretion in faeces increases up to 40–50% of total intake (steatorrhoea). This is because of the fact that due to deficiency of bile salts emulsification and absorption of fat is inadequate in the intestine.

II. Tests to assess metabolic functions of liver

1. Galactose tolerance test. This test is based on the principle that galactose after absorption from the gut gets converted into glycogen in the liver. Therefore, in liver insufficiency its level in the blood rises.

Procedure. In a fasting individual 40 g galactose is administered orally and blood samples are collected at half an hour interval for 2 h. From the blood galactose levels in these samples, then galactose index (GI) is calculated. Galactose index is sum of the four values of blood galactose levels in mg/dL. Its normal value is 68–160 mg/dL. In hepatic insufficiency, GI value markedly increased.

2. Blood glucose level. The normal fasting blood glucose level is 70–90 mg/dL. In hepatic insufficiency its level decreases.

3. Blood and urine amino acids levels. Blood and urine amino acids levels are estimated to assess protein metabolism. In liver damage blood amino acid levels (normal 30–65 mg/dL) and urine amino acid levels are increased.

4. Lipid profile. In hepatic insufficiency, lipid profile is affected as:

- Plasma NEFA and FFA : Increase (normal 10–30 mg/dL)
- Serum cholesterol : Decrease (normal 150–240 mg/dL)
- Serum triglycerides : Decrease (normal 30–150 mg/dL)
- Serum phospholipids : Decrease (normal 150–300 mg/dL)
- Total lipids : Decrease (normal 350–800 mg/dL)
- Ketone bodies : Increase (normal 7–15 mg/dL)

Abnormalities of serum lipid levels are sensitive but nonspecific indicators.

III. Tests to assess synthesis functions of liver

1. Estimation of plasma proteins

- *Albumin.* Liver cell damage causes hypoalbuminaemia (normal 6.4–8.3 g/dL).
- *Globulins.* Hyperglobulinaemia is usually associated with hypoalbuminaemia.
- *A:G ratio.* In liver disorder, there occurs reversal of A:G ratio (normal 1.7:1).

2. Serum levels of liver enzymes

- *Transaminases.* The activity of transaminase like SGPT and SGOT increases in the hepatic insufficiency. Serum SGPT levels are increased by 10–1000 times in an acute phase (normal value < 40 IU%).
- *Alkaline phosphatase.* Alkaline phosphatase is not a liver-specific enzyme, but secreted into the bile. In obstructive jaundice, its level is markedly increased (< 30 KA units).

3. Blood urea. Liver is the main site for urea formation from ammonia. Decreased level of blood urea (normal 20–40 mg/dL) and raised blood ammonia level (normal 20–80 mg/dL) occur in the liver insufficiency.

4. Coagulation factors. Factors II, V VII, IX and X and Vita min K needed to activate these factors are synthesized by the liver. The integrity and activity of these factors is determined by prothrombin time test (PTT). Prolonged prothrombin time (PT) (normal 10–16 s) indicates severe liver disease.

IV. Tests to assess detoxication functions of liver

1. Bromsulphthalein excretion test. BSP is taken up by the liver cells from the blood and detoxified and excreted in the bile. The rate of removal of BSP from the blood depends on the functional efficiency of liver and rate of hepatic blood flow.

V. Tests to assess hepatic cellular integrity

1. Ultrasonography is done to detect diffuse disease of parenchyma of the liver (cirrhosis liver, fatty liver), abscess, cysts, tumours, gall stones and dilatation of biliary system proximal to site of obstruction.

2. Computed tomography (CT). CT scan has the same diagnostic significance as the ultrasonography except that it can also detect even smaller lesions.

3. Radionucleotide imaging. Technetium (^{99m}Tc) is a sulphur colloid, which is easily taken up by the liver cells, monocyte macrophage system (Kupffer's cells) and emits γ-rays. Gamma camera picks up these rays and is used to find out the size of the liver and the lesions of the liver (such as filling defect, diffuse liver disease and portal hypertension).

4. Liver biopsy. This is performed by a special needle passed through intercostal space under local anaesthesia to obtain tissue for histopathological examination.

5. Fine needle aspiration. Very fine needle is usually guided by an ultrasound and material is aspirated for cytological, histopathological and bacteriological examination.

6. Cholecystography is done to assess the functions and diseases of gall bladder. Iodinated compounds given orally are concentrated in the gall bladder and excreted in bile. The gall bladder gets opacified on cholecystography. Therefore, conditions like gall stones, non-functioning gall bladder fail to produce opacification. Nowadays, this test is less commonly performed than ultrasonography.

7. Other elaborate tests include:

- Percutaneous cholangiography,
- Portal venography and
- Endoscopic retrograde cholangiography.

VI. Miscellaneous tests

Serological tests to detect hepatitis viruses, antigens and antibodies.

Chapter 7.5

Physiological Activities in Small Intestine

FUNCTIONAL ANATOMY
- Gross anatomical considerations
- Structural characteristics of small intestine

SMALL INTESTINAL SECRETIONS
- Composition and formation
- Regulation
- Functions

MOTILITY OF SMALL INTESTINE
- Interdigestive period
- Digestive period
- Motility reflexes
  - Gastroileal reflex
  - Intestinointestinal reflex

FUNCTIONS OF SMALL INTESTINE
- Mechanical functions
- Digestive functions
- Absorptive function
- Hormonal functions
- Activator function
- Protective function
- Hydrolytic function

APPLIED ASPECTS
- Paralytic ileus
- Intestinal obstruction

FUNCTIONAL ANATOMY

GROSS ANATOMICAL CONSIDERATIONS

The small intestine is convoluted tube which extends from the pylorus to the ileocaecal valve, where it joins with caecum, the first part of large intestine. It is about 6–7 m in length. It is divided into three parts: the duodenum, the jejunum and the ileum (see Fig. 7.1-1).

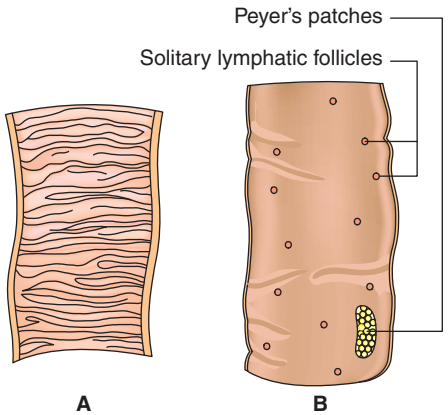


Fig. 7.5-1 The mucosal surface of jejunum, A and ileum, B.

Duodenum. The first and *shortest* part (25 cm long) of the small intestine is also the *widest* and most fixed part. It is C-shaped and for descriptive purposes is divided into four parts: superior (1st) part, descending (2nd) part, horizontal (3rd) part and ascending (4th part). Superior part of the duodenum is also called *duodenal cap* or bulb. It is the region which is struck by the acidic gastric contents when they pass through pylorus and is a *common site for peptic ulcer.* The bile and pancreatic ducts open by a common hepatopancreatic ampulla of Vater on the posteromedial wall of descending (2nd) part of duodenum.

Jejunum and ileum. Jejunum and ileum form respectively, the proximal 2/5th and distal 3/5th of the remaining part of small intestine. There is no sharp demarcation between jejunum and ileum. The inner mucosal surfaces of jejunum and ileum, however, can be differentiated from each other.

STRUCTURAL CHARACTERISTICS OF SMALL INTESTINE

Histologically, the wall of small intestine is made up of four layers, which from within to outwards consist of mucosa, submucosa, muscle coat and serosa (for details see page 452).

Characteristic features of mucous membrane of small intestine

Although the small intestine is about 6 m long, it has an absorptive area of over 250 m^2. This larger surface is created by:

- Numerous folds of the intestinal mucosa *plicae circulares* (Fig. 7.5-1),
- Densely packed villi, which line the entire mucosal surface,
- *Microvilli,* which protrude from the surface of intestinal cells, and the presence of numerous depression (crypts of Lieberkuhn) that invade the lamina propria.

Plica circulares

The mucosal surface shows numerous circular folds *(plicae circulares* or valvulae conniventes) which are absent in first 2 in of the duodenum (Fig. 7.5-1). Each fold is made up of all layers of the mucosa (lining epithelium, lamina propria, and muscularis mucosa). The submucosa also extends into the fold. The circular folds serve following functions:

- Increase surface area for absorption and also slow down the passage of contents through small intestine which facilitate absorption.

Villi

Villi are finger-like projections of mucous membrane seen throughout the length of small intestine (Fig. 7.5-2).

Total number of villi is about 5 million and they are distributed about 20–40 villi/mm^2. Each villus is about 0.5–1 mm long.

Structure. Each villus is covered by a single layer of columnar epithelial cells called enterocytes. The core of each villus contains (Fig. 7.5-3):

- An arteriole, a venule and a lymphatic vessel called lacteal, (which carry the absorbed fats to the thoracic duct), few smooth muscle fibres extending from the muscularis mucosa and
- A fine network of nerves which has connections with submucosal and myenteric plexus.

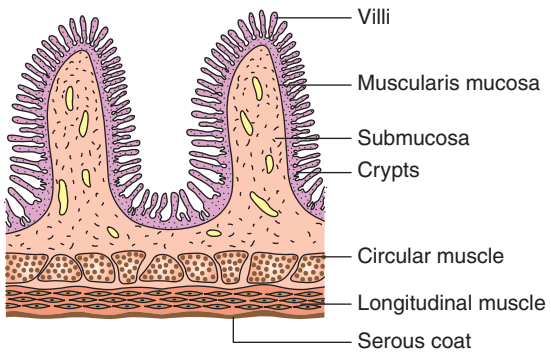


Fig. 7.5-2 Longitudinal section of small intestine showing plica circulares and villi.

Activity. During digestion and absorption, the villi contract quickly with an irregular rhythm and relax slowly. Their muscular fibres serve to pump the lymph from core of villi towards the submucosal lacteals.

The crypts of Lieberkuhn

The crypts of Lieberkuhn are tubular intestinal glands which invaginate deep into the lamina propria, present between the villi throughout the length of small intestine (Fig. 7.5-3). These glands are lined by undifferentiated columnar cells and also contain *goblet cells*, *argentaffin cells* and *Paneth cells*.

Epithelial cells of the mucous membrane of the small intestine and intestinal glands

The mucous membrane of small intestine is made up of following types of epithelial cells:

1. Absorptive columnar cells or the so-called enterocytes. These cells are specialized for the absorptive function. The luminal surface of each enterocyte shows small multiple projections of the cell membrane called the *microvilli* or the *brush border* (Fig. 7.5-3), which increase the surface area to some 30 fold.

2. Undifferentiated columnar cells line the crypts of Lieberkuhn. The cells of the lower parts of the crypts *proliferate actively* by mitosis. The newly formed cells migrate upwards from the crypt to reach the walls of villi.

3. Goblet cells. Fairly, a large number of mucous secreting goblet cells can be seen among the epithelial cells of the mucous membrane. They increase in number in lower parts of intestine, being few in the duodenum and most numerous in the terminal ileum.

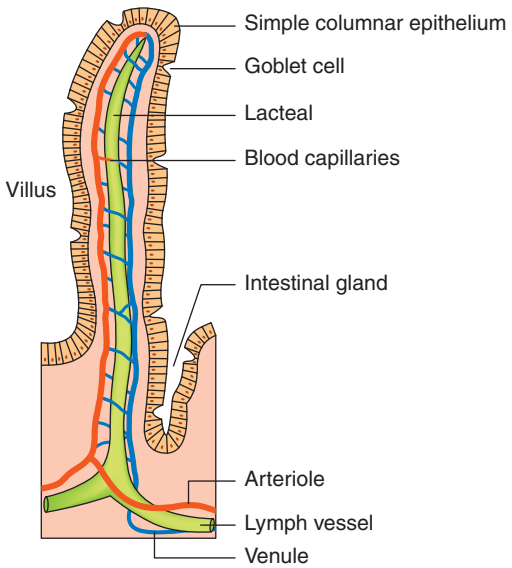


Fig. 7.5-3 Structure of an intestinal villus, crypts of Lieberkuhn and an enterocyte.

4. Argentaffin or enterochromaffin cells are also present in the intestinal mucous membrane, being most numerous near the *lower ends of crypts.* These cells secrete 5-hydroxytryptamine (5-HT, serotonin).

5. Zymogen cells or Paneth cells are found only in the deeper parts of intestinal crypts. They are large acidophilic cells containing secretory granules. They are known to produce lysozyme which destroys bacteria. They may also produce other enzymes.

6. Duodenal glands of Brunner are limited only to the duodenum. These are compound tubuloalveolar glands present in submucosa of duodenum. Their ducts pass through the muscularis mucosa to open into the crypts of Lieberkuhn. They are situated mostly near the pylorus, beyond pylorus their number greatly diminishes. Secretion of these glands contains *mucus and* HCO_3^-, which neutralizes gastric acid entering the duodenum and thus protects its mucosa.

7. Peyer's patches. The ileum contains aggregates of lymphatic follicles known as Peyer's patches (Fig. 7.5-1B).

SMALL INTESTINAL SECRETIONS

COMPOSITION AND FORMATION

The intestinal juice also called *succus entericus* comprises the intestinal secretions which include:

- Aqueous component (water and electrolytes),
- Intestinal enzymes and
- Mucus.

Aqueous component of intestinal juice

Aqueous component of intestinal juice primarily refers to the water and electrolytes secreted by the epithelial cells of small intestine, especially those present in the crypts of Lieberkuhn. About 2 L of secretion is produced per day by these cells, whose chemical composition is almost similar to the extracellular fluid except that it is slightly more alkaline (pH 7.5–8.6). This fluid is colourless, however, becomes slightly cloudy due to admixture of mucus, shedded epithelial cells and cholesterol.

Intestinal enzymes

Brush border of epithelial cells covering the villi contains a large number of intracellular digestive enzymes. The enzymes which have been identified in the brush border are:

1. *Peptidases* (proteolytic enzymes), which digest peptides into amino acid, e.g. aminopeptidases, dipeptidases, nuclease and related enzymes and so on.
2. *Disaccharidases,* such as sucrase, maltase and lactase, which split the respective disaccharidases into the monosaccharides.
3. *Intestinal lipases* that split triglycerides present in small amount.
4. *Enterokinase* or enteropeptidase, which activates trypsinogen to trypsin.

Mucus

Mucus in the small intestine is secreted by:

1. Brunner's glands, which secrete thick alkaline mucoid secretion that serves a protective role, preventing HCl and chyme from damaging the duodenal mucosa.

2. Goblet cells also secrete a lot of mucus, which protects the intestinal mucosa and lubricates the chyme.

REGULATION OF SMALL INTESTINAL SECRETIONS

1. Local stimuli. Mechanical distension of the intestinal mucosa by the food or irritation by chemicals, via local myenteric reflexes, increase the volume and total enzyme output of the small intestine; that is why, the greater is the chyme, greater is the secretion of intestinal secretion.

2. Role of vasoactive intestinal polypeptide (VIP). Though the secretion of the crypts of Lieberkuhn is mainly regulated by the local stimuli, but the local hormone VIP is also reported to increase its secretion.

3. Secretion of Brunner's gland is increased by:

- Vagus stimulation,
- Direct tactile stimulation or irritation of the duodenal mucosa and
- Secretin.

FUNCTIONS OF INTESTINAL JUICE

See functions of small intestine (page 502).

MOTILITY OF SMALL INTESTINE

Motility of small intestine can be described as:

A. *Motility of the small intestine during interdigestive period.*

B. *Motility of the small intestine during digestive period* (related to meals), which includes:

- Mixing movements, such as segmentation contractions and pendular movements,
- Propulsive movements, such as peristaltic contractions and peristaltic rush and
- Movements of villi.

C. *Motility reflexes*, which include:
- Peristaltic reflex,
- Gastroileal reflex and
- Intestinointestinal reflex.

MOTILITY OF SMALL INTESTINE DURING INTERDIGESTIVE PERIOD

Migrating motor complexes

- The migrating motor complex (MMC) is the name given to the peristaltic wave that begins in the oesophagus and travels through the entire gastrointestinal tract during the interdigestive period.
- The MMCs sweep out the chyme remaining in the small intestine.
- The MMCs occur every 60–90 min and last for about 10 min (see page 469).

MOTILITY OF SMALL INTESTINE DURING DIGESTIVE PERIOD

1. Mixing movements

The mixing movements of small intestine are responsible for the proper mixing of chyme with digestive juices like pancreatic juice, bile juice and intestinal juice. The mixing movements of small intestine include:

- Segmentation contractions and
- Pendular movements.

(i) Segmentation contractions

Features
- Segmentation is the most common type of intestinal contractions which occur throughout the digestive period in a rhythmic fashion, and hence also called *rhythmic segmentation contractions.*
- During segmentation contraction, a section of the small intestine (about 2–5 cm) contracts, sending the intestinal contents (chyme) in both oral and caudal directions. That section of the small intestine then relaxes and the contents move back into the segment. At the same time, the adjoining segment which was relaxed, now contracts (Fig. 7.5-4A).
- The alternate contracted and relaxed segments give a ring-like appearance resembling the chain of sausages (Fig. 7.5-4B).

Function. This back-and-forth movement of chyme produced by the segmentation contractions causes thorough mixing without any net forward movement of chyme (Fig. 7.5-4A).

Rate and duration. Segmentation contractions occur about 12 times/min in the duodenum and 8 times/min in the ileum. The contractions last for 5–6 s.

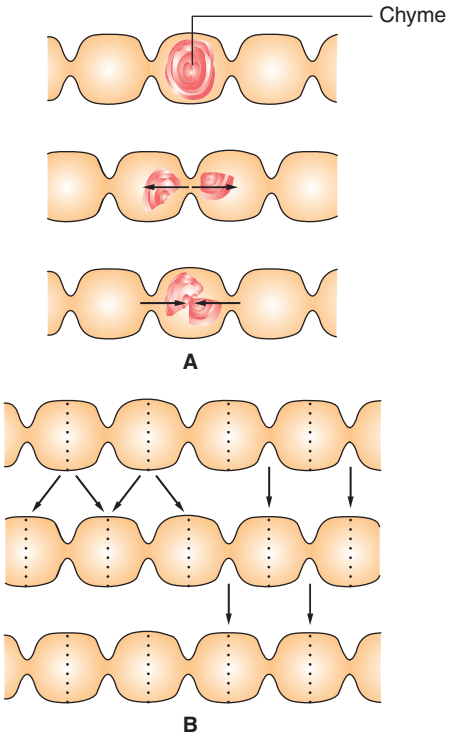


Fig. 7.5-4 Segmentation contraction of small intestine: A, steps of segmentation are shown to reveal back and forth movements of chyme; B, the alternate contracted and relaxed segments give a ring-like appearance resembling the chain of sausages.

Types. Two types of segmentation contractions have been described:

- *Eccentric contractions.* They consist of contraction located in a localized segment less than 2 cm in length and are eccentric in appearance. They are mainly due to contraction of outer longitudinal smooth muscle layer.
- *Concentric contractions.* The segments are usually longer than 2 cm and are of relatively uniform circumference. They are mainly due to contraction of inner circular smooth muscle layer.

Control
- *Initiation.* Segmentation contractions can occur only if the slow waves (basal electrical rhythm, i.e. BER) produce spikes, or action potentials. The slow waves (BER) are initiated by the pacemaker cells located in the second part of duodenum near the entry of common bile duct.
- *Frequency* of segmentation contractions is directly related to the frequency of slow waves and is thus controlled by pacemaker cells within the wall of small intestine and is not influenced by the neural activity or circulating hormones.
- *Strength* of segmentation contractions is proportional to the frequency of spikes generated by slow waves which in turn is controlled by the amplitude of slow waves. The slow wave amplitude and thus the strength

of segmentation contraction is controlled by the hormones released during digestion:

- Slow wave amplitude is increased by gastrin, cholecystokinin (CCK), motilin and insulin.
- Slow wave amplitude is decreased by secretin and glucagon.

(ii) Pendular movements

These are small constrictive waves which sweep forward and backward or upward and downward in a pendular fashion. These mixing movements can be noticed only by close observation.

(iii) Tonic contractions

These are the variant of segmental contractions which last for somewhat longer time. During these contractions, one segment of the intestine is isolated from the other, which permits longer contact of the chyme with enterocytes and thus facilitates absorption.

2. Propulsive movements

The propulsive movements of small intestine are involved in pushing the chyme towards the aboral end of intestine. These include:

- Peristaltic contractions and
- Peristaltic rush.

(i) Peristaltic contractions

Characteristic features. The peristaltic contractions are highly co-ordinated and typically a peristaltic contraction involves contraction of a segment behind the bolus and simultaneous relaxation of the segment in front of the bolus, causing the chyme to be propelled caudally (Fig. 7.5-5).

Each contraction travels for a variable but short distance and then dies out. A new contraction is then initiated from a site little distal to the site of origin of the previous contraction. In this way, a continuous peristaltic wave is set up in the intestine. Several of these wave-like contractions occur simultaneously along the length of the intestine, resulting in vermiform movements (worm-like movements).

Law of intestine. The peristaltic waves always travel from the oral end towards the aboral end of the intestine. This phenomenon has been labelled as the *Law of the intestine* by Starling in 1901. The other names which have been given to this phenomenon are: 'Polarity of intestine' 'Polar conduction of intestine', 'Electrical activity of intestine', 'Law of gut' and 'Theory of receptive relaxation'.

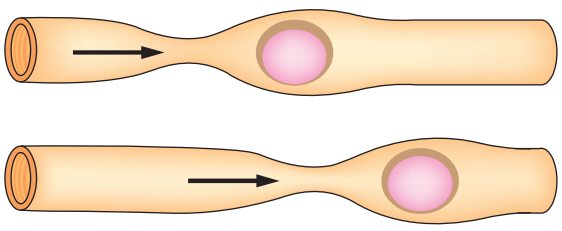


Fig. 7.5-5 Peristaltic contraction moves the food through intestine by pushing bolus ahead of muscle contraction.

Functions subserved by the peristaltic waves are:

- Help to propel the intestinal contents aborally.
- Also help in digestion and absorption of the food particles because different types of nutrients are digested and absorbed in different segments of the small intestine.

Control of peristaltic contractions. The co-ordinated peristaltic activity is dependent on the integrity of enteric nerve plexus. The usual stimulus for peristalsis is distension. This response to stretch is called *myenteric reflex.* The local stretch releases serotonin, which activates sensory neurons that stimulate the myenteric plexus.

IMPORTANT NOTE

Activity of the myenteric plexus from a stimulus point travels in either direction to activate neurons that release:

- *Acetylcholine and substance P* above the point of stimulus, producing a circular constriction and
- *Nitric oxide, VIP and ATP* below the point of stimulus producing receptive relaxation.

(a) Neural control. Though the peristaltic contractions can occur in the absence of extrinsic innervation, but their magnitude is affected by neural influences:

- Parasympathetic stimulation increases intestinal motility through vagus as seen during strong emotions.
- Sympathetic stimulation decreases intestinal movements as seen during anger and pain.

(b) Hormonal control. Certain hormones also affect the magnitude of peristaltic contraction:

- Intestinal motility is enhanced by gastrin, CCK, 5-HT, thyroxine and insulin.
- Intestinal motility is decreased by secretin and glucagon.

(ii) Peristaltic rush

- Peristaltic rush refers to a very powerful peristaltic contraction which occurs when intestinal mucosa is irritated intensely as in some infectious processes.
- This type of powerful contraction begins in duodenum and passes through entire length of small intestine and finally reaches the ileocaecal valve within few minutes. Thus, they sweep the contents of intestine into the colon, thereby relieving the small intestine of irritant or excessive distension, as it occurs in cases of diarrhoea.

3. Movements of villi

Movements of villi consist of alternate shortening and elongation of the villi caused by contraction and relaxation of

the muscle. Villikinin, a hormone secreted from the small intestinal mucosa is believed to play an important role in increasing the movements of villi.

Functions. Movements of villi help in emptying lymph from the central lacteal into the lymphatic system. The surface area of villi is increased during elongation. This helps in absorption of digested foodstuffs from the lumen of intestine.

MOTILITY REFLEXES

1. Gastroileal reflex

- Gastroileal reflex refers to a marked increase in the peristaltic contractions of ileum associated with relaxation of ileocaecal sphincter which occur immediately after the meals. As a result, the intestinal contents are delivered to the large intestine.
- This reflex is initiated by the distension of stomach by the food.
- The peristaltic contractions are caused by reflex stimulation of vagus and the relaxation of ileocaecal sphincter seems to be produced by the hormone gastrin.

2. Intestinointestinal reflex

- Intestinointestinal reflex refers to the relaxation of smooth muscles of the rest of the small intestine in response to overdistension of one segment of the intestine.

FUNCTIONS OF SMALL INTESTINE

The functions of small intestine can be summarized as:

1. Mechanical functions. The mixing and propulsive movements of the small intestine help in thorough mixing of chyme with the digestive juices (pancreatic juice, bile juice and succus entericus) and propel it towards the large intestine.

2. Digestive functions of small intestine are carried out by the digestive enzymes present in the succus entericus (see page 499), pancreatic enzymes (see page 482) and bile (see page 489).

3. Absorptive function is accomplished by the huge surface area created by the presence of plicae circulares, villi and microvilli. The end products of digestion of carbohydrates, proteins and fats are absorbed through portal system or through the lymph. For details of absorption (see page 511).

4. Hormonal functions. The small intestine secretes certain hormones which exert their effect on the secretions and motility of gastrointestinal tract. These hormones include *enterogastrone, secretin* and CCK.

5. Activator function. The enzyme enterokinase secreted by small intestine activates trypsinogen into trypsin, which in turn activates other enzymes.

6. Protective function. The mucus secreted into the succus entericus protects the intestinal wall from the gastric acid chyme.

7. Hydrolytic function. The aqueous component of the succus entericus provides water and thus helps in all the hydrolytic processes of enzymatic reactions of digestion of various food particles.

APPLIED ASPECTS

Paralytic ileus

Paralytic ileus or the adynamic ileus refers to a condition in which the intestinal motility is markedly decreased leading to retention of its contents (because the contents cannot be propelled into the colon). This produces irregular distension of the small intestine by pockets of gas and fluid.

Causes. Paralytic ileus may occur due to:

1. Direct inhibition of smooth muscles of small intestine due to handling of intestine:
- During intra-abdominal operations.
- During trauma.

2. Reflex inhibition of smooth muscles of small intestine due to increased discharge of noradrenergic fibres in splanchnic nerves, as seen in irritation of peritoneum (in patients with peritonitis and injury to peritoneum).

Intestinal obstruction

Causes. Obstruction of the lumen of small intestine may occur due to many causes, such as tumours, strictures and fibrotic bands in the abdomen.

Features. The intestinal obstruction is characterized by:

- Intestinal colic, i.e. severe abdominal pain. Pain is caused by the peristaltic rush (intense peristaltic wave initiated due to irritation of intestinal mucosa at the site of obstruction).
- Stimulation of visceral afferent nerves by the increased intraluminal pressure may cause sweating, hypotension and severe vomiting.

Chapter 7.6

Physiological Activities in Large Intestine

FUNCTIONAL ANATOMY

GROSS ANATOMICAL CONSIDERATIONS

Functional organization

The large intestine is a tube about 6 cm in diameter and 100 cm in length. It normally arches around and encloses the coils of small intestine and tends to be more fixed than the small intestine. It is divided into following parts (Fig. 7.6-1):

- *Caecum* is a blind-ended sac into which opens the lower end of ileum. The ileocaecal junction is guarded by the ileocaecal valve which allows inflow but prevents back-flow of the intestinal contents.
- *Appendix* is a worm-shaped tube that arises from the medial side of caecum, which in human being is a vestigial organ.
- *Ascending colon* extends upward from the caecum along the right side of abdomen up to the liver. On reaching the liver it bends to the left, forming the right hepatic flexure.
- *Transverse colon* extends from the right hepatic flexure to the left splenic flexure.
- *Descending colon* extends from the left splenic flexure to the pelvic inlet below.
- *Sigmoid colon* begins at the pelvic inlet as a continuation of the descending colon and joins the rectum in front of the sacrum.
- *Rectum* descends in front of the sacrum to leave the pelvis by piercing the pelvic floor. Here it becomes continuous with anal canal in the perineum.
- *Anal canal* opens to the exterior through the anus, the opening which is guarded by two sphincters.

Ileocaecal valve

Structure. Ileocaecal valve functioning occurs due to the invagination of ileum into the caecum at the ileocaecal junction and a very small ileal opening (only 2–3 mm in diameter).

Functions. The principal function of the ileocaecal valve is to prevent back flow of the faecal matter from the caecum into ileum. The valvular mechanism works in such a way that when the caecal pressure is increased the ileocaecal opening is closed.

Role of ileocaecal sphincter. Ileocaecal sphincter refers to a thickened band of circular muscle coat of the terminal part of ileum just above the ileocaecal junction. The rhythmic contractions of ileocaecal sphincter leading to rhythmic opening and closing occur after every 30 s after a meal. During every rhythmic opening, a small jet of ileal fluid (approximately 15 mL) escapes into the caecum. The ileocaecal sphincter slows down the emptying of ileal contents into the caecum and thus, helps in the completion of the absorption of nutrients in the ileum.

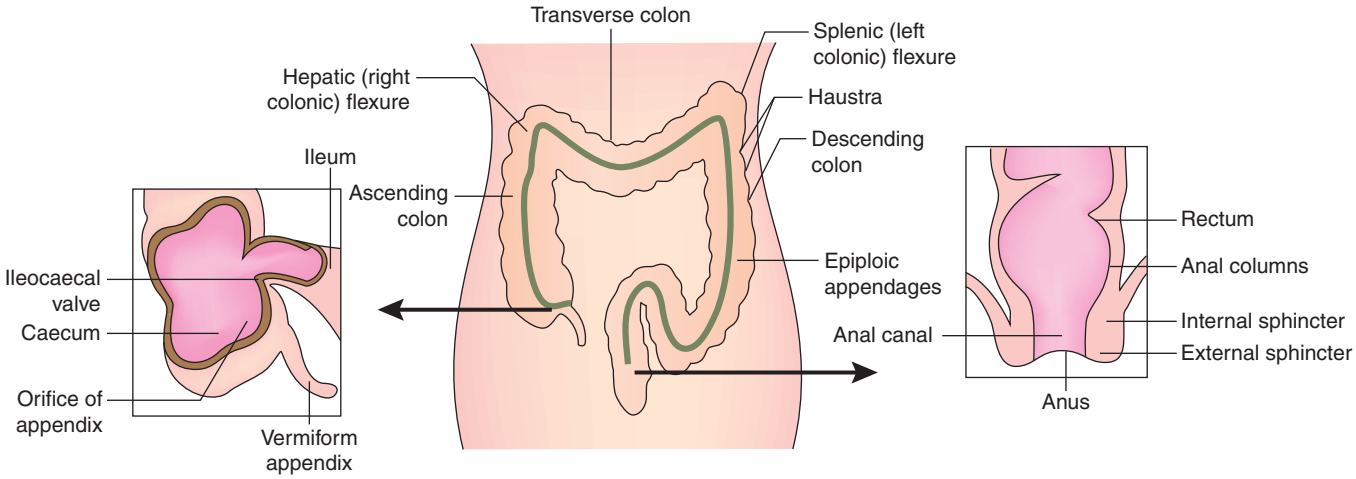


Fig. 7.6-1 Functional organization of the large intestine.

Control. Gastrin produces relaxation and secretin causes contraction of the ileocaecal sphincter.

Note. It is important to note that these hormones show opposite effects on cardiac sphincter.

STRUCTURAL CHARACTERISTICS

Histological structure of large intestines is similar to that described in general (see page 452) with following special characteristics. Mucosa of large intestine is characterized by:

- Absence of plica circulares and villi (seen in small intestine).
- A large number of simple tubular glands (crypts of Lieberkuhn) lined by simple columnar epithelial cells with large number of goblet cells which secrete mucus (Fig. 7.6-2). Epithelial cells contain no enzyme.
- The epithelium overlying solitary lymphatic follicles (in ascending colon, caecum and appendix) contains M-cells similar to those seen in the small intestine. Microfold (M) cells are specialized epithelial cells overlying the Peyer's patches.

Longitudinal layer of muscle coat of colon is unusual. Most of the fibres in it are collected to form three thick bands, the taenia coli which can be seen through the serous layer (Fig. 7.6-1). A thin layer of longitudinal fibres is present in the intestines between the taenia.

The taenia coli are shorter in length than other layers of the wall of colon. This results in the production of *sacculations* (also called *haustrations*) on the wall of colon.

Serous layer is missing over the posterior aspect of the ascending and descending colon. At places small peritoneal bags of fat, called appendices epiploicae, project from the colonic serosa.

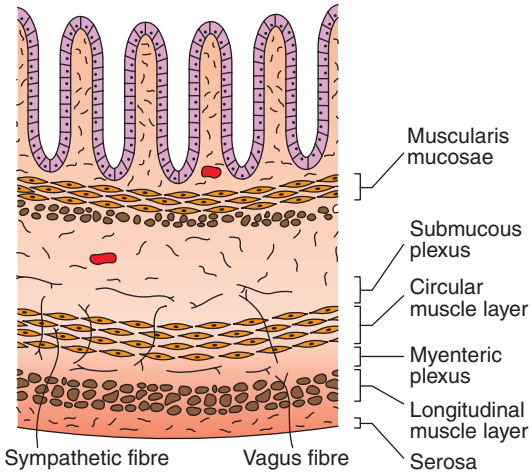


Fig. 7.6-2 Histological structure of the colon.

LARGE INTESTINAL SECRETIONS AND BACTERIAL ACTIVITY

LARGE INTESTINAL SECRETIONS

- The large intestinal secretions mainly comprise mucus secreted by the goblet cells; and some water and lot of HCO_3^- are secreted by the glands of Lieberkuhn.
- The mucus lubricates the faecal matter and also protects the mucous membrane of the large intestine by preventing the damage caused by the mechanical injury or chemical substances.
- The alkaline nature (pH 8.0) of the mucoid secretions of the large intestine is due to the presence of HCO_3^-. It serves to neutralize the acids formed by the bacterial action on the faecal matter.
- Large quantities of water and electrolytes are secreted by the mucosa of large intestine only when it is intensely irritated.

INTESTINAL BACTERIAL ACTIVITY

Bacterial flora. At birth, the colon is sterile, but the colonic bacterial flora becomes established early in life and includes:

- *Harmless bacteria,* such as *E. coli* and *Enterobacter aerogenes* and
- *Potentially dangerous bacteria,* such as *Bacteroides fragilis,* various types of cocci and gas gangrene bacilli. These bacteria can cause serious disease in tissues outside the colon.

Intestinal bacterial activities can be grouped as:

- Beneficial bacterial activities,
- Indifferent bacterial activities and
- Detrimental bacterial activities.

Beneficial bacterial activities include:

- *Synthesis of vitamins,* such as vitamin C, a number of B-complex vitamins and folic acid.
- *Trophic effects on colonic mucosa.* Unabsorbed carbohydrates are converted to short-chain fatty acids by colonic bacteria. Some of the short-chain fatty acids produced have a trophic effect on the colonic mucosa.
- *Play a role in cholesterol metabolism* by decreasing plasma cholesterol and LDL levels.

Indifferent bacterial activities include:

- *Production of intestinal gases.* The colonic bacteria produce gas in large volumes up to 7–10 L/day, which contribute towards flatus. The gas is produced chiefly through breakdown of undigested nutrients that reach the colon. The gases produced by colonic bacteria include carbon dioxide (CO_2), hydrogen sulphide (H_2S), hydrogen (H_2) and methane (CH_4) which contribute to flatus. Nitrogen gas (N_2) derived from the swallowed air accounts for most of the flatus passed through rectum, or other gases diffuse readily through the intestinal mucosa. Therefore, the volume of flatus expelled is reduced to about 600 mL/day.
- The absorption of protein antigens (bacterial or viral protein) occurs through the M cells. The M cells pass on antigens to the lymphoid cells which respond by the secretion of IgA antibodies. Thus secretory immunity plays an important role in localized protection of the intestinal mucosa.
- *Organic acids* formed by the colonic bacteria from the carbohydrates are responsible for the slight acidic reaction of the stools (pH 5–7).
- *Substances responsible for the odour of the faeces,* such as indole, skatole, mercaptans are synthesized by colonic bacteria.
- *Pigments formed* by the colonic bacteria from the bile pigments are responsible for the known colour of stools.

Detrimental bacterial activities include:

- *Consumption of nutrients* like vitamin C, vitamin B_{12} and choline by some bacteria may lead to deficiency symptoms, unless these are supplemented in adequate amounts in the diet.
- *Production of ammonia.* Colonic bacteria also produce ammonia, which is absorbed by blood and is normally detoxified quickly by the liver. However, in liver dysfunction, hyperammonaemia results, producing neurological symptoms (hepatic encephalopathy).

MOTILITY OF LARGE INTESTINE

SLOW WAVE ACTIVITY

Like other parts of gastrointestinal tract, the motility of large intestine is also co-ordinated by the 'basal electrical rhythm' or the so-called 'slow wave activity'. However, the frequency of slow wave activity gradually increases down the large intestine (from 9/min at the ileocaecal valve to 16/min at the sigmoid colon).

MOVEMENTS OF LARGE INTESTINE

Functions

The contractile activity of the large intestine serves two main functions:

- It increases the efficiency of colon for water and electrolyte absorption.
- Promotes the excretion of the faecal matter remaining in the colon.

Types of movements

The different types of movements (most accepted nomenclature) of colon are:

1. Haustral shuttling

- The haustral shuttling or haustral contractions are similar to the segmentation contractions of small intestine, which vigorously mix the contents of colon and, by exposing more of the contents to mucosa (facilitate absorption).
- Contraction of circular and longitudinal muscles in the large intestine cause haustrations to develop as:
  - Contraction of circular muscle produces constriction rings at regular intervals,
  - Contraction of longitudinal muscle (taenia coli) causes the unstimulated portion of large intestine in between the constriction rings to bulge in bag-like sacs called haustration.
  - Contraction disappears within 60 s. After a few minutes, haustral contractions are initiated in a nearby area. The dynamic formation and disappearance of

haustrations squeeze the chyme, moving it back and forth in a manner similar to that described for the segmentation contractions in the small intestine.

2. Peristalsis

Peristalsis is a progressive contractile wave preceded by a wave of relaxation. In the colon, the peristalsis waves are very small pressure waves of prolonged duration.

Function. They propel the contents towards rectum very slowly (5 cm/h). It can take up to 48 h for the chyme to traverse the colon.

3. Mass movements

- The mass movements are special type of peristaltic contractions which are observed in the colon only.
- These occur 3–4 times a day generally after meals and each contraction lasts for about 3 min.
- The mass movements force the faecal material rapidly in mass down the colon. They also move material into the rectum and rectal distension initiates the defaecation reflex.
- A mass movement can be initiated by:
  - Gastrocolic or duodenocolic reflexes,
  - Intense stimulation of the parasympathetic nerves or
  - Overdistension of a segment of colon.

Gastrocolic reflex

Gastrocolic reflex refers to the contraction of colon induced by entry of food into the stomach. This reflex results in an urge to defaecate after a meal. Because of this, defaecation after meals is a rule in children. However in adults, the bowel training suppresses this reflex.

Initiation. It has been reported that perhaps this reflex consists of two phases:

- *The early* or rapid phase (which occurs within 10 min of meals) is initiated by distension of stomach and is conducted through the extrinsic nerves of the autonomic nervous system. It can be abolished by anticholinergic drugs.
- *The late* or slow phase is considered to be mediated by gastrointestinal hormones like gastrin and cholecystokinin, which are secreted into the blood stream in significant amounts shortly after a meal.

Transit time in the gut

The transit time in various parts of the gut, studied after a test meal is:

- Up to *caecum* – 4 h
- Up to *hepatic flexure* – 6 h
- Up to *splenic flexure* – 8–9 h
- Up to *pelvic colon* – 12 h
- From *pelvic colon to anus,* transport is made slower and as much as a quarter (25%) of the residue of a test meal may still be in the rectum for up to 3 days.
- Complete expulsion of the meal in stool takes more than a week.
- It has been observed that a high residue diet passes more rapidly through the entire gut. This is mainly because of its effect on the colonic movements.

DEFAECATION

FUNCTIONAL ANATOMY

A brief description of functional anatomy of the anal sphincters, which play most important role in the process of defaecation. Features of anal sphincters are depicted in Table 7.6-1.

THE ACT OF DEFAECATION

Defaecation, the process of excretion of faecal material, involves both voluntary and reflex activity. The events

Table 7.6-1 Functional anatomy of the anal sphincters

	Internal or involuntary anal sphincter	External or voluntary anal sphincter
Muscle type	Formed by thickened circular smooth muscle	Formed by somatic skeletal muscle
Nerve innervation	• Parasympathetic (pelvic splanchnic) nerves • Sympathetic nerves	Pudendal nerves which maintain the sphincter in a state of tonic contraction
Stimulation	• Relaxes by reflex in response to stimulation of stretch receptors in the rectum wall. When the rectum is sufficiently distended by faeces, the internal anal sphincter relaxes through innervation by the pelvic nerve	• Mild to moderate distension of the rectum increases its force of contraction • Moderately severe distension of the rectum will initiate a reflex which inhibits the discharge of somatic pudendal nerves to cause sphincter relaxation. Therefore, the sphincter can be voluntarily relaxed

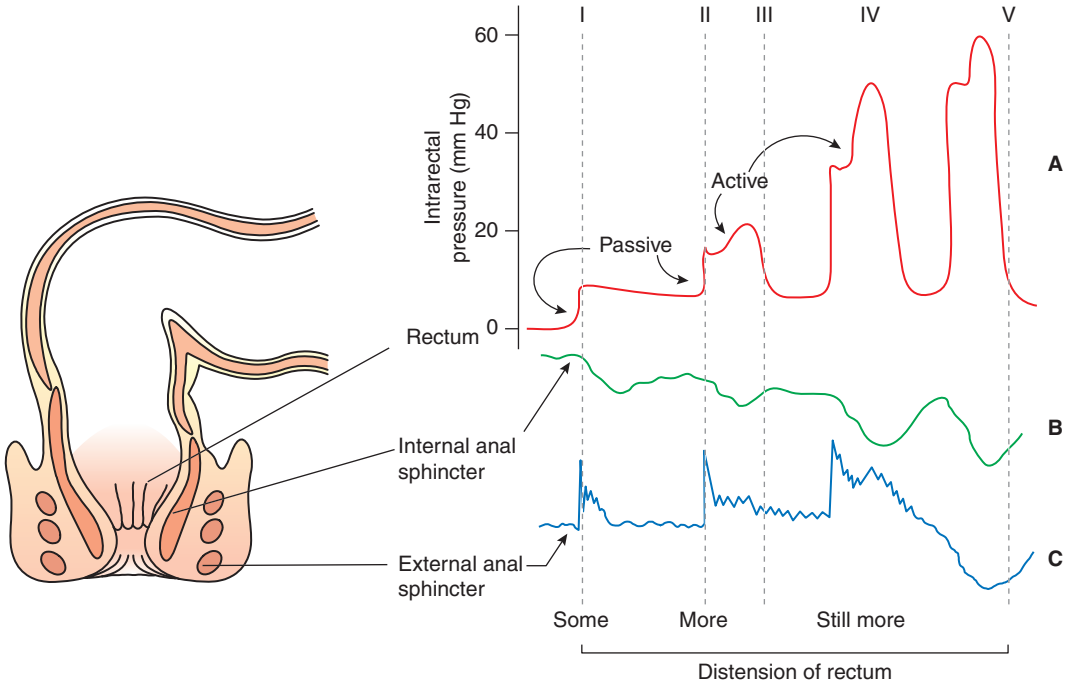


Fig. 7.6-3 Changes in: A, intrarectal pressure; B, tone of internal anal sphincter and C, tone of external anal sphincter during distension of rectum by the faeces.

associated with the process of defaecation proceed as follows:

Distension of rectum

Usually, once or twice a day gastrocolic reflex drives the faeces into the rectum, which increase the intrarectal pressure passively (Fig. 7.6-3A I & II).

Defaecation reflexes

As the rectum starts filling, the resultant rise in the intrarectal pressure stimulates the stretch receptors, sets up defaecation reflexes and produces an urge to defaecate (when intrarectal pressure increases to about 18 mm Hg). The voluntary external anal sphincter which normally remains tonically contracted further contracts when there is moderate rise in the rectal pressure (Fig. 7.6-3C I & III).

Intrinsic reflex. It is mediated by an intrinsic nerve plexus. Distension of rectum with faeces initiates afferent signals that spread through the myenteric plexus and

- Initiates peristaltic waves in the descending colon, sigmoid colon and rectum causing the active contraction of smooth muscles and further raising intrarectal pressure thus forcing the faeces towards the anus (Fig. 7.6-3A III & IV).
- Relaxation of internal anal sphincter occurs by inhibitory signals from the myenteric plexus, when the peristaltic wave approaches the anus (Fig. 7.6-3B).

The intrinsic defaecation reflex functioning by itself is relatively weak. To be effective in causing defaecation, this reflex usually fortified by a spinal cord reflex.

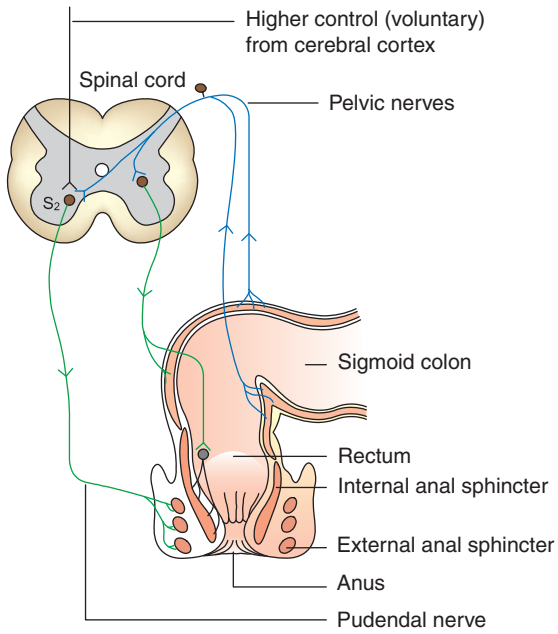


Fig. 7.6-4 Pathway of spinal defaecation reflex and its voluntary control.

Spinal cord reflex. Distension of rectum by faeces causes transmission of afferent impulses through the pelvic nerves to sacral segments of spinal cord. This induces reflex parasympathetic discharge (mainly from S_2) and the pelvic splanchnic nerves (Fig. 7.6-4) to cause:

- Intensification of colonic peristaltic contraction further raising the intrarectal pressure (Fig. 7.6-3A-V),

- When the rectal pressure reaches to about 55 mm Hg there occurs,
- Further relaxation of internal anal sphincter and relaxation of external sphincter as well (Fig. 7.6-3C-V).

Role of voluntary control on defaecation

Once the above described reflex effects are obtained, the voluntary control mechanism depending upon the convenience may or may not allow the act of defaecation to occur:

When defaecation is not allowed, the voluntary control mechanism maintains the contraction of external anal sphincter (which is composed of skeletal muscle innervated by the pudendal nerves). Soon, the internal anal sphincter also closes and the rectum relaxes to accommodate the faecal matter within it. Once the defaecation reflex dies out, it recurs after some hours.

When it is convenient to defaecate

- The external anal sphincter is relaxed voluntarily. Thus both internal and external sphincters are relaxed.
- The intra-abdominal pressure is increased by the contraction of abdominal and diaphragmatic muscles (a process of expiring against closed glottis, i.e. Valsalva manoeuvre).
- The smooth muscles of the distal colon and rectum contract forcibly, propelling the faecal matter out of the body through the anal canal.

Voluntary initiation of defaecation. As per convenience, before the pressure that relaxes the external anal sphincter is reached (i.e. below 55 mm Hg but above 18 mm Hg), the defaecation can be voluntarily initiated. This is done by voluntarily relaxing the external sphincter and contracting the abdominal muscles (straining); thus aiding the reflex emptying of distended rectum.

APPLIED ASPECTS

Applied aspects of defaecation

Defaecation in infants. In infants, defaecation reflex causes automatic emptying of lower bowel without normal voluntary control on external anal sphincter. The voluntary control of the reflex by higher centres is attained by social training as the child grows.

Defaecation in individuals with spinal cord transection. In individuals with spinal cord transection, initially there occurs retention of faeces. But defaecation reflex returns quickly. However, reflex evacuation occurs automatically, without voluntary control, when the rectal pressure increases to about 55 mm Hg.

Role of dietary fibres. Dietary fibres increase bulk of faeces, this plays a role in defaecation reflex by distending the rectum.

FAECES

Composition. Faeces or the faecal matter is derived mainly from the intestinal secretion and partly from the undigested material. The faecal matter consists of: *water*, forms the main bulk of faeces (75%) and *solids*, contribute 25% to total faecal matter weight. These include inorganic material, mostly calcium and phosphate, undigested plant fibres, epithelial cells, dead bacteria, constituents of intestinal secretions including bile pigments, fats and proteins. It is important to note that:

- *Proteins* in the stools are not of dietary origin but comes from bacteria and cellular debris.
- *Fats* in the stools come some from the dietary intake but most of it is also derived from the desquamated epithelial cells and from the bacterial synthesis. On an average intake of (about 100 g/day) fat, only 5–6 g is lost in faeces.

pH of stools is slightly acidic (5–7) due to the organic acids formed from the carbohydrates by colonic bacteria.

Brown colour of stools is due to the pigment *urobilin*, which is formed from oxidation of urobilinogen which is colourless. Urobilinogen is formed from the bile pigments by the intestinal bacteria. Oxidation of residual urobilinogen in the stools accounts for the darkening of faeces, which occurs upon standing in the air. When the bile fails to enter the intestine, stools become white (*acholic stools*), as seen in obstructive jaundice.

Odour of stools is due to the presence of substances like indole, skatole, mercaptans and hydrogen sulphide. These substances are formed by the action of colonic bacteria on the food.

FUNCTIONS OF LARGE INTESTINE

The functions of large intestine can be summarized as:

1. Secretory functions. The large intestinal secretion mainly comprises mucin, which helps to lubricate the faecal matter. The alkaline nature (pH 8) of the secretion serves to neutralize the acids formed by the bacterial action on the faecal matter.

2. Synthesis functions. The bacterial flora of the large intestine synthesizes folic acid, vitamin B_{12} and vitamin K.

3. Absorptive functions. Absorption of water and electrolytes is the chief function of proximal part of the colon. Organic substances like glucose, alcohol and some drugs like anaesthetic agents, sedatives and steroids can also be absorbed in large intestine. The vitamin K and a number of B-complex vitamins which are synthesized in colon by bacterial flora are also absorbed in the large intestine.

4. Excretory functions. Heavy metals like mercury, lead, bismuth and arsenic are excreted by large intestine through the faeces.

APPLIED ASPECTS

ROLE OF DIETARY FIBRES

Physiological role of dietary fibres on intestinal food transit

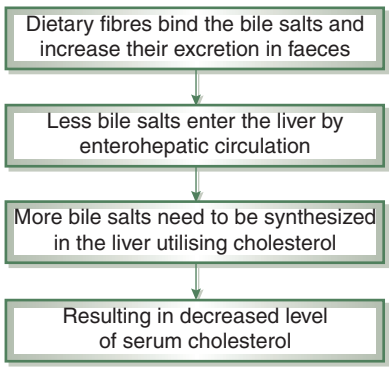
Dietary fibres are constituted by the cellulose, hemicellulose and lignin components of the vegetable products in diet.

- In human beings, there is no appreciable digestion of the dietary fibres, at all. The ingested dietary fibres reach the large intestine in an essentially unchanged state and thus add bulk to the faeces, and thus play a role in defaecation reflex by distending the rectum.

Role of dietary fibres in prevention of diseases

Epidemiological evidences indicate that groups of people who consume a diet which contains large amounts of vegetable fibres have a low incidence of diverticulitis, cancer of colon, diabetes mellitus and coronary artery disease. Probably, the dietary fibres might be playing role by their following effects:

- *Reduction in absorption of digested foodstuffs* is caused by dietary fibres by forming a mechanical barrier between the nutrients and absorptive surface. Due to this effect, the dietary fibres reduce chances of post-prandial hyperglycaemia and are thus especially useful in diabetics.
- *Reduction in blood cholesterol level* by dietary fibres is caused by increasing excretion of bile salts in faeces as summarized:



Therefore, dietary fibres are especially useful in the patients with atherosclerosis, obesity, hypercholesterolaemia and diabetes mellitus.

Therapeutic role of dietary fibres

The daily recommended intake of dietary fibres is about 25–35 g/day. High-fibre supplements have therapeutic role in following conditions:

- *In constipation*, the dietary fibres work as bulk laxatives by providing a larger volume of indigestible material to the colon. *Plantago lanata* or isabgol, rich in hemicellulose, is being used since ages as ancient Indian medicine for constipation.
- In *spastic colon* and *diverticular disease*, the dietary fibres are useful by making the stools softer and thus lowering the intraluminal pressure.
- *In diabetes and high cholesterol levels*, the role of dietary fibres have already been discussed.
- *In diarrhoea*, *complete avoidance of dietary fibres* is useful by increasing the transit time, decreasing the frequency and volume of stools.

DISORDERS OF LARGE INTESTINE MOTILITY

1. Hirschsprung's disease

Hirschsprung's disease, or the aganglionic megacolon, refers to the congenital absence of Auerbach's plexus in the wall of rectosigmoid region. This leads to the blockage of both the peristalsis and mass contractions at the aganglionic segment. Therefore, the faeces pass the aganglionic segment with difficulty and accumulate in the large intestine leading to dilatation of the colon (megacolon).

2. Constipation

Constipation refers to the failure of voiding of faeces which produces discomfort. It results from infrequent mass movement in the colon. As a result, the faecal matter remains in the colon for longer time, so a large amount of fluid is absorbed and the faeces become hard and dry.

3. Diarrhoea

Diarrhoea is a condition which is characterized by an increase in frequency of defaecation with increased water content of the faeces.

Chapter 7.7

Digestion and Absorption

DIGESTION AND ABSORPTION OF CARBOHYDRATES

DIETARY CARBOHYDRATES

Dietary intake of carbohydrates is 250–850 g/day which represents 50–60% of the diet. Major carbohydrates in the human diet are present in following forms:

1. Polysaccharides. These may be present in following forms:
- Starch is the carbohydrate reserve of plants.
- *Glycogen.* It is available in non-vegetarian diet and so often referred to as animal starch. It has glucose molecules which are mostly long chain (1:4α linkages and 1:6α linkages) at branching points.
- *Cellulose* (plant polysaccharide), which is present in diet in large amounts. But there is no enzyme in the human gastrointestinal tract (GIT) to digest it.

2. Oligosaccharides. Based on the number of monosaccharide units present, oligosaccharides are further subdivided into di, tri, tetra and pentasaccharide.

Disaccharides include:
- Sucrose (glucose + fructose) is also known as table sugar (cane or beet sugar).
- Lactose (glucose + galactose) is also called milk sugar.
- Maltose (glucose + glucose). It is a product of starch hydrolysis. It is present in germinating seeds.

3. Monosaccharides. Monosaccharides consumed mostly in human diet are:

Hexoses such as:
- Glucose (in fruits, vegetables and honey) and
- Fructose in fruits.

Pentoses do not occur in free form, but are found in nucleic acid and in certain polysaccharides, such as pentosans of fruits and gums.

Other carbohydrates, which may be present in the human diet are alcohol, lactic acid, pyruvic acid pectin, dextrin and minor quantities of carbohydrate derivatives in the meat.

DIGESTION OF CARBOHYDRATES

The digestion of carbohydrates begins in mouth, continues in stomach but occurs mainly (almost all) in the small intestine.

Digestion of carbohydrates in the mouth

Initial starch digestion starts in the mouth by the enzyme α-amylase (ptyalin) present in the saliva. *α-amylase* present

in the saliva acts on the 1–4 linkages (but not on 1–6 linkages). It digests cooked starch to maltose.

Digestion of carbohydrates in the stomach

In the stomach there occurs minimal carbohydrates digestive activity. α-amylase (which enters the stomach with food) activity continues in the stomach for 20–30 min till the highly acidic gastric juice mixes with the food and makes it inactive. The optimum pH for the action of salivary amylase is 6–7 and its activity in the stomach completely stops when pH falls below 4.

The HCl of the gastric juice may hydrolyse some sucrose.

Digestion of carbohydrates in the small intestine

In the small intestine the carbohydrates are digested by:

Pancreatic α-amylase is present in the pancreatic juice which is poured into the duodenum acts on boiled as well as unboiled starch and variety of other carbohydrates except cellulose. Pancreatic amylase acts in an alkaline medium and its digestive activity is increased by the presence of bile salts. It converts the starch (polysaccharides) into oligosaccharides, such as maltose, maltotriose and dextrin.

$$\underset{\text{(e.g. starch and glycogen)}}{\text{Polysaccharides}} \xrightarrow{\text{Pancreatic amylase}} \underset{\text{(e.g. maltose, dextrin etc.)}}{\text{Oligosaccharides}}$$

Brush border enzymes of small intestine. The carbohydrate splitting brush border enzymes of small intestine include *dextrinase, maltase, sucrase* and *lactase.* These brush border enzymes digest the oligosaccharides into monosaccharides on the surface of epithelial cells of villi as below:

- *α-limiting dextrinase.* It is the only enzyme in GIT, which attacks 1,6 α-glycoside linkage, at the branching points of α-limit dextrins. It also attacks 1,4 α-glycosidic linkages resulting in a sequential removal of glucose monomers from the dextrins (the breakdown products of starch by the enzyme amylase).

$$\text{Dextrin} \xrightarrow[\text{dextrinase}]{\alpha\text{–limiting}} \text{Glucose}$$

- *Maltase, sucrase and lactase* hydrolyse the corresponding disaccharides into monosaccharides as below:

$$\text{Maltose} \xrightarrow{\text{Maltase}} \text{Glucose}$$
$$\text{Sucrose} \xrightarrow{\text{Sucrose}} \text{Glucose} + \text{Fructose}$$
$$\text{Lactose} \xrightarrow{\text{Lactase}} \text{Glucose} + \text{Galactose}$$

End products of carbohydrate digestion

- The end products of carbohydrates are monosaccharides, such as glucose, fructose and galactose.
- A little amount of pentoses are the end products of digestion of nucleic acids and partial digestion of pentosans.

ABSORPTION OF CARBOHYDRATES

Carbohydrates are absorbed from the GIT in the form of monosaccharides.

Site of absorption

Most of the monosaccharides are absorbed from the mucosal surface of jejunum and upper ileum.

Mechanism of absorption

Various monosaccharides are absorbed by following mechanisms:

- *Glucose and galactose* are absorbed by a common Na^+-dependent active transport system;
- *Fructose* is absorbed by *facilitated diffusion* and
- *Pentoses* are absorbed by simple diffusion.

Absorption of glucose and galactose

Glucose and galactose are absorbed into the epithelial cells (enterocytes) lining the mucous membrane of the small intestine from their brush border surface (luminal surface) by an *active transport mechanism*—the *sodium co-transport* mechanism. Salient points of glucose absorption are (Fig. 7.7-1):

Binding of glucose and Na^+ to carrier protein. The carrier protein (present in the cell membrane) has two binding sites, one for sodium and another for glucose. It is called *sodium-dependent glucose transporter-1.* The conformational change in the carrier protein occurs only when the binding sites are occupied by the sodium and glucose

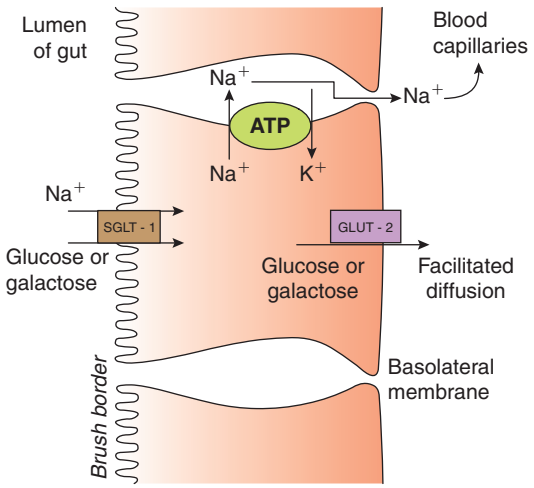


Fig. 7.7-1 Mechanism of glucose absorption across intestinal epithelial cell.

present in the gut lumen forming the sodium–glucose–carrier complex.

Creation of electrochemical gradient across the epithelial cell. The active transport of sodium by Na^+–K^+–ATPase pump through the basolateral membrane into the paracellular spaces lowers the intracellular Na^+ concentration. This creates an electrochemical gradient.

Movement of sodium and glucose inside the cell. Because of the electrochemical gradient created, the sodium moves into the cell (downhill transport). The flow of sodium ions down the gradient is so forceful that glucose (or galactose) molecule attached to the carrier protein also enters the cell even against concentration gradient for glucose (uphill movement). The energy is required for Na^+–K^+ pump activity to maintain the sodium gradient.

Transport of glucose into blood capillaries. From the epithelial cell, the glucose is transported into the interstitial space and thence to blood capillaries of portal system *through facilitated diffusion* by *glucose transporter-2.*

Rate of absorption of monosaccharides is variable, being:
- Fastest with glucose and galactose
- Intermediate with fructose
- Slowest with mannose or pentoses

FATE OF GLUCOSE IN THE BODY

1. Storage as glycogen. About 5% of the total glucose absorbed is stored as glycogen in the liver and muscles.

2. Catabolism to produce energy. About 50–60% of the glucose absorbed is catabolised in the body tissues to produce energy.

3. Conversion into fat. About 30–40% of glucose is converted into fat and is stored in the fat depot.

ABNORMALITIES OF CARBOHYDRATE DIGESTION AND ABSORPTION

Lactose intolerance

Congenital lactose intolerance refers to a condition in which lactose (milk sugar) cannot be digested due to congenital deficiency of enzyme lactase.

- The undigested lactose acts as osmotic particles and draws excessive fluids into the intestine resulting in *diarrhoea.*
- The diarrhoea so produced can lead to life-threatening dehydration and electrolyte imbalance.
- Avoidance of milk and milk products prevents the symptoms from developing if the infant can be fed by synthetic milk containing sucrose instead of lactose.

Secondary lactase deficiency, occurring in adults is very common. It produces intestinal distension, diarrhoea and flatulence. For adults, it is usually not a problem, as they can easily avoid milk and milk products.

DIGESTION AND ABSORPTION OF PROTEINS

SOURCES OF PROTEINS

The proteins that are digested and absorbed in the GIT come from two sources: exogenous and endogenous.

1. Exogenous (dietary) proteins

- *Daily requirement* of dietary proteins for adults is 0.5–0.7 g/kg body weight and for children (1–3 years), it is 4 g/kg.
- *Sources of dietary proteins* with high biological value are meat, fish, eggs, cheese and other milk products. Soyabeans, wheat and various types of pulses are also rich source of proteins.
- *Structure of dietary proteins.* The dietary proteins are made of long chains of amino acids bound together by peptide linkages.

2. Endogenous proteins

Endogenous proteins, totaling 30–50 g/day, are the proteins which reach the intestine through various gastrointestinal secretions and those which are present in the desquamated epithelial cells of the gut.

DIGESTION OF PROTEINS

Proteins are digested by the proteolytic enzymes to amino acids and small polypeptides before they are absorbed. Digestion of proteins does not occur in the mouth, as there are no proteolytic enzymes in the saliva. Digestion of proteins, thus begins in the stomach and is completed in the small intestine.

Digestion of proteins in the stomach

Pepsin, secreted by chief cells of the main gastric glands in an inactive form (pepsinogen), is responsible for digesting about 10–15% proteins entering the GIT.

- Pepsinogen is converted into pepsin (active form) by the action of HCl or preformed pepsin.
- Pepsin splits proteins into proteoses, peptones and polypeptides (Fig. 7.7-2).
- It is important to note that the optimum pH for the action of pepsin is 2.0; therefore HCl secretion by the stomach is as essential as pepsinogen secretion for the digestion of proteins.

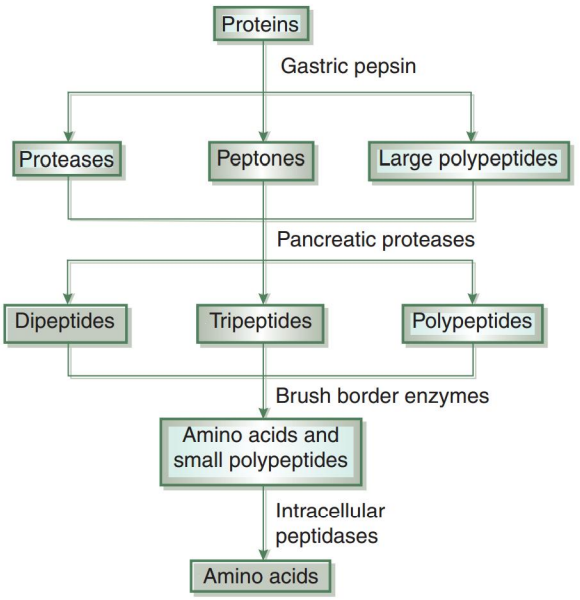


Fig. 7.7-2 Digestion of proteins.

Digestion of proteins in the small intestine

In the small intestine, the proteins are digested by the pancreatic proteases, brush border peptidases and intracellular peptidases.

Pancreatic proteases or proteolytic enzymes of pancreas play a major role in protein digestion. These can digest all the proteins, even if gastric pepsin is absent.

- *Various types of proteases* along with their functions are described on page 483.
- *Pancreatic proteases* digest the proteins and split them into dipeptides, tripeptides and small polypeptides, which are further digested by brush border peptidases (Fig. 7.7-2).
- Some of the dipeptides and tripeptides are absorbed directly into the epithelial cells of mucosa of small intestine and are further digested by the intracellular enzymes into the amino acids.

Brush border peptidases include aminopeptidases, dipeptidases, tripeptidases, nuclease and related enzymes. These enzymes continue the digestive process begun by the pancreatic proteases, eventually converting the proteins to small polypeptides and amino acids (Fig. 7.7-2).

Intracellular peptidases are the proteolytic enzymes present in the cytosol of epithelial cells of small intestine. These peptidases are specific for linkages between the various amino acids. Within minutes, these digest the last dipeptides and tripeptides into amino acids which then enter the blood.

Digestion of nucleic acid and nucleoproteins

Nucleic acid and nucleoproteins are found in abundance in the foodstuffs which are rich in nuclei, such as liver, kidney, pancreas, yeast, etc.

In the stomach, HCl hydrolyses the nucleoproteins, removing proteins which are digested together with other proteins as described above.

$$\text{Nucleoproteins} \xrightarrow{\text{HCl}} \begin{matrix}\text{Proteins}\\ + \\ \text{Free nucleic acid}\end{matrix}$$

In the small intestine, the free nucleic acids are digested by the pancreatic enzymes and brush border enzymes.

- *Pancreatic enzymes,* such as ribonuclease and deoxyribonuclease in the duodenum digest free nucleic acids into nucleotides and nucleosides.

$$\begin{matrix}\text{Freen ucleic acids}\\ \text{(RNA and DNA)}\end{matrix} \xrightarrow[\text{Deoxyribonuclease}]{\text{Ribonuclease}} \begin{matrix}\text{Nucleotides}\\ \text{and}\\ \text{nucleosides}\end{matrix}$$

- Brush border enzymes, such as nucleases, nucleotidases and nucleosidases convert nucleotides and nucleosides into pentoses (purine and pyrimidine).

$$\begin{matrix}\text{Nucleotides and}\\ \text{Nucleosides}\end{matrix} \xrightarrow[\text{Nucleosidases}]{\begin{matrix}\text{Nuclease}\\ \text{Nucleotidase}\end{matrix}} \text{Pentoses (purine and pyrimidine)}$$

End products of protein digestion

The protein digestion which starts in the stomach is completed in the enterocyte of the small intestine.

The end products of protein digestion are amino acids.

ABSORPTION OF PROTEINS

Mechanisms of absorption into the intestinal epithelial cells

The end products of protein digestion (amino acids, dipeptides and tripeptides) are absorbed through the luminal membrane of the epithelial cells of small intestine. Following mechanisms of absorption are known:

1. Na^+-dependent active transport mechanism. The levo amino acids, dipeptides and tripeptides are absorbed by a Na^+-dependent active transport mechanism.

- Separate transporters (carriers) are present for the absorption of basic, acidic and neutral amino acids. At least two different polypeptide transporters exist.
- Steps of active transport mechanism are similar to those described for glucose absorption (see page 511). These include (Fig. 7.7-3):
  - Binding of amino acid and Na^+ to carrier protein.
  - Creation of electrochemical gradient across the epithelial cells.
  - Movement of Na^+ and amino acids inside the cell.

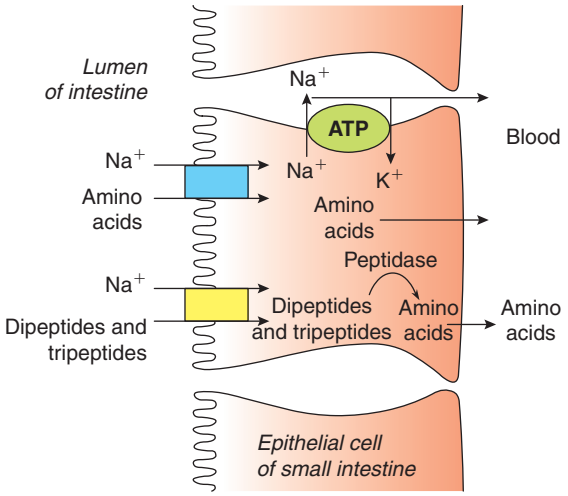


Fig. 7.7-3 Mechanism of absorption of amino acids, dipeptides and tripeptides by intestinal epithelial cells.

APPLIED ASPECTS

- Hartnup disease occurs due to the congenital defect in the transporters for neutral amino acids in intestine and renal tubules.
- Cystinuria occurs due to congenital defect in the transporters for basic amino acid.

2. Simple diffusion. The dextro amino acids are absorbed solely by the passive diffusion.

3. Endocytosis. Small amounts of larger polypeptides are absorbed by endocytosis. Proteins absorbed by endocytosis usually excite immunological/allergic reaction. In newborn infants, immunoglobulins present in the colostrum are absorbed in the intestinal mucosa by endocytosis and impart passive immunity to child.

Further digestion in the epithelial cells

Once amino acids and polypeptides are absorbed into the intestinal epithelial cells, the intracellular peptidases break the remaining linkages of tripeptides, and dipeptides causing release of amino acids.

Transport of amino acids into blood capillaries

From inside the epithelial cells, the amino acids are transported into the interstitial space across the basolateral membrane of the cells by facilitated or simple diffusion. From the interstitium, the amino acids enter the capillaries of villus by simple diffusion, and then via portal vein, they reach the liver and general circulation.

Note. It is important to note that almost all proteins ingested are absorbed. About 2–5% of proteins which escape digestion and absorption in the small intestine enter the colon and are finally digested by bacterial digestion. Therefore, the proteins that appear in the stool are not of dietary origin, but are derived from the bacterial and cellular debris.

ABNORMALITIES OF PROTEIN DIGESTION AND ABSORPTION

1. Inadequate absorption of proteins, due to lack of trypsin is a common consequence of pancreatic diseases.

2. Malabsorption of amino acids due to lack of transporters is relatively rare.

DIGESTION AND ABSORPTION OF FATS

DIETARY FATS

Types of fats. Fats are of three types:

- Simple fats or neutral fats, e.g. triglycerides and cholesterol.
- Compound fats, e.g. phospholipids.
- Associated fats, e.g. steroids and fat-soluble vitamins.

Dietary fat is of both vegetable and animal origin. Mostly, it is in the form of neutral fat (triglycerides). It also includes small amounts of phospholipids, cholesterol, some free fatty acids, lecithin and cholesterol esters.

Daily intake of fats in the diet varies widely, from about 25–160 g.

DIGESTION OF FATS

Site of digestion

Although lipolytic enzymes are secreted in the mouth *(lingual lipase)* and stomach *(gastric lipase),* their action is so insignificant that practically digestion of all the dietary fats occurs in the small intestine. Under normal conditions, gastric lipase is soon inactivated by gastric juice at pH 2.5 and acts at an optimum pH of 4.5.

Some fat digestion in stomach may occur under following exceptional circumstances:

- Achlorhydria (i.e. gastric juice cannot inactivate gastric lipase),
- Regurgitation of pancreatic lipase from the duodenum into the stomach and
- In young suckling animals which ingest large quantities of milk, the fat of milk is present in an emulsified

form and digested and inhibits the secretion of gastric juice.

Mechanism of digestion of fats

The digestion of fat includes three steps:

- Emulsification of fat by bile salts,
- Hydrolysis of fat by pancreatic and intestinal lipolytic enzymes, and
- Acceleration of fat digestion by micelle formation.

1. Emulsification of fat by bile salts

Emulsification, i.e. breaking of large fat drops into smaller droplets is a prerequisite for the action of pancreatic lipase. It is so, because the pancreatic lipase being water soluble acts only on the oil–water interface of fat. The surface area available for the action of lipase is increased many thousand times by the emulsification of fats.

Emulsification of fat is caused by the bile salts because of their property of lowering the surface tension (detergent-like action). With the lowered surface tension of the fats, the segmentation movements of small intestine break up large fat globules into fine droplets (1 µm in diameter). Lecithin (a component of bile) which has a stabilization action on the emulsions greatly enhances the emulsifying action of bile salts. The bile salts surround the fine fat droplets in such a way that their lipophilic non-polar ends are towards the fat and their hydrophilic polar ends separate the fat droplets from the aqueous phase (Fig. 7.7-4).

2. Hydrolysis of fat droplets by pancreatic and intestinal lipolytic enzymes

Pancreatic juice is markedly alkaline (pH 7.8–8.4). When it mixes with the acidic chyme (pH 6.0) coming from the stomach into the duodenum, the pH of chyme is adjusted to about 7 (which is optimal pH for the action of pancreatic lipases).

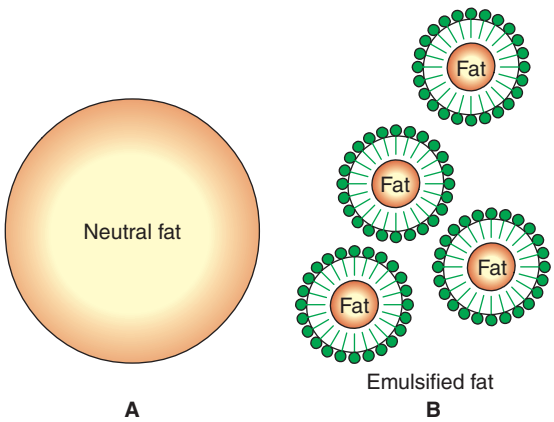
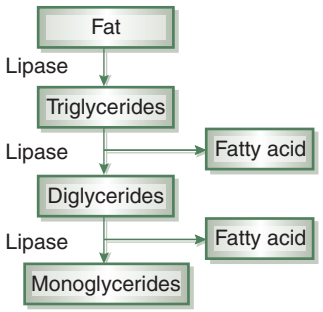


Fig. 7.7-4 Emulsification of fats by bile salts: A, a large fat particle and B, small fat particles surrounded by the bile salts.

Pancreatic lipolytic enzymes. Pancreatic juice contains three types of lipolytic enzymes. Their hydrolysing effects on fats are given:

(i) Pancreatic lipase. Pancreatic lipase is a very powerful lipolytic enzyme. The colipase, a protein present in the pancreatic juice, displaces the bile salts from the fat droplet and allows the action of lipase. The pancreatic lipase hydrolyses almost all the triglycerides (neutral fat) of the food to produce two fatty acids and a 2-monoglycerides.

(ii) Cholesterol ester hydrolase. Most of the dietary cholesterol is in the form of cholesterol esters which are hydrolysed to cholesterol and fatty acid by the cholesterol ester hydrolase.



(iii) Phospholipase A_2. It is secreted in an inactive form prophospholipase A_2 and gets converted to an active form. It hydrolyses phospholipids and separates fatty acid from them.

Intestinal lipolytic enzymes. Brush border of epithelial cells contain small amount of lipase and cholesterol esterase. Their effects though minor, but are similar to that of the pancreatic lipase.

3. Acceleration of fat digestion by micelle formation

The micelles are small water-soluble cylindrical disc-shaped particles. Each micelle is composed of a central fat globule surrounded by about 30 molecules of bile salts in such a way that their lipid-soluble non-polar ends are in the central fat globule and water-soluble polar ends fan out to form the outer covering of micelle. The monoglycerides and free fatty acids released from the digestion of fat are quickly incorporated into the central fatty portion of the micelles forming, what are known as the *mixed micelles* (Fig. 7.7-5). In this way, bile salts accelerate the fat digestion by allowing the lipolytic action to continue.

ABSORPTION OF FATS

Most of the fat absorption occurs in the duodenum; almost all the digested lipids are totally absorbed by the time the chyme reaches the mid jejunum. Absorption of fats is accomplished by following steps (Fig. 7.7-6):

1. Transportation as micelles to the brush border membrane. The bile salt micelle acts as a transport vehicle for

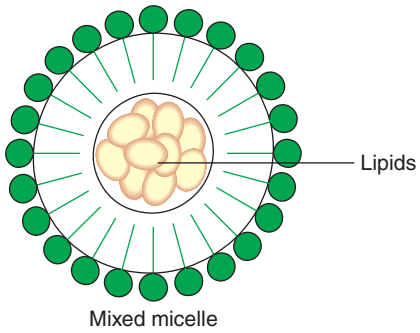


Fig. 7.7-5 Structure of a mixed micelle composed of lipids (monoglycerides, fatty acids, cholesterol) in the centre surrounded by bile salts.

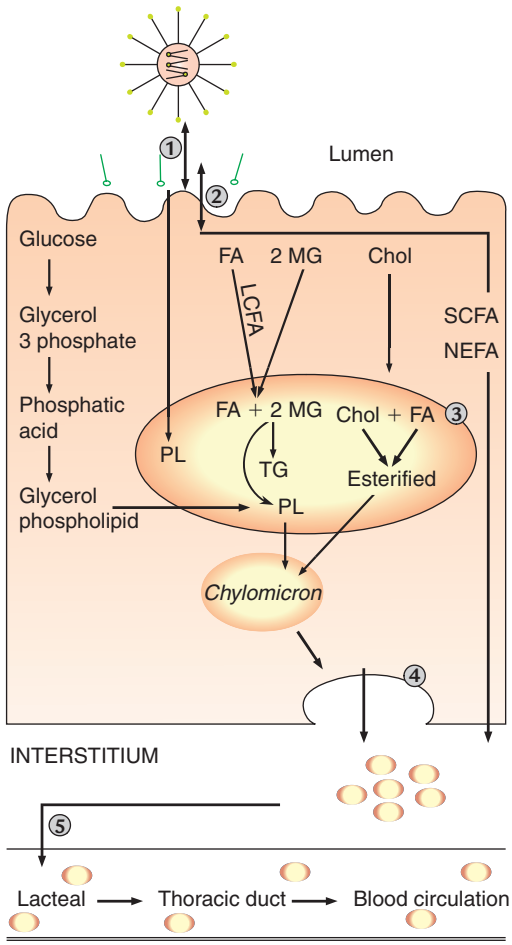


Fig. 7.7-6 Steps of fat absorption: 1, transportation of micelle to enterocytes brush border; 2, diffusion of lipids across the enterocyte membrane leaving bile salt in the lumen; 3, formation of chylomicron in the endoplasmic reticulum; 4, release of lipids into interstitium by exocytosis and 5, diffusion of lipids from interstitium into lacteal (from where lipids enter into lymphatic circulation) and through thoracic duct into circulation. (FA = fatty acid; MG = monoglycerides; chol = cholesterol; TG = triglycerides; LCFA = long-chain fatty acid; SCFA = short-chain fatty acids; NEFA = non-esterified fatty acids; PL = phospholipid.)

the products of fat digestion. As described above (Fig. 7.7-6), the outer surface of micelle is formed by water-soluble polar ends of bile salts, which helps the micelle to diffuse through the aqueous medium to reach the brush border membrane.

2. Diffusion of lipids across the enterocyte cell membrane. Once the micelle comes in contact with the cell membrane, the monoglycerides, free fatty acids, cholesterol and fat-soluble vitamins (being soluble in the cell membrane) *diffuse passively* at a rapid speed through the enterocyte cell membrane to the interior of the cell, leaving bile salts in the intestinal lumen. Thus the *rate-limiting* step in lipid absorption is the formation and migration of the micelles from the intestinal chyme to the microvilli surface.

The bile salts released from the micelle after diffusion of their associated lipids are absorbed in the terminal ileum by a Na^+-dependent active transport process.

IMPORTANT NOTE

It is important to note that the bile salts must be present in certain minimum concentration called *critical micellar concentration* before micelles are formed.

3. Transport of lipids from inside the enterocytes to the interstitial space. Once inside the cell, the end products of fat digestion enter the interstitium by two mechanisms:

(i) Diffusion across the basal border of enterocyte. The small chain fatty acids with less than 12–14 carbon atoms are able to diffuse across the basal border of enterocytes to enter the interstitium.

(ii) Formation and excretion of chylomicrons from the enterocytes by exocytosis. The large-chain fatty acids, cholesterol and lysophosphatides enter the smooth endoplasmic reticulum, where they are reconstituted:

- 2-Monoglycerides are combined with fatty acids to produce triglycerides,
- Lysophosphatides are combined with fatty acids to form phospholipids,
- Cholesterol is re-esterified.

The reformed lipids coalesce to form a small lipid droplets (about 1 nm in diameter) called chylomicrons, which are lined by β-lipoproteins synthesized. The chylomicrons are then excreted into the interstitium by *exocytosis* from the basolateral membrane of enterocyte. Covering of β-lipoproteins is essential for the exocytosis to occur.

4. Transport of lipids into circulation. After exiting the enterocytes (i.e. in the interstitium), the chylomicrons

merge into larger droplets that vary in size from 50 to 500 nm, depending on the amount of lipid being absorbed. From the interstitium, the lipids diffuse into the *lacteals,* from which they enter the lymphatic circulation and via thoracic duct gain access into the blood circulation.

Movements of villi compress the lacteals and capillaries, thus helps in mobilisation of absorbed lipids towards thoracic duct and portal vein, respectively.

APPLIED ASPECTS

Lipid malabsorption

- Lipid malabsorption is much more common than carbohydrate and protein malabsorption.
- Causes of lipid malabsorption include:
  - Deficiency of pancreatic lipase in certain pancreatic diseases
  - Bile deficiency in disorders of liver and gall bladder.
- Steatorrhoea, i.e. an increased amount of fat in the stools is common manifestation of fat malabsorption.

Serum lipid profile

Lipids are present as lipoprotein complexes. Depending upon the density, the lipoproteins are of following types:

(i) Very low-density lipoproteins. Density is, 1.060
(ii) Low-density lipoproteins
(iii) High-density lipoproteins. Density is 1.060–1.200)

Normal values

- Serum triglycerides: 30–150 mg/dL
- Serum cholesterol: 150–240 mg/dL
- Serum phospholipids: 150–300 mg/dL
- Serum-free fatty acids (FFA or NEFA = 10–30 mg/dL).

IMPORTANT NOTE

- On moderate fat intake, only 5–6% of fat is passed in the stools.
- At birth, the fat absorption process is not fully matured, therefore, in infants the faecal fat content is 10–15% of the ingested fat.

ABSORPTION OF WATER, ELECTROLYTES, MINERALS AND VITAMINS

ABSORPTION OF WATER

Water balance in the GIT

- The GIT receives about 9 L of water per day, which includes about 2 L of ingested water and about 7 L contained in salivary, gastric, biliary, pancreatic and intestinal secretions (Table 7.7-1).

Table 7.7-1 Daily water balance in GIT

Input (L)		Absorption (L)		Faecal excretion (L)
Water ingested	: 2	Jejunum (60%)	: 5.5	0.2
Water in GIT	: 7	Ileum (25%)	: 2.0	
Secretions • Saliva: *1.50* • Gastric juice: *2.50* • Bile: *0.75* • Pancreatic juice: *0.75* • Intestinal juice: *1.50*		Colon (10–15%)	: 1.3	
TOTAL	9		8.8	0.2

- The GIT absorbs about 8.8 L of water (about 95% of total water received) per day. About 60% of absorption occurs in jejunum, 20–25% in ileum and 10–15% in colon (Table 7.7-1).
- The gastrointestinal tract excretes about 0.2 L of water in the faeces per day.

Mechanism of water absorption

In general, water is absorbed passively and iso-osmotically across the gastrointestinal mucosa following the osmotic gradient created by the active absorption of electrolytes and nutrients.

- Only a small amount of water moves across the gastric mucosa, but water moves in both directions across the mucosa of small intestine and colon in response to the osmotic gradient.
- In the duodenum, the osmotic pressure created by the entering chyme causes water to flow into it.
- In the jejunum and ileum, reabsorption of sodium chloride (NaCl) creates an osmotic gradient favouring the reabsorption of water.

ABSORPTION OF SODIUM

Sodium balance in GIT

Gastrointestinal tract receives about 40 g of sodium per day, out of which about 10 g is ingested with food and about 30 g is contained in the gastrointestinal secretions. All of it is reabsorbed.

Site of absorption

Though sodium can be reabsorbed in the entire length of the intestine, but maximum absorption occurs in the jejunum.

In the small intestine, Na^+–glucose co-transport, Na^+–amino acid co-transport and Na^+–H^+ exchange mechanisms are most important (these co-transport and exchange mechanisms are similar to those in renal proximal tubule). Thus, the presence of glucose in the intestinal lumen facilitates the reabsorption for Na^+. Because of this reason, in the treatment of Na^+ and water loss in diarrhoea, glucose is added to the orally administered NaCl solution. Cereals containing carbohydrates are also useful in the treatment of diarrhoea.

In the colon, passive diffusion via Na^+ channels is most important. These channels of the colon are similar to those in the renal distal tubules and are stimulated by aldosterone (which greatly enhances sodium absorption). This mechanism is especially useful in dehydration, which leads to aldosterone secretion by the adrenal medulla.

Transport of Na^+ out of the enterocytes into interstitium occurs against its electrochemical gradient across the basolateral membrane by Na^+–K^+–ATPase active transport system.

ABSORPTION OF CHLORIDE

In the jejunum and proximal ileum, most of the Cl^- is absorbed passively through the enterocytes down the electrochemical gradient established by the active transport of Na^+. The mechanisms involved in the transport of Cl^- are:

- Passive diffusion by a paracellular route through the leaky (permeable) junction between the enterocytes and
- Neutral Na^+–Cl^- co-transport system.

In the distal ileum and large intestine, the Cl^- is absorbed by an active Cl^-–HCO_3^- exchange mechanism. In this mechanism, Cl^- is absorbed from the lumen in exchange of HCO_3^- which is secreted into the lumen. Bicarbonate (HCO_3^-) secreted into the lumen helps to neutralize the acidity produced by the action of colonic bacteria on the food.

ABSORPTION OF POTASSIUM

Passive diffusion via paracellular route down its electrochemical gradient is the mechanism involved in absorption of dietary K^+ from the small intestine.

Net movement of K^+ across the intestinal mucosa is directly proportional to the potential difference between the blood and the intestinal lumen.

ABSORPTION OF CALCIUM

Body calcium. Calcium is the most abundant among the minerals in the body. The total content of calcium in an adult man is about 1–1.5 kg of which about 99% is present in the bones and teeth. A small amount (10%) found outside the skeletal tissue performs a wide variety of functions (see page 563).

Dietary calcium. Best sources of dietary calcium are milk and milk products. Good sources of calcium are beans, leafy vegetables, fish, cabbage and egg yolk.

Dietary requirements of calcium are:

- Infants (<1 year): 300–500 mg/day,
- Children (1–18 years): 800–1200 mg/day,
- Adult men and women: 800 mg/day,
- Women during pregnancy, lactation and post-menopause: 1500 mg/day.

Site of absorption. Most of the ingested calcium is absorbed in the upper small intestine (duodenum and jejunum).

Mechanism of absorption. Normally, about 75–80% of the daily intake (about 1000 mg) of calcium is absorbed from the upper small intestine. Most of the calcium is absorbed by an active transport mechanism.

Regulation of calcium absorption

Calcium absorption from the small intestine is well regulated to maintain the plasma calcium *(homeostasis of calcium)* levels within a narrow range (9–11 mg/dL). Vitamin D and parathyroid hormone play main role in the regulation of calcium absorption.

Factors promoting calcium absorption include:

- Vitamin D (through its active form calciferol) promotes calcium absorption by inducing synthesis of calcium-binding protein.
- Parathyroid hormone enhances Ca^{2+} absorption by influencing synthesis of calciferol.
- Low pH is more favourable for Ca^{2+} absorption.
- Lactose promotes Ca^{2+} intake by the intestinal cells.
- Amino acids (lysine and arginine) facilitate Ca^{2+} absorption.

Factors inhibiting calcium absorption are:

- Phytates and oxalates inhibit Ca^{2+} absorption by forming insoluble salts with Ca^{2+} in the intestine.
- High content of dietary phosphate also prevents Ca^{2+} absorption by forming insoluble calcium phosphate. The dietary ratio of Ca^{2+} and P between 1:2 and 2:1 is ideal for optimum Ca^{2+} absorption.
- Free fatty acids inhibit Ca^{2+} absorption by forming insoluble calcium soaps. It occurs particularly when the fat absorption is impaired.
- High pH (alkaline conditions) is unfavourable for Ca^{2+} absorption.
- Dietary fibres in high content interfere with calcium absorption.

Importance of calcium-phosphorus (Ca:P) ratio. The ratio of Ca:P is important for calcification of bones. The product of Ca × P (in mg/dL) in children is about 50 and in adults it is around 40. This product is less than 30 in rickets.

ABSORPTION OF IRON

- Absorption occurs mainly in the duodenum and upper jejunum.
- Normally, about 10% of the 15–20 mg iron ingested each day is actually absorbed in a healthy adult male. This absorption is more in menstruating women.

Mechanism of iron absorption

Mechanism of iron absorption for the purpose of understanding can be described under three headings:

A. Transport of iron across the brush border of enterocyte,
B. Fate of iron in the enterocyte and
C. Transport of iron in the plasma.

A. Transport of iron across brush border of enterocyte

In the diet, iron may be present as haem (derived from meat) or non-haem iron (Fig. 7.7-7).

1. Absorption of haem iron. Haem iron is the iron present in myoglobin, haemoglobin and related compounds. From these compounds, the haem is released by the proteolytic enzymes in the gut. From the lumen, the haem is transported inside the enterocyte across the brush border membrane by a *haem transport* protein. Inside the cell, the ferrous iron (Fe^{2+}) is released from the haem by the enzyme haemoxygenase.

2. Absorption of non-haem iron. Most of the dietary non-haem iron is present in ferric form (Fe^{3+}), whereas iron can be absorbed more efficiently in ferrous form (Fe^{2+}).

- Iron has got tendency to form insoluble complexes with dietary phytates, phosphates and dietary fibres. Gastric HCl tends to break insoluble iron complex apart and thus facilitates iron absorption. This explains the occurrence of iron-deficiency anaemia in patients with deficient gastric acid secretion (achlorhydria).
- Ascorbic acid and other reducing agents promote iron absorption by reducing ferric iron to ferrous form and also by preventing iron from forming insoluble iron complexes within the chyme.
- Ferrous iron (Fe^{2+}) is transported across the brush border by the *iron transport protein* or *receptors* present on the cell membrane (Fig. 7.7-7). Once inside the enterocyte, the fate of non-haem ferrous iron is same as that of the haem iron.

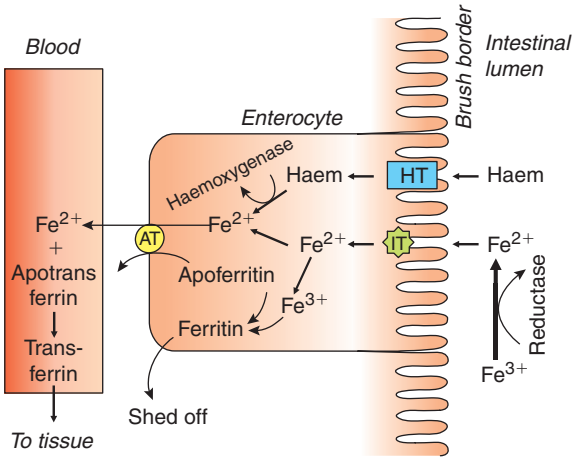


Fig. 7.7-7 Absorption of Iron. Haem is carried across brush border of enterocyte by a haem transport protein (HT), and Fe^{2+} of non-haem iron by iron transport protein (IT), inside the enterocyte some iron binds to ferritin and some crosses the basolateral membrane by active transport process (AT). In the blood, iron binds to the transport protein transferrin (TF).

B. Fate of iron in the enterocyte

As shown in Fig. 7.7-7, in the cytosol of enterocyte the free ferrous iron (Fe^{2+}) has two fates:

- A part of Fe^{2+}, depending upon the body's requirement, is actively transported across the basolateral membranes of the enterocytes into the interstitium, from where it enters the blood.
- Rest of the ferrous iron is oxidized to ferric form and bound to apoferritin forming ferritin. It is difficult to release iron from this storage form, and in general the ferritin stays in the enterocyte until the cell is sloughed off at the tip of villus.

C. Transport of iron in the blood

Normally, the iron absorbed into the blood binds with a betaglobulin (apotransferrin) to form the *transferrin* and is transported in this form in the plasma. Iron combines loosely in the globulin apotransferrin and can be released easily to enter any of the tissue cells of any point in the body.

Factors affecting absorption of iron

Factors affecting absorption of iron from the gut are:

1. Form of dietary iron. Iron may be present as haem iron or non-haem iron. Haem is absorbed directly and in *non-haem iron* the ferrous (Fe^{2+}) form is better absorbed than ferric (Fe^{3+}) form. Therefore, reducing agents, such as vitamin C enhances iron absorption by converting ferric into ferrous.

2. Meat and fish in the diet considerably enhance absorption of non-haem iron. The exact mechanism is, however, unknown.

3. Human breast milk improves the iron absorption.

4. The acid gastric juice (HCl from stomach) favours absorption of non-haem iron by causing its solubilization and reduction. Therefore, absorption of ferric iron is impaired in subjects with gastrectomy or achlorhydria.

5. Dietary factors inhibiting non-haem iron absorption are:

- *Phytates* in foods (cereals) reduce iron absorption by forming insoluble iron salts.
- *Phosphates,* calcium, egg white and bovine milk proteins inhibit iron absorption.
- *Phenols* present in legumes, tea, coffee and wine cause poor absorption of iron.

6. Iron stores in the body affect iron absorption as:

- *Decrease* in iron store of the body (e.g. in iron deficiency anaemia or when erythropoiesis is increased due to hypoxia) enhances iron absorption.
- *An increase in iron storage* in the body reduces iron absorption by the gut mucosa.

Mucosal block theory of absorption

This theory states that:

- Iron absorption is increased when body iron stores are depleted or when erythropoiesis is increased and decreased under the reverse conditions.
- As compared to the normal conditions (Fig. 7.7-8A), in iron deficiency states a larger percentage of dietary iron enters the circulation and a smaller amount forms ferritin in the enterocytes (Fig. 7.7-8B).
- In the presence of iron overload, more ferritin is formed in the enterocytes and shed with these cells in the stools (Fig. 7.7-8C).

APPLIED ASPECTS

Iron deficiency results in iron-deficiency anaemia. For details see page 118.

Iron excess. Iron overload may occur when its absorption exceeds its excretion. Such a condition occurs due to:

- *Excessive intake* of iron and
- *Idiopathic* or congenital failure of mucosal feedback mechanism controlling iron absorption.
- *Excessive* destruction of erythrocytes may also be associated with siderosis.

The excess of iron in the body results in accumulation of haemosiderin in the tissues producing the so-called *haemosiderosis.* The excess of haemosiderin damages the tissues and produces the condition of **haemochromatosis,** which is characterized by:

- Pigmentation of the skin,
- Diabetes due to pancreatic damage (bronze diabetes),
- Cirrhosis of liver,
- Carcinoma of the liver and
- Atrophy of gonads.

ABSORPTION OF VITAMINS

Absorption of fat-soluble vitamins

- Fat-soluble vitamins (A, D, E and K) become part of micelle formed by the bile salts and are absorbed along with other lipids in the upper part of the small intestine.
- The absorption of fat-soluble vitamins is deficient if fat absorption is depressed because of lack of pancreatic enzymes or if bile is excluded from the intestine by obstruction of bile duct.

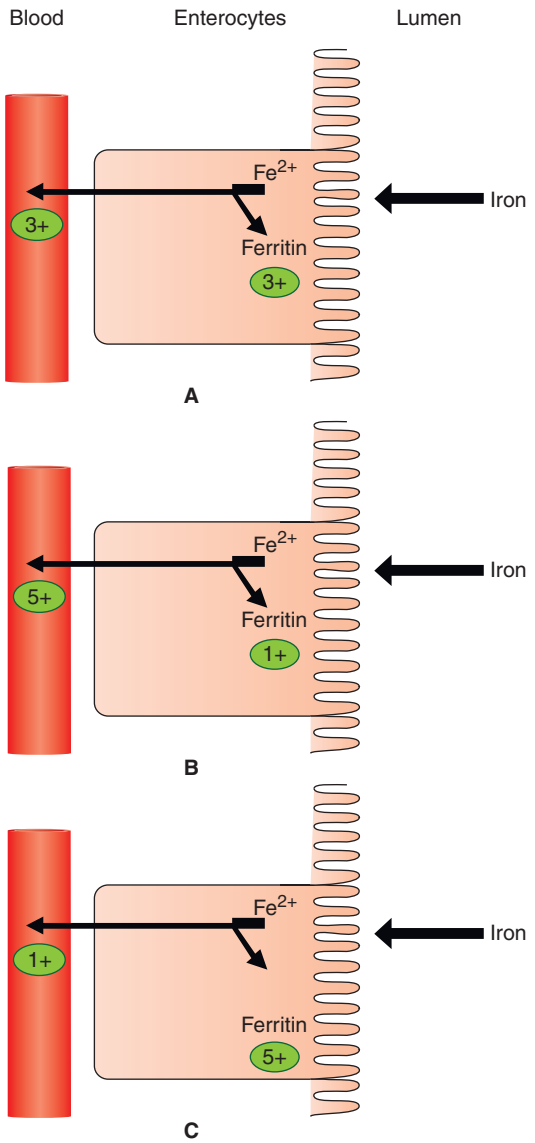


Fig. 7.7-8 Mucosal block theory of regulation of iron absorption: A, normally equal amount (3^+) of iron is bound to form ferritin and enters the blood across basolateral membrane; B, in iron deficiency states less iron forms ferritin (1^+) and more enters the blood (5^+); and C, in iron overload more ferritin is formed (5^+) and less enters the blood (1^+).

Absorption of water-soluble vitamins

- Absorption of water-soluble vitamins is rapid as compared to fat-soluble vitamins.
- Most vitamins are absorbed in the upper part of the small intestine (jejunum) except the vitamins B_{12}, which is absorbed in the ileum.
- Most water-soluble vitamins, e.g. vitamin C and the vitamins B (biotin, folic acid, nicotinic acid B_8, i.e. pyridoxine, B_2, i.e. riboflavin, and B_1, i.e. thiamine) are absorbed by facilitated transport or by a Na^+-dependent active transport system in the proximal small intestine.
- Vitamin B_{12} absorption is most complex than that of other water-soluble vitamins and needs separate description.

Absorption of vitamin B_{12} involves following steps (Fig. 7.7-9):

- *In the stomach,* vitamin B_{12} is exposed to specific binding protein R and vitamin B_{12} binding protein called intrinsic factor (IF). As the affinity of R protein for vitamin B_{12} is much more than that of IF, so most of the vitamin B_{12} gets bound to R protein in the stomach.
- *In the lumen of intestine,* the pancreatic proteases cleave vitamin B_{12} from the R protein. Then, the vitamin B_{12} binds to IF to form a complex (IF–B_{12}).

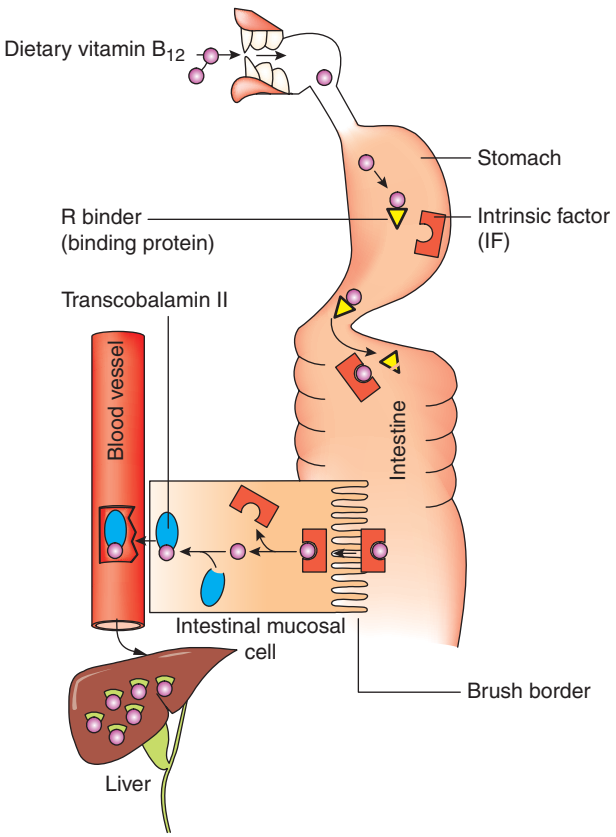


Fig. 7.7-9 Schematic diagram showing absorption and transport of vitamin B_{12}.

- *On the brush border of enterocyte,* IF–B_{12} complex become bound to the specific receptors. Following this, the vitamin B_{12} is transported into the cytosol of enterocyte by endocytosis, leaving behind IF at the brush border. It is important to note that absorption of vitamin B_{12} from the IF–B_{12} complex can occur only after the complex binds to the receptors. In the absence of intrinsic factor, vitamin B_{12} absorption is markedly decreased and patient may develop pernicious anaemia (see page 119).
- *From the basolateral border of enterocyte,* vitamin B_{12} enters the portal circulation after binding with plasma globulin called transcobalamine-II.
- *In the liver,* vitamin B_{12} is stored in large amounts after binding with another globulin called *transcobalamin-1.* The storage of water-soluble vitamin is unique to vitamin B_{12}. Liver may store up to 3 years of supply. Vitamin B_{12} is also stored in muscles for some extent. Whenever required, it is transported from the liver to the bone marrow.

APPLIED ASPECTS

MALABSORPTION SYNDROME

Malabsorption syndrome is not a simple disease but a group of disorders in which multiple nutritional deficiency states are produced.

General features of malabsorption are:

- Deficient absorption of amino acids, fats and carbohydrates results in general weakness.
- Malabsorption of vitamins may produce anaemia and signs of hypovitaminosis.
- Malabsorption of iron results in iron-deficiency anaemia.
- Malabsorption of fats produces steatorrhoea (see page 495).
- Water and electrolyte depletion may result in dehydration.

Common conditions which can produce malabsorption are:

- Coeliac disease
- Sprue
- Lactose intolerance (see page 512)
- Crohn's disease
- Resection of small intestine
- Malabsorption after gastric surgery (see page 479)
- Blind loop syndrome
- Chronic pancreatitis (see page 485)
- Obstruction of common bile duct (see page 494).

Coeliac disease

Aetiopathogenesis. It occurs due to the deficiency of the enzyme gluten hydrolase. As a result, gluten, the principal

protein of wheat, rye, barley and oats is not properly hydrolysed. Consequently, gliadine, a toxic polypeptide, is formed, which produces an inflammatory response in the intestinal mucosa leading to the destruction of microvilli.

Clinical features of gluten-induced enteropathy are those of generalized malabsorption. It may occur as:

- Congenital disease manifesting usually within first 3 years of life and
- Acquired disease in adults due to unknown aetiology.

Treatment consists of withdrawal of wheat and other sources of gluten in the diet.

Sprue

Sprue or the tropical sprue is a disorder of malabsorption, which is particularly characterized by features of failure of absorption of folate with or without associated malabsorption of vitamin B_{12}. So, there occurs general features of malabsorption with megaloblastic anaemia which is conspicuous.

Crohn's disease

Aetiopathogenesis. It is an inflammatory bowel disease characterized by idiopathic non-specific granulomatous inflammation of the bowel.

Clinical features are varied depending upon the part and extent of bowel involved. Common clinical features are off and on fever, chronic diarrhoea, abdominal discomfort and pain, and weight loss is frequent and many patients have moderate anaemia and other features of malabsorption. Ultimately, patient may develop narrowing and obstruction of intestinal lumen, fistula formation or intestinal perforation.

Resection of small intestine

Removal of short segment from the jejunum or ileum generally does not produce any severe symptoms. Because there occurs compensatory hypertrophy and hyperplasia of remaining mucosa (*intestinal adaptation*), the capacity of the jejunum to adapt is less than that of the ileum.

Removal of ileum produces greater degree of malabsorption as compared to the removal of jejunum. Because ileal resection prevents absorption of bile salts causing decreased fat absorption, the entry of unabsorbed bile salts in the colon inhibits Na^+ and water reabsorption producing diarrhoea.

Removal of large segment of ileum leaving behind duodenum, jejunum and a very small length of ileum produce malabsorption which is characterized by:

- Normal carbohydrate absorption (99% of ingested carbohydrates are absorbed).
- Adequate protein absorption (70% of ingested proteins are absorbed).
- Markedly decreased fat absorption which may produce:
  - Steatorrhoea, i.e. increase in faecal fat (see page 495).
  - Deficiency symptoms of fat-soluble vitamins (A, D, E and K).
  - Fatty infiltration of liver and cirrhosis.
- Markedly decreased calcium absorption due to the formation of insoluble calcium salts. Decreased serum calcium (hypocalcaemia) may produce tetany.

Gastrocolic fistula

In this condition, chyme enters directly into the transverse colon from the stomach. It is characterized by following additional features over and above the features of large segment resection of ileum described above:

- Pernicious anaemia due to failure of absorption of vitamin B_{12}.
- Hypovitaminosis due to both water-soluble as well as fat-soluble vitamins.
- Amino acid malabsorption which produces hypoproteinaemia (causing generalized oedema) and marked muscular weakness with wasting.

Blind loop syndrome

Blind loop syndrome is characterized by the formation of the areas of the intestine where bacteria can proliferate without being subjected to movement down the intestine.

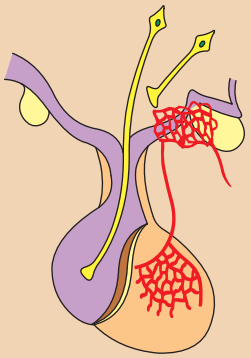
Causes of blind loop formation are multiple diverticula in the small intestine, afferent loop after partial gastrectomy, areas of disordered peristalsis in small intestine and fistula from the upper small intestine to the colon.

Features. Colonisation of small bowel by bacteria in blind loop syndrome may produce:

- Malabsorption of fat (steatorrhoea). It occurs due to deconjugation of bile salts by bacteria.
- Megaloblastic anaemia due to vitamin B_{12} deficiency (which is taken up by bacteria).
- Amino acid deficiency (due to consumption by the bacteria) resulting in weakness and hypoproteinaemia.
- Diarrhoea and other nutritional deficiency.

SECTION 8

Endocrinal System



The biological functions of the multicellular living organisms are very well co-ordinated. This co-ordination is achieved by two main control systems, the nervous system and the endocrinal system.

Nervous system is principally related with functions of the body in external and internal environment. The nervous system co-ordinates the body functions through transmission of impulses via nerve fibres.

Endocrinal system is mainly concerned with different metabolic functions of the body, especially the chemical reactions and transport of various substances. The endocrinal functions are accomplished through a wide range of chemical messengers, the hormones.

Relationship between endocrine and neural physiology

In a conceptual sense, the nervous system and the endocrine system have important functional similarities. Each is basically a system for signaling. In fact, the nervous system and the endocrine system often respond together to incoming stimuli so as to integrate the organism's response to changes in its external and internal environment. The co-ordinated function of these systems is well illustrated by following example.

A significant decrease in the circulating blood volume is sensed by baroreceptors, the cardiac atria, the kidney and the brain. The sympathetic nervous system, a neurohormone from the posterior pituitary gland and hormones from the cardiac atria and ventricles, the adrenal medulla, the adrenal cortex and the kidneys act on target cells in the blood vessels and kidneys to restore blood volume.

Organization of endocrine system

The endocrinal system consists of various endocrine glands and neurosecretory cells located in the hypothalamus. The neurosecretory cells of hypothalamus secrete certain neurohormones called releasing and inhibitory factors, which influence the secretion of hormones from other endocrine glands. Certain other substances act as neurotransmitters in the brain and influence the secretion of neurosecretory cells of the hypothalamus. The environmental factors through these neurotransmitters influence the whole endocrine system. The various endocrine glands present in the body are:

1. Pituitary gland (hypophysis). Pituitary gland is also known as hypophysis, which in Greek means undergrowth of the brain. It has two main parts: adenohypophysis and neurohypophysis. *Adenohypophysis* secretes growth hormone (GH) or somatotropins, follicle-stimulating hormone (FSH), luteinizing hormone (LH), prolactin, thyrotropin or thyroid-stimulating hormone (TSH) and corticotropin or adrenocorticotropic hormone (ACTH). The *neurohypophysis* stores the antidiuretic hormone (ADH) or vasopressin and oxytocin synthesized by the hypothalamus.

2. Thyroid gland. The thyroid gland is present in the neck in front of the trachea. It secretes thyroxine (T_4) and triiodothyronine (T_3). The C cells or parafollicular cells secrete calcitonin.

3. Parathyroid glands. These are four in number, very small glands situated behind the lobes of the thyroid gland and secrete parathormone.

4. Adrenal glands. These are situated on the upper poles of the two kidneys, hence also called suprarenal glands. The outer cortex region of the adrenal glands secretes cortisol, aldosterone and sex steroids and the inner medullary region secretes catecholamines (adrenaline and noradrenaline).

5. Pancreatic islets (islets of Langerhans). These are small groups of cells, which secrete insulin, glucagon and somatostatin.

6. Gonads. These include ovaries in females and testes in males. The ovaries secrete oestrogens and progesterone (female sex steroids) and testes secrete testosterone (male sex hormone).

7. Pineal gland. It is a small gland present in the roof of third ventricle in the brain. It secretes melatonin and other biogenic amines.

8. Placenta. During pregnancy, placenta secretes various hormones like human chorionic gonadotropin (HCG), oestrogen, progesterone, somatotropins and relaxin.

9. Gastrointestinal mucosa also secretes various hormones collectively known as gastrointestinal (GI) hormones, e.g. gastrin, secretin, cholecystokinin–pancreozymin (CCK-PZ), etc.

10. Kidneys. In addition to their renal functions, the kidneys secrete erythropoietin, prostaglandins and 1,25-dihydroxycholecalciferol, and also help in the activation of angiotensin production.

11. Atrial muscle cells. These secrete atrial natriuretic peptides (ANP) and many other peptides.

12. Skin. This is also considered to act as an endocrine structure by producing vitamin D, which is now considered to be a hormone.

Chapter 8.1

General Principles of Endocrinal System

HORMONES: DEFINITION AND CLASSIFICATION

DEFINITION

The word hormone is derived from the Greek word *hormaein,* which means to execute or to arouse. In the classic definition, hormones are secretory products of the ductless glands, which are released in catalytic amounts into the blood stream and transported to specific target cells (or organs), where they elicit physiologic, morphologic and biochemical responses. The chemical messengers which perform hormonal functions are defined as (Fig. 8.1-1):

Endocrine hormones. These are the chemical messengers whose function is the transmission of a molecular signal from a classic endocrinal cell through the blood stream to a distant target cell (Fig. 8.1-1A).

Neurocrine hormones. Neurohormones or peptides are released from a neurosecretory neuron into the blood stream and then carried to a distant target cells (Fig. 8.1-1B). Example of such neurocrine substances are oxytocin and antidiuretic hormone.

Paracrine hormones. These are chemical messengers which after getting secreted by a cell are carried over short distance by diffusion through the interstitial spaces (extracellular fluid) to act on the neighbouring different cell types (Fig. 8.1-1C). For example, in islets of Langerhans, somatostatin secreted by the delta cells acts on the alpha and beta cells.

Autocrine hormones. These refer to those chemical messengers which regulate the activity of neighbouring similar type of cells (Fig. 8.1-1D). Examples of autocrine hormones are prostaglandins.

CLASSIFICATION OF HORMONES

A. Depending upon the chemical nature

1. *Amines or amino acid derivatives;* e.g.
   - Catecholamines (epinephrine and norepinephrine) and
   - Thyroxine (T_4) and Triiodothyronine (T_3).
2. *Proteins and polypeptides*
   - Posterior pituitary hormones (antidiuretic hormone and oxytocin),
   - Insulin,
   - Glucagon,
   - Parathormone and
   - Other anterior pituitary hormones.
3. *Steroid hormones.* These include:
   - Glucocorticoids,
   - Mineralocorticoids,
   - Sex steroids and
   - Vitamin D.

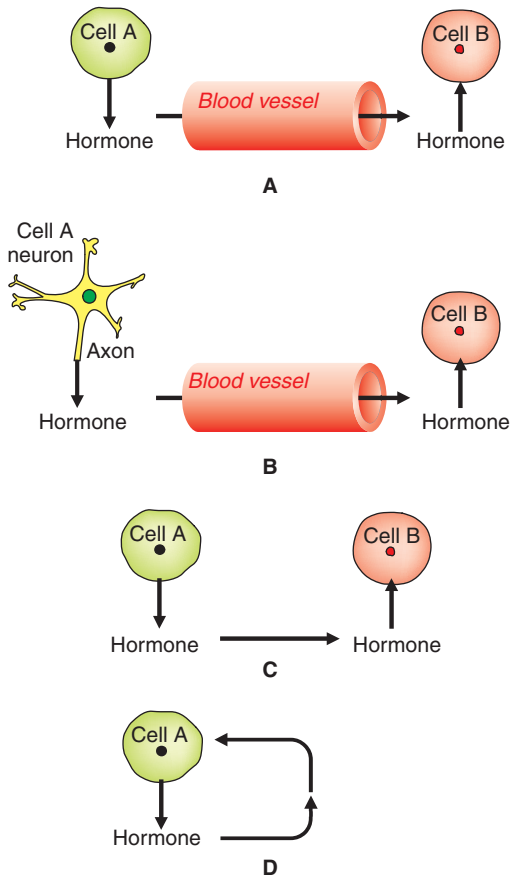


Fig. 8.1-1 Different types of hormones (by their mechanism of action) are: A, endocrine hormone; B, neurocrine hormone; C, paracrine hormone and D, autocrine hormone.

B. Depending upon the mechanism of action

1. Group I hormones. These act by binding to intracellular receptor and mediate their actions via formation of a hormone–receptor complex. These include steroid, retinoid and thyroid hormones.

2. Group II hormones. These involve second messenger to mediate their effect. Depending upon the chemical nature of the second messengers, group II hormones are further divided into four subgroups: A, B, C and D (Table 8.1-1).

HORMONES: GENERAL CONSIDERATIONS

HORMONE TRANSPORT, PLASMA CONCENTRATION AND HALF-LIFE

Hormone transport

After secretion into the blood stream, the hormones may circulate in two forms:

Unbound form. Some hormones circulate as free molecule, e.g. catecholamines and most peptide and protein hormones circulate unbound.

Bound form. Some hormones, such as steroids, thyroid hormones and vitamin D circulate bound to specific globulins

Table 8.1-1 Types of group II hormones based on the chemical nature of second messenger involved in their mechanism of action

Group	Second messenger	Hormones
Group II-A	Cyclic AMP (cAMP)	Adrenocorticotropic hormone (ACTH) Antidiuretic hormone (ADH) Angiotensin II Calcitonin Corticotropic hormones (CRH) Catecholamine (α2 adrenergic) Follicular-stimulating hormone (FSH) Glucagon Luteinizing hormone (LH) Parathormone (PTH) Somatostatin Thyroid-stimulating hormone (TSH)
Group II-B	Cyclic GMP (cGMP)	Atrial natriuretic factor (ANF) and nitric oxide
Group II-C	Calcium/or phosphatidyl inositol/ or both	Acetylcholine (ACh), catecholamines (α_1 adrenergic), gastrin, oxytocin, thyrotropin-releasing hormone (TRH), gonadotropin-releasing hormone (GnRH) and platelet-derived growth factor (PDGF).
Group II-D	Kinase or phosphatase cascade	Human chorionic somatotropin (HCS), erythropoietin, growth hormone, insulin and insulin like growth factors (IGF-I and IGF-II), nerve growth factor (NGF), prolactin and other growth factors.

that are synthesized in the liver. The binding of hormones to proteins is advantageous as it:

- Protects the hormone against clearance by the kidney,
- Slows down the rate of degradation by the liver and
- Provides circulating reserve of the hormone.

Some hormones are carried in the blood as inactive forms with proteins. They become active at the target site only. Only unbound hormones pass through capillaries to produce their effects or to degrade.

Plasma concentration

Hormones are usually secreted into the circulation in extremely low concentrations:

- *Peptide hormone concentration is between* 10^{-12} and 10^{-10} mol/L.
- *Epinephrine and norepinephrine concentrations are* 2×10^{-10} and 13×10^{-10} mol/L, respectively.
- *Steroid and thyroid hormone concentrations are* 10^{-9} and 10^{-6} mol/L, respectively.

Half-life

Most hormones are metabolized rapidly after secretion. In general:

- Peptide hormones have a short half-life and
- Steroids and thyroid hormones have significantly longer half-life because they are bound to the plasma proteins. Table 8.1-2 depicts half-life of some of the hormones.

Table 8.1-2 Half-life of some of the important hormones

Class of hormone	Hormone	Half-life
Protein and peptide hormones	ADH	<1 min
	Oxytocin	<1 min
	Insulin	5 min
	Prolactin	12 s
	Growth hormone	<30 min
	ACTH	15–25 s
	LH	15–45 min
	FSH	180 min
Amines	Epinephrine	10 s
	Norepinephrine	15 s
	Thyroxine (T_4)	5–7 days
	Triiodothyronine (T_3)	1–3 days
Steroid hormones	Aldosterone	30 min
	Cortisol	90–100 min
	1,25-Dihydroxycholecalciferol	15 h
	25 Hydroxycholecalciferol	15 days

FUNCTIONS OF HORMONES

Hormones regulate existing fundamental processes but do not initiate reactions de novo.

1. Regulation of biochemical reactions. Hormones regulate the metabolic functions in a variety of ways:

- They stimulate or inhibit the rate and magnitude of biochemical reactions by controlling enzymes and thereby cause morphologic, biochemical and functional changes in target tissues.
- They modulate energy producing processes and regulate the circulating levels of energy-yielding substances (e.g. glucose, fatty acids). However, they are not used as energy sources in biochemical reactions.

2. Regulation of bodily processes. Hormones regulate different bodily processes, such as growth, maturation, differentiation, regeneration, reproduction and behaviour. Thus, main function of the endocrine glands is to maintain homeostasis in an internal environment. For these functions the hormones do not act directly on the intracellular machinery.

HORMONE DISPOSAL

Mechanisms of hormone disposal

The circulating hormones are disposed off by following mechanisms:

- Target cell uptake and intracellular degradation,
- Metabolic degradation/inactivation and
- Urinary or biliary secretion.

1. Target cell uptake and intracellular degradation. The interaction of hormones with their target cells is followed by intracellular degradation.

- *Degradation of protein and amine hormones* occurs after binding to membrane receptors and then internalization of hormone–receptor complex.
- *Degradation of thyroid and steroid hormones* occurs after binding the hormone–receptor complex to the chromatin.

2. Metabolic degradation/inactivation. Only a small fraction of the circulating hormone is removed by the target tissue cells, most of the hormone extraction and degradation occurs in the liver and kidneys. Metabolic degradation occurs by enzymatic processes that include proteolysis, oxidation, reduction, hydroxylation, decarboxylation and methylation. Virtually, all the hormones are extracted from the plasma and degraded to some extent by the liver. In addition, glucuronization and sulfation of hormones or their metabolites may be carried out and the conjugates are subsequently excreted in the bile or the urine.

REGULATION OF HORMONE SECRETION

The quantity of hormones secreted is regulated in accordance with their requirement. General mechanisms that govern the secretion of hormone include:

- Feedback control,
- Neural control and
- Chronotropic control.

1. Feedback control

Feedback control is of two types:

- Negative feedback control and
- Positive feedback control.

Negative feedback control. Generally, the influence of blood concentration of the hormone concerned or its effect is to inhibit further secretion of the hormone and is called negative feedback control (Fig. 8.1-2A).

Positive feedback control. It is less common, acts to amplify the initial biological effects of the hormone (Fig. 8.1-2B).

Depending upon the product involved the feedback mechanism may be:

- Hormone–hormone feedback and
- Substrate–hormone feedback.

(i) Hormone–hormone feedback control

The best example of hormone–hormone negative feedback control is the regulation of hormone secretions by the hypothalamus and pituitary, which involves three loops (Fig. 8.1-3):

- ***Long-loop feedback*** (Fig. 8.1-3A). The peripheral gland hormone (e.g. thyroid, adrenocortical, and gonads) can exert long-loop negative feedback control on both the hypothalamus and the anterior lobe of pituitary.
- ***Short-loop feedback*** (Fig. 8.1-3B). The pituitary tropic hormones decrease the secretion of hypophysiotropic hormone (e.g. GHRH, GHIH, TRH, GnRH, etc.) by short-loop feedback.

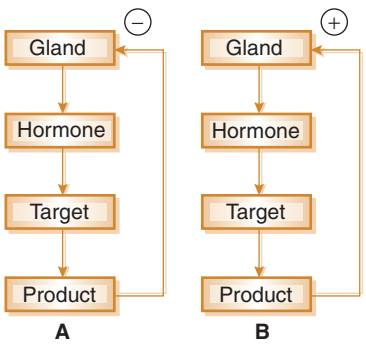


Fig. 8.1-2 Hormonal regulation by feedback control mechanism: A, negative feedback and B, positive feedback.

- ***Ultra-short-loop feedback*** (Fig. 8.1-3C). The hypophysiotropic hormones may inhibit their own synthesis and secretion via an ultra-short-loop feedback mechanism.

(ii) Substrate–hormone feedback control

The best example of substrate–hormone feedback control is regulation of insulin secretion from the pancreatic beta cells of the islets of Langerhans and glucagon secretion from the alpha cells by blood glucose levels. A rise in blood glucose level promotes the secretion of insulin while a fall in blood glucose promotes secretion of glucagon. These responses keep the blood glucose level within narrow limits in spite of variation in carbohydrate intake in diet.

2. Neural control

Neural control acts to evoke or suppress hormone secretion in response to both external and internal stimuli.

External stimuli, which can modulate hormone release through neural mechanisms, may be visual, auditory, olfactory, gustatory and tactile.

Internal stimuli, which influence hormonal release through neural mechanism, include pain, emotion, sexual excitement, fright, stress and changes in blood volume.

Neural control depending upon the type of nerve fibres involved may be:

- Adrenergic,
- Cholinergic,
- Dopaminergic,

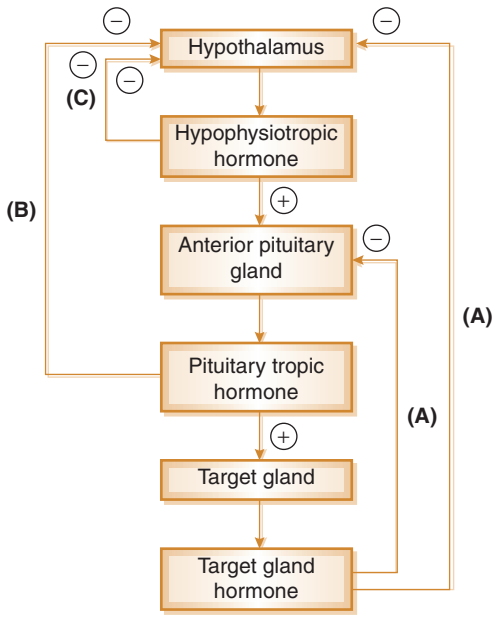


Fig. 8.1-3 Hormone–hormone negative feedback control by the hypothalamus and pituitary: A, long-loop feedback; B, short-loop feedback and C, ultra-short-loop feedback.

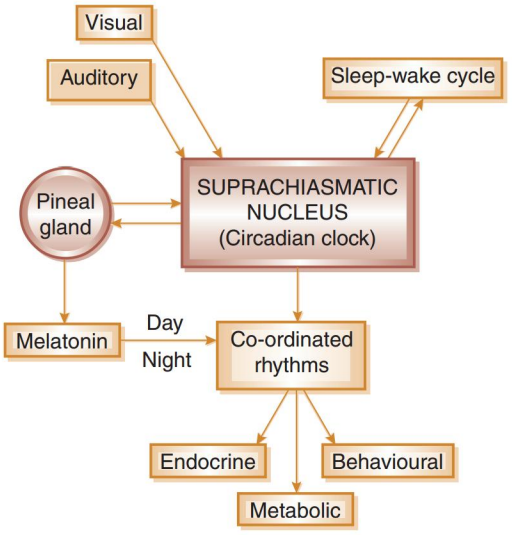


Fig. 8.1-4 The origin of circadian rhythms in endocrine gland secretion, metabolic process and behavioural activity.

- Serotoninergic and
- GABAergic.

Examples of neural control of hormones are:

- *Release of oxytocin,* which fills the milk ducts in response to the stimulus of suckling,
- *Release of aldosterone,* which augments the circulatory volume in response to upright posture and
- *Release of melatonin* in response to darkness.

3. Chronotropic control

Chronotropic control of hormone secretion accounts for:

- Oscillating and pulsatile release of certain hormones,
- Diurnal variation in hormonal levels,
- Menstrual rhythm,
- Seasonal rhythm and
- Developmental rhythm.

The source of regular oscillatory cycles is a pulse generator(s) located in the suprachiasmatic nucleus (SCN) of the hypothalamus (Fig. 8.1-4).

The intrinsic circadian clock is also located in the SCN, which is responsible for endocrinal, metabolic and behavioural co-ordinated rhythms.

HORMONE: RECEPTORS AND MECHANISM OF ACTION

HORMONE RECEPTORS

All hormones act through specific receptors. Almost all hormone receptors are large proteins present in hormone-sensitive target cells.

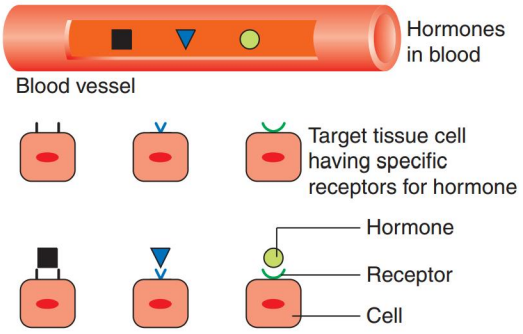


Fig. 8.1-5 Specificity of hormone action is because of specific receptors.

CHARACTERISTICS OF HORMONE RECEPTORS

Receptor specificity. There are specific receptors for each hormone. This is the reason that all hormones circulate to all parts of the body, yet each hormone has a specific target tissue for its action (Fig. 8.1-5).

Receptor location. Depending upon the location, receptors are of two types:

(i) Internal receptors are located inside the cells, for example, receptors for steroid hormones and thyroid hormones are localized within the nucleus of the target cells.

(ii) External receptors are located on the plasma membrane of the target cells, for example, peptide and protein hormones, amines and prostaglandins are interspersed within the phospholipid bilayer of the plasma membranes.

Change in receptor number. Number of receptors of a cell varies depending upon the situation. It is regulated by two mechanisms: down regulation and up regulation.

(i) Down regulation refers to a decrease in the number of active receptors. It occurs to regulate the hormone sensitivity when it is present in excess. For example, elevated ambient insulin concentration causes a loss or inactivation of insulin receptors in liver cells, fat cells and white blood cells.

(ii) Up regulation refers to an increase in the number of active receptors on a cell. It occurs to regulate the hormone action when its concentration is less. This phenomenon tends to reduce the effect of hormone deficiency.

MECHANISM OF ACTION OF HORMONES

The main mechanisms of hormone actions are:

- Action through change in the membrane permeability,
- Action through effect on gene expression by binding of hormones with intracellular receptors,
- Action through secondary messengers which activate intracellular enzymes when hormones combine with membrane receptors and
- Action through tyrosin kinase activation.

ACTION THROUGH CHANGE IN MEMBRANE PERMEABILITY

Certain hormones bind with the receptors present in the cell membrane (external receptors) and cause conformational change in the protein of the receptors, this results into either opening or closing of the ions channels (such as Na^+ channels, K^+ channels, and Ca^{2+} channels). The movement of ions through Ca^{2+} channels causes the subsequent effect, e.g. adrenaline, noradrenaline act by this mechanism.

ACTION THROUGH EFFECT ON GENE EXPRESSION BY BINDING OF HORMONES WITH INTRACELLULAR RECEPTORS

Group I hormones act by their effect on the gene expression include steroid hormones, retinoids and thyroid hormones. These hormones are lipophilic in nature and can easily pass across the cell membrane. They act through intracellular receptors located either in the cytosol or in the nucleus. The sequence of events involved is (Fig. 8.1-6):

1. *Transport.* After secretion, the hormone is carried to the target tissue on serum binding protein.
2. *Internalization.* Being lipophilic, the hormone easily diffuses through the plasma membrane.

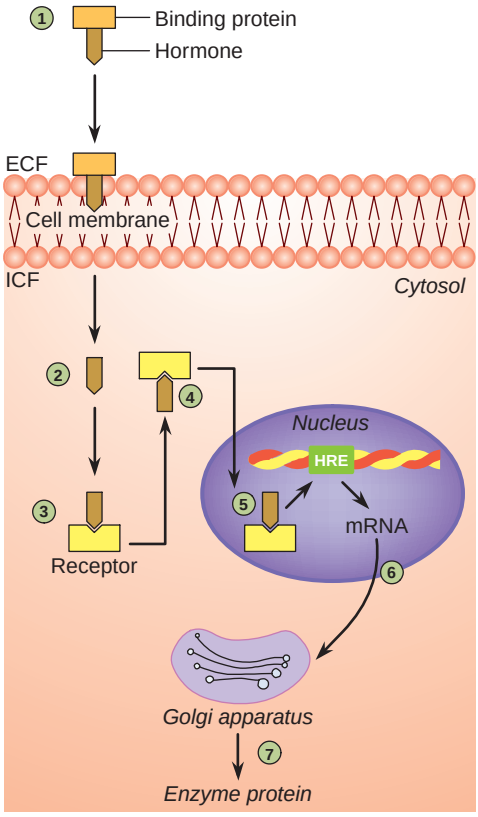


Fig. 8.1-6 Action of hormones through their effect of gene expression. Note 1–7 represent the steps involved in the process (for details see text).

3. *Receptor–hormone complex* is formed by binding of hormone to the specific receptor inside the cell.
4. *Conformational change* occurs in the receptor proteins leading to activation of receptors.
5. The activated receptor–hormone complex then diffuses into the nucleus and binds on the specific region on the DNA known as hormone responsive element (HRE), which initiates gene transcription.
6. *Binding of the receptor–hormone complex* to DNA alters the rate of transcription of messenger RNA (mRNA).
7. *The mRNA diffuses in the cytoplasm*, where it promotes the translation process at the ribosomes. In this way, new proteins are formed which result in specific responses. Some of the new proteins synthesized are enzymes.

IMPORTANT NOTE

Heat shock proteins (HSP) are intracellular proteins and on exposure to heat or other stresses their concentration increases. These proteins help the cells to survive a variety of stresses, therefore these are also called the stress proteins.

Mechanism of action

The heat shock proteins bind to the receptors of the hormone (mainly of glucocorticoid, progesterone, oestrogen) and cover the DNA binding domain. When the hormone binds to the receptor, a conformational change occurs that causes release of HSP and the DNA binding domain is exposed.

Note. The hormonal action mediated through intracellular receptors is comparatively slower. Therefore glucocorticoids may take hours to few days to achieve the therapeutic effect.

ACTION THROUGH SECOND MESSENGERS

The peptides and biogenic amines are two principal classes of hormones which act through second messenger and are classified as group II hormones (Table 8.1-1). Such hormones are also called first messengers. The release of second messenger is mediated by GTP binding proteins also called G-proteins.

Coupling by G-proteins

Events involved in coupling by G-protein which lead onto changes in the cellular concentration of the second messengers are summarized (Fig. 8.1-7):

- Group II hormones are water soluble and bind to the plasma membrane of the target cell via cell surface receptors.
- The hormone bearing receptor then interacts with a G-protein and activates it by binding GTP. There are two classes of G-proteins: stimulatory G-protein (Gs) and inhibitory G-protein (Gi).

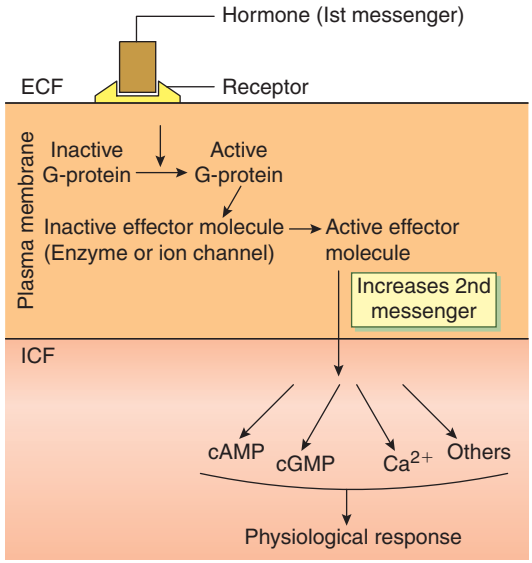


Fig. 8.1-7 Schematic mechanism of coupling by G-protein leading to increase in second messenger which mediates hormone's physiological response.

- In its activated ("on") state, the G-protein interacts with one or more of the effector protein (most of which are enzymes or ion channels such as adenylyl cyclase; Ca^{2+} or K^+ channels or phospholipase C, A_2 or D) to activate or inhibit them.
- The changed effector molecules, in turn, generate second messenger that mediates the hormone's intracellular action.

Second messenger systems

The second messenger systems that are activated through coupling of hormone–receptor complexes by G-protein include:

- Adenylyl cyclase–cAMP system,
- Guanyl cyclase–cGMP system,
- Membrane phospholipase–phospholipid system and
- Calcium–calmodulin system.

1. Adenylyl cyclase–cAMP system

The adenylyl cyclase–cAMP system was the first to be described by Sutherland in 1961 that initiated the concept of second messenger. The hormones which act through this system constitute the group IIA hormones (Table 8.1-1). The steps involved in the hormone action via adenylyl cyclase–cAMP system are summarized below (Fig. 8.1-8):

(i) Binding of hormone (Step 1) to a specific receptor in the cell membrane.

(ii) Activation of G-protein (Step 2). After formation of hormone–receptor complex, the GDP is released from the G-protein and is replaced by GTP, i.e. G-protein is activated.

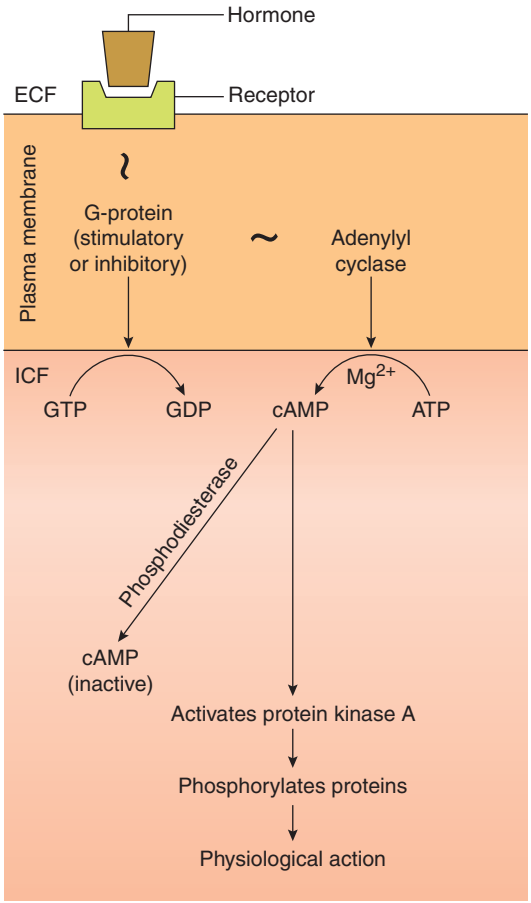


Fig. 8.1-8 Mechanism of action of hormone through adenylyl cyclase (cAMP) system as second messenger.

(iii) Activation of enzyme adenylyl cyclase (Step 3). The hormone–receptor complex via activated G-protein (stimulatory or inhibitory) either stimulates or inhibits the enzyme adenylyl cyclase, which is also located in the plasma membrane.

(iv) Formation of cAMP (Step 4). A part of the enzyme adenylyl cyclase protrudes through the inner surface of the cell membrane and when activated it catalyzes the formation of cAMP from cytoplasmic ATP with Mg^{2+} as cofactor. A stimulatory G-protein (Gs) therefore increases intracellular cAMP levels, whereas an inhibitory G-protein (Gi) decreases cAMP levels.

(v) Action of cAMP. The cAMP once formed stimulates a cascade of enzyme activation. One molecule of cAMP may stimulate many enzymes. Therefore, even a slightest amount of hormone acting on the cell surface can initiate a very powerful response. The cyclic AMP so formed initiates response by different mechanisms.

2. Guanylate cyclase–cGMP system

Group II-B hormones which act via second messenger cGMP include atrial natriuretic factor and nitric oxide.

(i) Synthesis of cyclic GMP is analogous to the formation of cAMP. Enzyme guanylate cyclase produces cGMP from GTP.

(ii) cGMP exerts its biochemical response through an enzyme protein kinase G, which when activated initiates a cascade of subsequent enzyme activations that is characteristic of this signaling system.

3. Membrane phospholipase–phospholipid system or IP_3 mechanism

Hormones which exert their response through this system constitute the so-called group II-C hormones (Table 8.1.1). Steps involved in this system are (Fig 8.1-9):

- Hormone binds to a receptor in the plasma membrane.
- The hormone–receptor complex via a G-protein activates the membrane enzyme phospholipase C.
- Activated phospholipase C then releases diacylglycerol and inositol triphosphate (IP_3) from the membrane phospholipid.
- Inositol triphosphate (IP_3) then mobilizes Ca^{2+} from the endoplasmic reticulum.
- Calcium ions (Ca^{2+}) and diacylglycerol together activate protein kinase C.
- Activated protein kinase C phosphorylates proteins and causes specific physiological action.
- Diacylglycerol also yields arachidonic acid, which serves as a substrate for rapid synthesis of prostaglandins that modulate cell response.

4. Calcium–calmodulin system

Hormones that act through this system as a second messenger are also included in the so-called group-II C hormone (Table 8.1-1). Steps involved in this system are (Fig 8.1-10):

- Hormone binds to a specific receptor in the plasma membrane, then
- The hormone–receptor complex, via G-protein opens the Ca^{2+} channels on the cell membrane and also activates mobilization of Ca^{2+} bound to the endoplasmic reticulum.
- Ca^{2+} binds to a specific binding protein the calmodulin in various proportions.
- The different calcium–calmodulin complexes activate or deactivate various calcium-dependent enzymes producing different physiological actions.

ACTION OF HORMONE VIA TYROSINE KINASE ACTIVATION

Certain hormones act by activating tyrosine kinase system and have been classified as group-II D hormones (Table 8.1-1). This mechanism of signal generation from the plasma membrane receptors does not require G-protein intermediaries. These receptors have an extracellular hormone binding portion, a single transmembrane portion and an intracytoplasmic C-terminal portion.

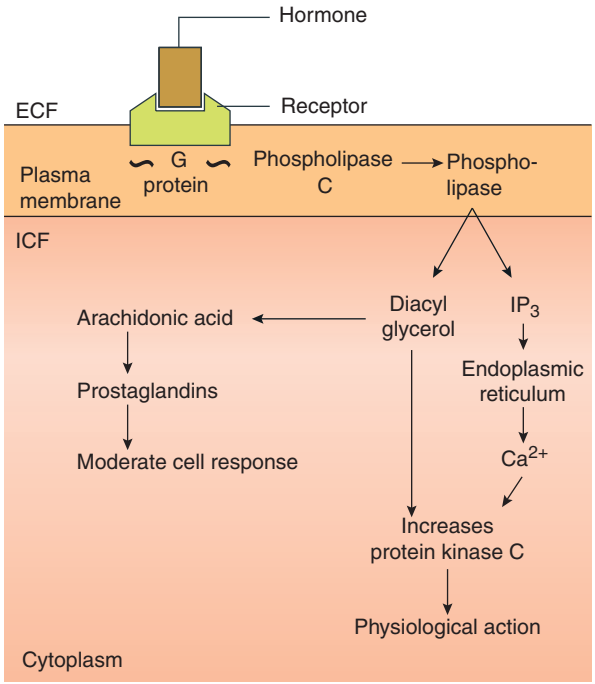


Fig. 8.1-9 Mechanism of action of hormone via membrane phospholipase–phospholipid system or IP_3 mechanism.

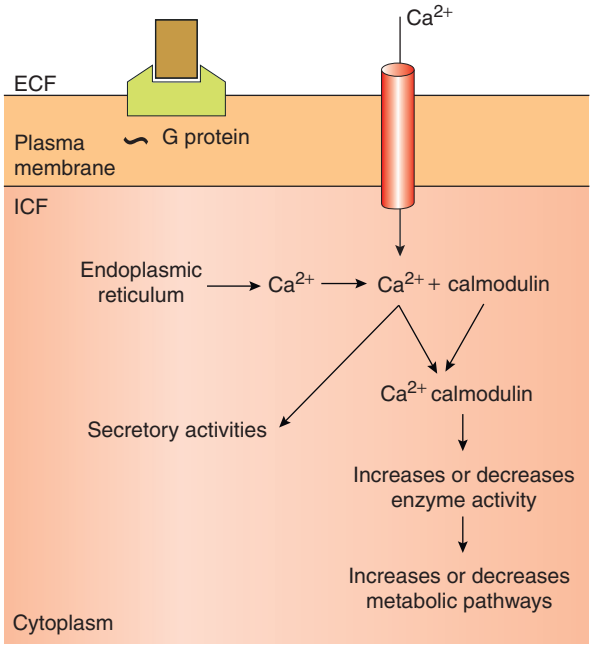


Fig. 8.1-10 Mechanism of action of hormone via calcium–calmodulin system.

The activation of tyrosine kinase occurs by two mechanisms:

1. Hormone receptors possessing intrinsic tyrosine activity, e.g. those for insulin and epidermal growth factor involve following steps (Fig. 8.1-11A):

- Binding of hormone to the receptor changes its conformation and exposes sites on its intracellular portion that are capable of receptor autophosphorylation at specific tyrosine sites.
- As a result, the receptor itself becomes a tyrosine kinase that phosphorylates tyrosine residue on the intracellular protein substrates.
- This latter activity sets into motion a cascade of events leading to an enzyme activation and gene transcription.

2. Hormone receptors that do not possess intrinsic tyrosine activity, e.g. those for growth hormone, prolactin-releasing hormones, cytokines, etc. act as follows (Fig. 8.1.11B):

- Hormone binding to extracellular portion of the receptor changes its intracytoplasmic tail.
- The changes produced in the intracytoplasmic tail of receptor exposes sites which attract and dock the intracytoplasmic tyrosine kinases [such as janus tyrosine kinases (JAK) and signal transducer and activator of transcription (STAT) kinases] and then activates them.
- The activated intracytoplasmic tyrosine kinases phosphorylate cytoplasmic substrates, such as transcription factor proteins and ultimately modulate gene expression.

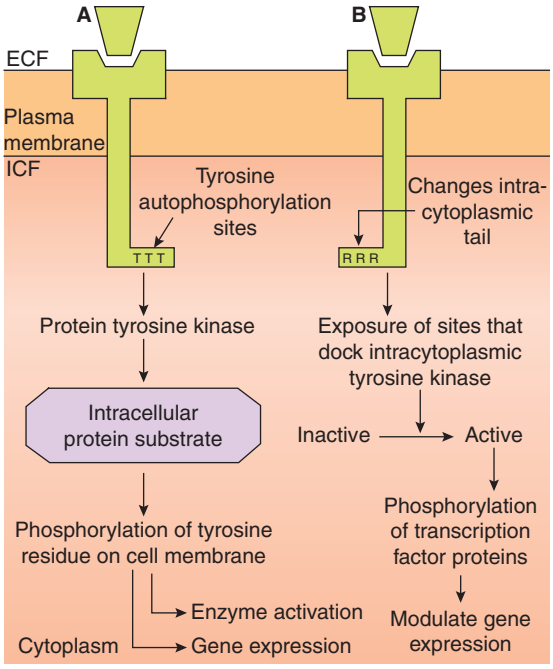


Fig. 8.1-11 Mechanism of action of hormone via tyrosin kinase activity: A, by receptors that possess intrinsic tyrosine activity and B, by receptors that do not possess intrinsic tyrosine activity.

MEASUREMENT OF HORMONES

Measurement of blood level of hormones is essential to confirm the endocrinal disorders associated with either deficiency or excess of a hormone. Since the hormones exist in the blood in very low concentration, the conventional methods of estimation such as colorimetry are not of much use. Therefore, they are measured by hormone assays and some special techniques which include:

- Bioassay,
- Immunoassay,
- Cytochemical assay and
- Dynamic tests.

BIOASSAY

In this method, hormone levels were assessed by injecting the unknown sample of plasma in the experimental animals and observing quantitatively the specific biological effect. The effect chosen was a characteristic action of the hormone for which a clear dose–response relationship existed.

IMMUNOASSAY

The immunoassay methods, frequently employed for estimation of hormone levels, include:

- Radioimmunoassay (RIA), and
- Enzyme-linked immunosorbent assay (ELISA).

1. Radioimmunoassay

The radioimmunoassay is performed as:

- An unknown sample of plasma in which the concentration of a particular hormone (H) to be estimated is mixed with commercially available purified specific antibody (anti-H) and an appropriate amount of the purified hormone tagged with radioactive isotope (H^+). The mixture is incubated in the cold.
- The antibodies have high affinity for the hormone. There occurs a competition between the free hormone (H) present in the unknown sample of plasma and the tagged hormone (H^+) for binding to the specific antibody (anti-H).

2. Enzyme-linked immunosorbent assay method

Enzyme-linked immunosorbent assay (ELISA) method is principally similar to RIA, i.e. it is also based on the principle of antigen–antibody reaction. Any antigen that is protein can be measured by this technique. In this method, radioactivity is not measured, instead specific antibody hormone (antigen) complex is stained with a suitable dye,

and the intensity of colour is measured by the spectrophotometer. This technique is useful in estimating peptide and steroid hormones.

CYTOCHEMICAL ASSAY

This test is much more sensitive than the immunoassay, but is cumbersome and time consuming and so rarely used. In this technique, genesis of hormone can be detected in slices cut out of the endocrine gland by incubating them in a culture medium. This test is very useful in measuring the minute basal levels of hormone secretion.

DYNAMIC TESTS

Dynamic tests are needed in certain situations when simple blood hormone level estimation is not enough. Two types of dynamic tests are:

Suppression type of dynamic tests are useful in certain conditions, e.g. to know whether a lung cancer is secreting ACTH.

Stimulation type of dynamic tests are useful in certain other conditions, e.g. metyrapone test is performed to know whether the corticotrophs of the pituitary (which secrete ACTH) are normally functioning or not.

Chapter 8.2

Endocrinal Functions of Hypothalamus and Pituitary Gland

INTRODUCTION AND FUNCTIONAL ANATOMY

The hypothalamic–pituitary unit forms a unique component of the entire endocrine system that regulates growth, lactation, fluid homeostasis and the functions of thyroid gland, adrenal glands and gonads.

GROSS ANATOMY AND DEVELOPMENT OF PITUITARY GLAND

Pituitary gland, also called hypophysis cerebri, is a small gland, weighs about 0.5 g and is approximately 1 cm in diameter. It is situated in the hypophyseal fossa (sella turcica) of the sphenoid bone.

Development of pituitary gland

Anterior pituitary is ectodermal in origin. It develops from the *Rathke's pouch,* which is an embryonic upward outpouching from the roof of the primitive oral cavity.

Posterior pituitary or neurohypophysis develops from a lowered outpouching of neuroectodermal tissue from the central areas of the hypothalamus (tuber cinereum and median eminence).

From the above, it is quite clear that the anterior and posterior pituitary develops independently from widely different origins, and it is only a coincidence that when fully formed, they happen to lie so close together that they are considered parts of the same organ.

Parts of pituitary gland

Physiologically, the pituitary gland consists of three distinct parts or lobes (Fig. 8.2-1):

- Anterior lobe or adenohypophysis,
- Posterior lobe or neurohypophysis and
- Intermediate lobe or pars intermedia.

Adenohypophysis. The glandular anterior lobe of the pituitary gland is called adenohypophysis. It constitutes about

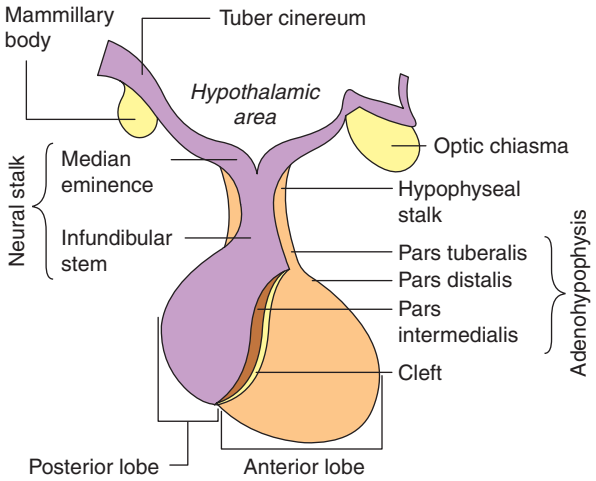


Fig. 8.2-1 Anatomical subdivisions of the pituitary gland.

80% of the pituitary gland. It can be further divided into three parts (Fig. 8.2-1):

- *Pars distalis.* It forms the main bulk of the anterior lobe and is highly vascular area.
- *Pars intermedia.* It is an avascular zone that lies between pars distalis and neurohypophysis. In human beings, this area is rudimentary, but in lower animals it forms the intermediate lobe of the pituitary.
- *Pars tuberalis.* It is the most vascular zone and contains many secretory cells. Superficially, it is surrounded by the pituitary stalk.

Neurohypophysis. The posterior lobe of the pituitary is a neural structure and hence called neurohypophysis. It consists of three parts (Fig. 8.2-1):

- *Pars posterior.* It is also called pars nervosa or neural lobe or infundibular process and forms the main bulk of neurohypophysis.
- *Infundibular stem.* It is the funnel-shaped extension arising from the median eminence at the floor of third ventricle.
- *Median eminence.* It is a small protrusion from the base of hypothalamus (tuber cinereum). It is situated just beneath the third ventricle and is highly vascular.

Pituitary stalk. The median eminence and infundibulum constitute the neural stalk. The posterior pituitary maintains its neural connection with the hypothalamus by this neural stalk. The neural stalk surrounded by pars tuberalis of adenohypophysis constitutes the pituitary stalk.

Intermediate lobe of pituitary gland is rudimentary in humans as well as in a few other mammalian species. In certain lower animals, this lobe secretes melanocyte-stimulating hormone (MSH) in response to changes in exposure to light and other environmental factors.

HISTOLOGICAL STRUCTURE OF PITUITARY GLAND

Adenohypophysis

Pars distalis consists of cords of cells separated by fenestrated sinusoids. The cells can be divided into two main types: the chromophobes and chromophils.

1. Chromophobes. These are agranular cells and it is considered that the chromophils are derived from the chromophobes.

2. Chromophils. These are granular cells, constitute 50% of the cells of anterior pituitary. Chromophils are further classified as: acidophils (35%) and basophils (15%).

(i) Acidophilic cells (α cells). The granules of these cells are acidophilic. Depending on the size and nature of granules, the acidophils are further divided into following subtypes:

- *Somatotrophs* and
- *Mammotrophs or lactotrophs.*

(ii) Basophilic cells (β cells). The granules of these cells are basophilic. The basophils are also further divided into functional subtypes:

- *Corticotrophs,*
- *Thyrotrophs* and
- *Gonadotrophs Delta (δ) cells.*

3. Folliculostellate cells. These cells send processes between the secretory cells. Recently, it has been demonstrated that these contain and secrete the cytokine 1L-6, but their physiological role is still not clear.

Pars tuberalis. It mainly consists of undifferentiated cells with few acidophils and basophils.

Pars intermedia. It contains β cells, few secretory cells and chromophobe cells.

Neurohypophysis

Histologically, posterior pituitary contains following structures:

1. Unmyelinated nerve fibres. These are the axons of the neurons located in the supraoptic and paraventricular nuclei of the hypothalamus. These carry precursor of posterior pituitary hormones and end as close terminals near the blood capillaries (Fig. 8.2-2).

2. Pituicytes are the special type of supporting cells, having long dendritic processes. These are present in between the axons.

3. Glial cells like astrocytes and oligodendrocytes are also seen.

BLOOD SUPPLY OF PITUITARY GLAND

Arterial supply

The arterial blood to the pituitary gland is supplied by the branches of:

- Internal carotid arteries (superior and inferior hypophyseal branches),
- Anterior cerebral artery and
- Posterior cerebral artery.

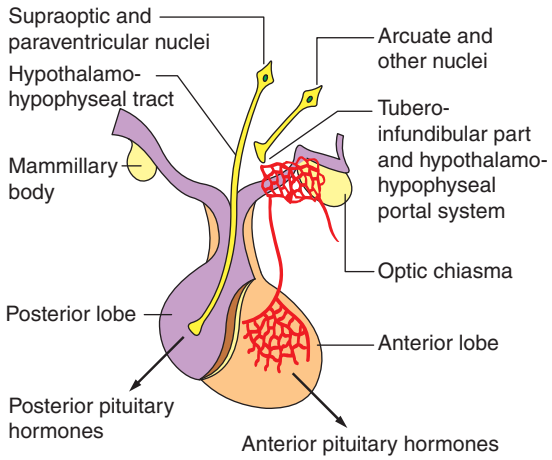


Fig. 8.2-2 Anatomical and functional relationship between the hypothalamus and pituitary gland through hypothalamo–hypophyseal tract and hypothalamo–hypophyseal portal system.

Hypothalamo–hypophyseal portal system (Figs 8.2-2 and 8.2-3)

- The branches from superior hypophyseal artery form a ring around the upper part of the pituitary stalk and further branched to form a *capillary network.*
- The blood from this capillary network is drained by *long portal veins* in the infundibulum.
- Then in the anterior lobe, these long portal veins break up into another set of capillary network and represented as *sinusoids of pars anterior.* This arrangement is called *hypothalamo–hypophyseal portal system.*
- The inferior hypophyseal (branch of internal carotid artery) branches to form a capillary network at the lower end of the infundibulum stem.
- The short portal vessels arise from this capillary network and supply blood mainly to the posterior pituitary and some part of the anterior pituitary.
- The short portal vessels provide link between the anterior and the posterior pituitary.

Venous drainage

The blood from the anterior pituitary is drained to the cavernous sinus and then into the jugular vein.

Note. The anterior pituitary lies outside the blood–brain barrier, hence it is accessible to influences from general circulation (hormones and neurotransmitter by the brain and hormones secreted in the general circulation).

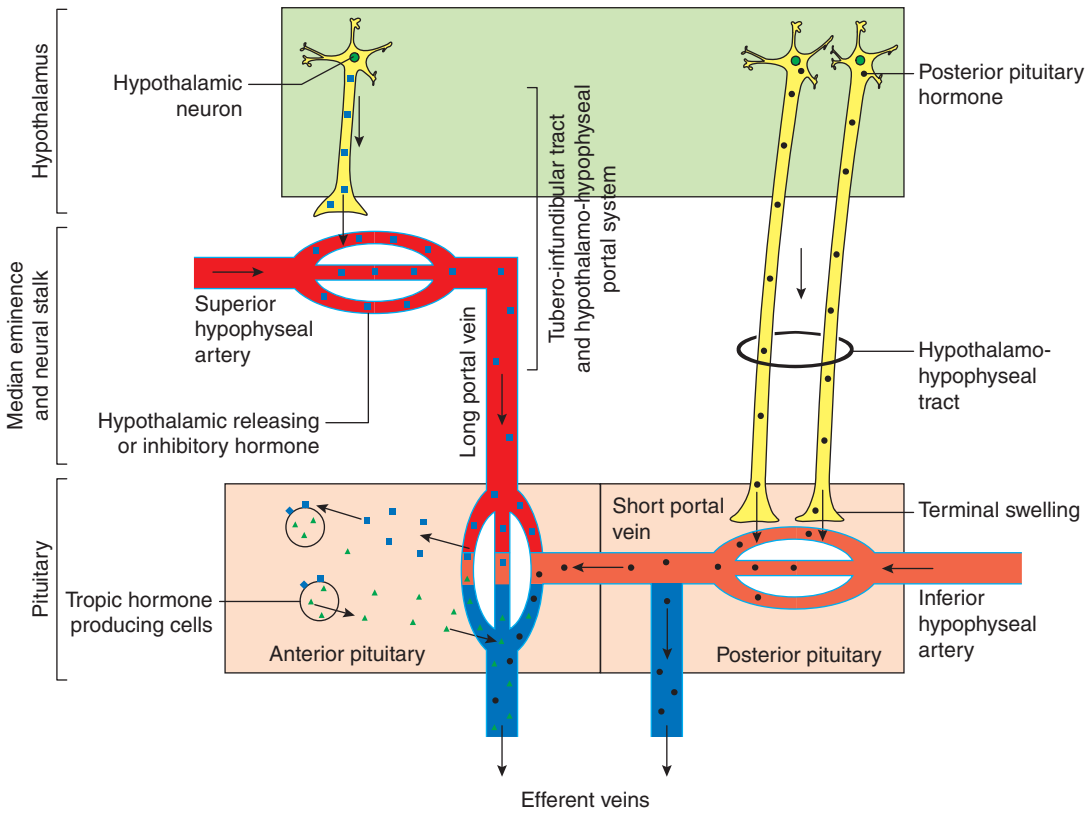


Fig. 8.2-3 Schematic diagram to explain anatomical and functional relationship between the hypothalamus and pituitary gland.

HYPOTHALAMIC–PITUITARY RELATIONSHIP

The influences from the hypothalamus are conveyed to the pituitary gland by two different tracts:

1. Hypothalamo–hypophyseal tract. It is composed of axons of the large neurosecretory cells of the supraoptic and paraventricular nuclei of the hypothalamus. These fibres pass to neurohypophysis through the infundibular stem and form a series of dilated terminals known as *Herring bodies* (Fig. 8.2-2). The neurosecretory cells of supraoptic and paraventricular nuclei secrete peptide hormones (vasopressin and oxytocin), which travel down their axons in the neurosecretory granules to be stored in the nerve terminals lying the neurohypophysis. Upon stimulation of the cell bodies, the granules are released from the axonal terminals by exocytosis. The peptide hormones then enter the peripheral circulation via the capillary plexuses of inferior hypophyseal artery (Fig. 8.2-3). Thus, a single neural cell performs the entire process of hormone synthesis, storage and release.

2. Tubero-infundibular tract and hypothalamo–hypophyseal portal system. It consists of fibres arising from the arcuate nuclei of the tuberal region of the hypothalamus and extends to the median eminence (Fig. 8.2-2). The cell bodies of these hypothalamic neurons synthesize certain releasing and inhibiting hormones, which are conveyed by the tubero-infundibular tract to the median eminence region where they are stored in the nerve terminals. After these hypothalamic neurons are stimulated by nerve impulses, the releasing or inhibiting hormones are discharged into the median eminence and enter the capillary plexus of the superior hypophyseal artery. From here they are transported down the portal vessels (long portal veins) and then exit from the secondary capillary plexus to reach the specific endocrine target cells in the adenohypophysis where they regulate the secretion of tropic hormones of anterior pituitary (Fig. 8.2-3).

ENDOCRINAL ASPECTS OF HYPOTHALAMUS

FUNCTIONAL ANATOMY

Hypothalamus is a specialized centre in the brain that functions as a master co-ordinator of hormonal action. It is a part of the brain situated below the thalamus and is very closely connected to the pituitary gland as described above. Thus hypothalamus provides an important link between the endocrine system and the nervous system. Before proceeding further see *details of functional anatomy of hypothalamus* at page 739.

ENDOCRINAL FUNCTIONS OF HYPOTHALAMUS

The hypothalamus serves its endocrinal functions through the neurosecretory cells, which are arranged in different nuclei of hypothalamus.

The main endocrinal functions of hypothalamus are:

- Control of anterior pituitary function and
- Control of posterior pituitary function.

1. Control of anterior pituitary function

Hypothalamus controls the functioning of anterior pituitary through various hypothalamic–hypophysiotropic hormones, i.e. various hypothalamic-releasing and -inhibiting hormones. The hypothalamic-releasing and -inhibiting hormones are released in response to neural stimuli. The hypothalamus receives afferent nerve tracts from the thalamus, the reticular-activating system, the limbic system, the eyes and remotely from the neocortex (Fig. 8.2-4). Through these inputs (Fig. 8.2-4), the pituitary functions can be influenced by pain, sleep, wakefulness, emotion, fright, rage, olfactory sensations, light and possibly even thought.

Various hypothalamic-releasing and -inhibiting hormones include:

- Growth-hormone-releasing hormone (GHRH),
- Growth-hormone-inhibiting hormone (GRIH), also called somatostatin,
- Corticotropin-releasing hormone (CRH),
- Thyrotropin-releasing hormone (TRH),
- Gonadotropin-releasing hormone (GnRH),
- Prolactin-releasing hormone (PRH) and
- Prolactin-inhibiting hormone (PIH).

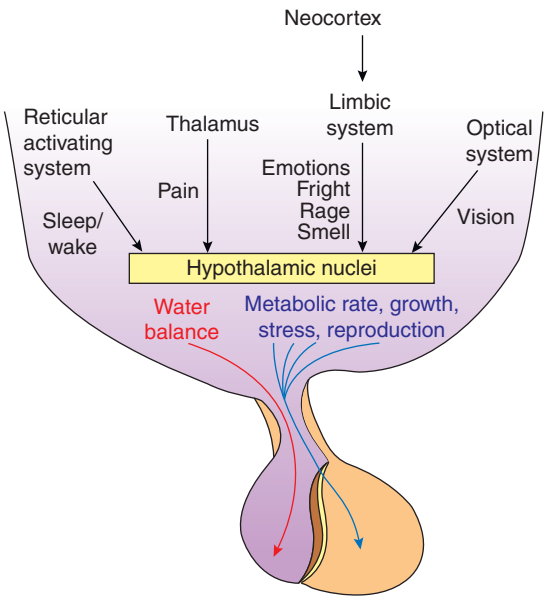


Fig. 8.2-4 Afferent impulses to different hypothalamic nuclei leading to release of hypothalamic releasing or inhibiting hormones controlling pituitary gland.

Functions of hypothalamic-releasing and -inhibiting hormones have been described at different places in the context [e.g. thyrotropin-releasing hormone, i.e. TRH, has been discussed along with thyroid-stimulating hormone (TSH) and thyroid hormones in Chapter 8.3 on 'Thyroid Gland']. In general, they control the release of various tropic hormones from the anterior pituitary.

2. Control of posterior pituitary function

The large (magnocellular) neurosecretory cells forming the supraoptic and the paraventricular nuclei of the hypothalamus are responsible for synthesis of the two posterior pituitary peptide hormones [oxytocin and antidiuretic hormone (ADH)]. These hormones reach the posterior pituitary through hypothalamic–hypophyseal tract described above (Figs 8.2-2 and 8.2-3).

ANTERIOR PITUITARY HORMONES

Anterior pituitary is truly the master endocrine organ. The hormones of adenohypophysis are broadly classified into three categories:

I. Hormones of growth hormone family include:
- Growth hormone (GH) and
- Prolactin (PRL).

II. Glycoprotein hormone family. The hormones of glycoprotein family secreted by the anterior pituitary include:

- Thyroid-stimulating hormone (TSH),
- Luteinizing hormone (LH) and
- Follicular-stimulating hormone (FSH).

III. Pro-opiomelanocortin peptides. The hormones of this group are:

- Adrenocorticotropic hormone (ACTH),
- Melanocyte-stimulating hormone (MSH), α-MSH and β-MSH are produced in the intermediary lobe (which is rudimentary in humans),
- β-lipotropin and
- β-endorphin.

Physiological aspects of growth hormone are discussed in detail in this chapter. Other hormones are discussed in different chapters (e.g. TSH with thyroid hormone and ACTH with adrenal gland).

GROWTH HORMONE

Growth hormone (GH), also called *somatotropin,* is the most important hormone for post-natal growth and development to adult size. It also helps to maintain lean body mass and bone mass in adults.

STRUCTURE, SYNTHESIS AND SECRETION

Structure. GH consists of a single unbranched chain containing 191 amino acids. Its molecular weight is 2000.

Species specificity. Growth hormones obtained from different species show chemical and immunological variations, i.e. exhibit species specificity. Because of species specificity, the bovine and porcine growth hormones do not even have a significant transient effect on growth in humans and monkeys. However, in humans, the human growth hormone and monkey growth hormone have similar biological activity. Therefore, growth hormone preparation from monkey origin is therapeutically effective.

Synthesis. GH is synthesized by acidophilic cells called *somatotrophs* of anterior pituitary.

Secretion. GH is released in *pulsatile* fashion.
- *Secretion is increased* by sleep, stress, hormones related to puberty, starvation, exercise and hypoglycaemia.
- *Secretion of GH is decreased* by somatostatin, somatomedins, obesity, hyperglycaemia and pregnancy.

REGULATION OF GH SECRETION

1. Hypothalamic control

Hypothalamus controls GH secretion by releasing two hormones, growth-hormone-releasing hormone (GHRH) and growth-hormone-release-inhibiting hormone (GRIH) (Fig. 8.2-5).

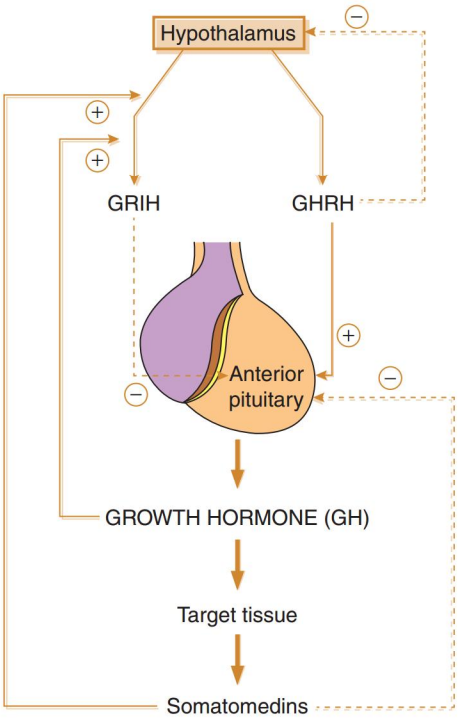


Fig. 8.2-5 Control of growth hormone secretion. GHRH (growth-hormone-releasing hormone) and GRIH (growth-hormone-release-inhibiting hormone).

Growth-hormone-releasing hormone (GHRH). It is a polypeptide with 44 amino acids. It stimulates the secretion of GH from the anterior pituitary.

Mechanism of action. GHRH acts through guanylate cyclase, which releases cyclic GMP, which in turn stimulates the release of GH from the anterior pituitary. Influx of Ca^{2+} into the pituitary cells is an essential event associated with the GHRH-stimulated release of GH.

Factors stimulating GHRH secretion and thus increasing GH release are:

- *Hypoglycaemia* increases GHRH secretion through the glucoreceptor cells in the ventromedial nucleus of the hypothalamus. Neurotransmitter involved is epinephrine.
- *Emotions, exercise and physical stress* (pain, trauma, cold, surgery, inflammation, etc.) stimulate GHRH release through nervous pathways (therefore, the effect is seen within a couple of minutes).
- *Slow wave phase of sleep* is associated with increase in GHRH. The neurotransmitter involved is serotonin.
- *Increase in the plasma levels of certain amino acids*, such as arginine (after protein meal or infusion of amino acids) increase GHRH secretion by α-adrenergic stimulation of the receptors in neurons that release GHRH.
- *Growth-hormone-releasing peptide* (GHRP) also called 'ghrelin' increases GHRH secretion. GHRP is synthesized by oxyntic glands of stomach. It increases GH release by its direct action on the anterior pituitary.

Growth-hormone-release-inhibiting hormone (GRIH), also called somatostatin, is a polypeptide with 14 amino acids. It inhibits the release of GH from the anterior pituitary.

Factors stimulating GRIH secretion and thus decreasing GH secretion are:

- Hyperglycaemia and
- High plasma free fatty acids (FFA) concentration.

2. Negative feedback control of GH secretion

The negative feedback control mechanism for GH involves the role of somatomedins, GH and GHRH (Fig. 8.2-5).

(i) Negative feedback control by somatomedins. Somatomedins are insulin-like growth factors (IGFs) that are produced when growth hormone acts on the target tissues. Somatomedins inhibit the secretion of GH directly or by stimulating the secretion of somatostatin from the hypothalamus.

(ii) Negative feedback control by GH. Growth hormone also inhibits its own secretion by stimulating the secretion of somatostatin from the hypothalamus.

(iii) Negative feedback control by GHRH. GHRH inhibits its own secretion from the hypothalamus. This mechanism called *ultra-short feedback loop.*

3. Other factors controlling GH secretion

Other factors which control GH secretion are:

- *Thyroxine and cortisol* at their basal levels synergistically stimulate GH.
- *Insulin* represses GH gene expression.
- *Placental GH and placental lactogen* are responsible for decreased GH secretion noted during later part of the pregnancy.
- *Obesity* is associated with dampened GH responses to all stimuli including GHRH itself.
- *Neurotransmitters,* dopamine, norepinephrine, acetylcholine, serotonin, GABA and histamine, all increase GH secretion by stimulating the release of GHRH or by blocking the release of somatostatin.
- *Oestradiol* increases GH secretion and explains the greater secretion of GH in pre-menopausal women than in men.

PLASMA LEVELS, BINDING AND METABOLISM

Plasma levels

Basal plasma GH level varies from 2 to 4 ng/mL. Its concentration graph shows fluctuations, i.e. after every 1–2 h interval there is rise in plasma GH level.

Diurnal variation in plasma levels of GH is noted. The nocturnal peak occurs 1–2 h after deep sleep (which corresponds to stage three or stage four of slow wave sleep). The nocturnal sleep bursts account for nearly 70% of the daily GH secretion. These secretory bursts are greater in children and decrease with age.

Variation in plasma GH levels with age

- *From birth to early childhood,* plasma GH levels increase progressively.
- *Children versus adults.* In general, children have only slightly higher plasma GH levels than adults.
- *Puberty* is associated with a peak period of plasma GH levels.
- *Senescence* is associated with a reduction in GH secretion in response to GHRH and other stimuli.

Circulation, half-life and metabolism

Circulation. Circulating GH is bound to a plasma protein (GH binding protein).

Half-life of circulating GH in humans is 0–20 min, and the daily GH output has been calculated to be 0.2–1.0 mg/day in adult.

Metabolism. Growth hormone is rapidly metabolized, probably at least in part in the liver. Metabolic clearance rate is 350 L/day. Daily urinary GH excretion correlates well with the integrated 24 h plasma GH profile.

GROWTH HORMONE RECEPTORS AND MECHANISM OF ACTION OF GH

Growth hormone receptors

Growth hormone receptors of various sizes are present on the cell membrane in target tissues, including the liver and adipose tissue. The GH receptor belongs to the cytokine family of receptors. It comprises a large extracellular portion, a transmembrane domain and a large intracellular cytoplasmic portion.

Mechanism of action of GH

Growth hormone promotes growth by its direct and indirect actions on cartilage. Growth hormone produces IGF locally and also converts stem cells to the cells that respond to IGF-I. The circulating as well as locally produced IGF-I makes the cartilage to grow.

Growth hormone stimulates IGFs production by gene expression by tyrosine phosphorylation of signal transducer and activator of transcription (STAT) (see page 533).

Insulin-like growth factors

As mentioned above, IGF ultimately exerts its growth promoting effect via peptide mediators [(insulin-like growth factors (IGFs) or somatomedins] that are produced in the liver and many GH target cells.

Structure. IGFs are closely related to insulin except that their C chains are not separated and they have an extension of A chain called D domain.

Types. In humans, two types of IGF are known IGF-I and IGF-II. Variants of IGF-I and IGF-II are also known.

The characteristic features of IGF-I and II are depicted in Table 8.2-1

ACTIONS OF GROWTH HORMONE

Growth hormone promotes growth and also influences the normal metabolism; therefore, besides acting on one specific organ, its actions are generalized (Fig. 8.2-6).

1. Growth promoting actions of GH

Growth hormone promotes linear growth of an individual by its effects on the bone, cartilage and other connective tissues.

Effects on cartilage. GH stimulates the proliferation of chondrocytes (cartilage cells) present in the epiphyseal end plates of long bones.

Effects on bone. GH stimulates the osteoblastic activity which converts cartilage into bone. This process continues up to adolescence till there is fusion of the epiphyseal end plate with shaft of the bone. The bone mass also increases during this period.

2. Metabolic actions of GH

(i) Effects on protein metabolism. Growth hormone has an anabolic effect on the protein metabolism. It promotes the protein deposition in the tissues by following effects:

- Increases the rate of amino acid uptake into the cells.
- Increases protein synthesis in the ribosomes.
- Stimulates transcription (RNA synthesis from DNA).

The overall effect of GH on the protein metabolism is positive nitrogen balance that leads to an increase in the body weight. In addition, GH also decreases protein breakdown as well as the rate of amino acid degradation for energy purposes.

(ii) Effects on fat metabolism. GH promotes lipolysis in adipose tissue (catabolic effect) and then increases fat utilization for energy.

(iii) Effects on carbohydrate metabolism. GH is antagonistic to insulin and produces hyperglycaemia by following effects on the carbohydrate metabolism:

- Increases gluconeogenesis, i.e. increases hepatic glucose output,
- Decreases the uptake as well as utilization of glucose by the tissues for energy production and
- Inhibits glycolysis and thus glycogen stores tend to increase. This occurs as a consequence of increased mobilization and use of FFA for energy production.

Table 8.2-1 Characteristics of insulin-like growth factors

IGF-I	IGF-II
1. IGF-I is also called Somatomedin-C	IGF-II is known as Multiplication Stimulating Activity (MSA).
2. Secretion • Before birth, IGF-I secretion is independent of GH and • After birth, its secretion is stimulated by GH and its plasma concentration is correlated with GH (i.e. low during childhood, reaches to peak at puberty and declines with increasing age)	Secretion of IGF-II is independent of GH and its level remains constant.
3. Receptors of IGF-I are similar to the insulin receptors.	Receptors of IGF-II are mannose-6 phosphate receptors.
4. Its major role is in the skeletal and cartilage growth.	IGF-II plays a major role in fetal growth.

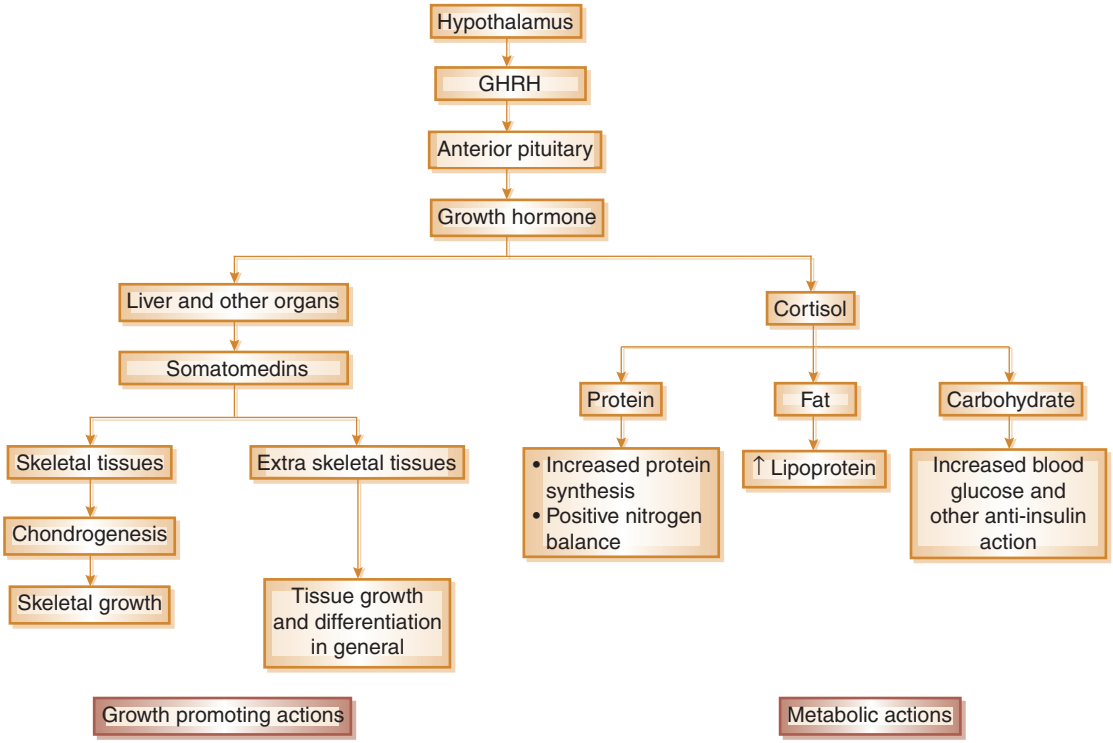


Fig. 8.2-6 Growth promoting and metabolic actions of growth hormone.

(iv) Effects on mineral metabolism. Growth hormone promotes bone mineralization in growing children. This effect of growth hormone is probably mediated through IGF-I, which causes positive balance of calcium, phosphate and magnesium. It promotes renal absorption of Ca^{2+}, phosphate and Na^+. It also promotes the retention of Na^+, K^+ and Cl^- in the body.

3. Effect on lactation

Growth hormone enhances milk production in lactating animals. Growth hormone acts like prolactin, therefore this action is referred as prolactin-like effect of growth hormone.

HUMAN PROLACTIN

STRUCTURE, SECRETION AND PLASMA CONCENTRATION

Structure and secretion. Human prolactin is a single peptide chain, secreted by acidophilic cells of anterior pituitary gland.

Plasma concentration. The prolactin secretion is pulsatile, shows diurnal variations (secretion increases about one hour after the onset of sleep and continues throughout the sleep period). Its basal average value varies in different conditions (Table 8.2-2).

During pregnancy, prolactin secretion starts rising from eighth weeks onwards and peak value (200–400 ng/mL) is reached at term. The sources of prolactin during pregnancy are placenta, amniotic fluid and maternal anterior pituitary gland. The prolactin secretion during pregnancy and during lactation is affected by *oestrogen*. Prolactin secretion parallels with secretion of oestrogen, i.e. 7–8th weeks gestation onwards, oestrogen secretion rises along with prolactin. This is due to oestrogen inhibition of hypothalamic prolactin inhibitory factor (PIF).

Table 8.2-2 Range of plasma concentration of prolactin

Condition	Plasma conc. (ng/mL)
• Prepubertal period and after menopause	2–8
• Fertile period (16–45 years)	9–14
• Early pregnancy (8 weeks)	10–25
• Late pregnancy (at term)	200–500
• Lactation period	
– Immediately after birth to 10 days	200–400
– 10–90 days (1st week–3 months)	70–200
– 90–180 days (3–6 months)	100–250
– 6 month–1 year	30–40

Amniotic fluid concentration may be up to 10,000 mg/mL.

CONTROL OF PROLACTIN SECRETION

Hypothalamic control. Secretion of prolactin from anterior pituitary is controlled by the hypothalamus. A PIF formed in

the arcuate nucleus of the hypothalamus is transported through hypothalamo–hypophyseal portal system to anterior pituitary where it checks the synthesis and release of prolactin.

IMPORTANT NOTE

Prolactin inhibitory factor has been identified as dopamine. Therefore, the substances like dopamine agonists (bromocriptine) and serotonin antagonists block the secretion of prolactin. Therapeutically, bromocriptine is used during post-partum period for reducing prolactin level to inhibit lactation.

Factors enhancing the release of prolactin are:

- *TRH.* There is prolactin-releasing factor for prolactin, but thyrotropin-releasing hormone (TRH) also causes release of prolactin from the anterior pituitary.
- *Stress.* Psychological stress, physiological stress (exercise, pregnancy and lactation) and pathological stress increase prolactin secretion.
- *Substances* like dopamine antagonists (phenothiazine and tranquillizers), adrenergic blockers, serotonin agonists stimulate prolactin release.
- *Role of oxytocin.* Oxytocin acts directly on the acidophilic cells of anterior pituitary to stimulate prolactin release.
- *Sectioning of pituitary stalk* or lesion, which interfere with pituitary portal circulation also increases prolactin secretion (because secretion is tonically inhibited by hypothalamus).

PHYSIOLOGICAL EFFECTS OF PROLACTIN

1. Breast growth. During pregnancy, it increases the breast growth particularly of alveolar tissue in the form of alveolar distension, dilatation of mammary vessels and formation of new capillaries (see page 675).

2. Lactogenic effect. Prolactin acts on the alveolar epithelium and stimulates the secretory activity. For lactogenic effect, prolactin acts by two ways:

- Directly by attaching on the surface of the alveolar epithelial cells and
- By binding on to the receptors on the membrane of epithelial cells.

During pregnancy, the lactogenic effect is suppressed by high concentration of oestrogen and progesterone. The exact mechanism involved in suppressing the lactogenic effect is not known but probably by inhibiting the binding of prolactin to its receptors and onto the surface of the cell or by inhibiting the translocation of prolactin into the nucleus of the cell.

After parturition, the lactogenic effect of prolactin is enhanced because of following reasons:

- The inhibitory factors are withdrawn,
- Oxytocin level is increased.

Note. The women who do not wish to feed their babies or when baby dies immediately after birth, in these situations, oestrogen is administered to stop lactation.

3. Suppression of ovarian cycle in nursing mothers. Prolactin inhibits the secretion of gonadotropin-releasing hormone (GnRH) from the hypothalamus. Therefore, gonadotropin (FSH and LH) secretion from the anterior pituitary also decreases. Thus in nursing mothers due to low levels of gonadotropins the ovarian cyclic changes do not occur. For details see page 678.

APPLIED ASPECTS: ABNORMALITIES OF ANTERIOR PITUITARY HORMONES

The abnormalities related to pituitary hormones occur either due to excess or deficiency of the hormones secreted. The most common causes of pituitary hormone disturbances are pituitary tumours, which may cause symptoms of excess of one or more hormones and simultaneous deficiency of other hormones, hence a mixed picture may evolve. The various hormones of pituitary, their site of action and disease produced by them are given in Table 8.2-3.

Pituitary disorders seen in clinical practice are:

- Hypopituitarism.
- Abnormalities of growth hormone.
- Prolactin deficiency.
- Cushing's syndrome (see page 593).

HYPOPITUITARISM

Hypopituitarism is a clinical condition of hyposecretion of one or more pituitary hormones. Hypopituitarism can be due to hypothalamic causes or pituitary causes. Since anterior pituitary has a large reserve, the endocrine abnormalities are produced only when large part of pituitary is destroyed.

Effects of hypopituitarism

Since anterior pituitary has a large reserve, the endocrine abnormalities are produced only when large part of pituitary is destroyed. The effects of hypopituitarism are:

1. GH deficiency. This appears first of all with a progressive loss of pituitary tissue. Effects of hyposecretion of GH are described on page 545.

2. Gonadotrophin secretion. This is decreased when 70–90% of anterior pituitary is destroyed. It leads to gonadal atrophy decreasing sex hormone levels which causes:

- In males, loss of spermatogenesis, loss of libido, impotency and gynaecomastia.

Table 8.2-3 Pituitary hormones: site of action and diseases associated with their deficiency and excess

Hormone	Site of action	Diseases	
		Excess	Deficiency
Anterior pituitary			
Growth hormone (GH)	All somatic cells	Gigantism (in adolescents) Acromegaly (in adults)	Dwarfism
Adrenocorticotropic hormone (ACTH)	Adrenal cortex	ACTH-dependent Cushing's syndrome	Hypoadrenalism (rare)
Thyroid-stimulating hormone (TSH)	Thyroid	Hyperthyroidism	Hypothyroidism
Prolactin	Breast	Hyperprolactinaemia	–
Gonadotropins	Gonads	Hypergonadism	Hypogonadism
Melanocytic-stimulating hormone (MSH)	Skin	Hyperpigmentation	–
Posterior pituitary			
Antidiuretic hormone	Kidneys	–	Diabetes insipidus

- In females, abolition of ovulation and stoppage of menstrual cycle results in sterility.
- Some of the secondary sex characteristics disappear (specially loss of axillary and pubic hair in both the sexes).
- Urinary gonadotropin excretion stops within 2 weeks–2 months.

3. Thyrotropic hormone secretion. This is decreased leading to impairment of thyroid function when 90–95% of anterior pituitary is destroyed. Clinical features of hypothyroidism due to hypopituitarism are less marked. Tolerance to cold *is rare*. Frank myxoedema is not seen.

4. Adrenocorticotropic hormone deficiency. This leading to an atrophy of adrenal cortex and adrenal insufficiency occurs when almost whole of the anterior pituitary is destroyed. Adrenocorticotropic hormone (ACTH)-dependent Addison's disease is produced. For details see page 595. In brief, there occurs:

- Pallor of the skin due to decreased ACTH.
- Sensitivity to the stress is increased due to decreased glucocorticoids.
- Mineralocorticoid deficiency does not occur, as secretion of aldosterone is controlled by renin secreted from juxtaglomerular apparatus. So, salt loss and hypovolaemic shock do not occur.

5. Effect on water metabolism. Though deficiency of ADH, the posterior pituitary hormone produces diabetes insipidus, but removal of both anterior and posterior pituitary usually cause no more than a transient polyuria. This occurs because of following effects:

- Fewer osmotically active products of catabolism are filtered (because ACTH deficiency decreases rate of protein catabolism and TSH deficiency decreases metabolic rate).
- GH deficiency also contributes to the depression of the glomerular filtration rate.

As a consequence of the above, the urine volume decreases, even in the absence of ADH. The diuretic activity of anterior pituitary thus can be explained in terms of the effects of decreased ACTH, TSH and GH levels.

6. Effect on insulin sensitivity. Sensitivity to insulin is markedly increased in hypophysectomized animals. It occurs because of two reasons:

- Due to deficiency of adrenocortical hormones and
- Due to lack of anti-insulin effect of GH.

ABNORMALITIES OF GROWTH HORMONE SECRETION

The abnormalities of growth hormone secretion include:

- Hypersecretion of GH and
- Hyposecretion of GH.

Hypersecretion of GH

Hypersecretion of GH occurs in tumours of acidophilic cells (particularly of somatotrophs) of the anterior pituitary. Depending upon the age of an individual, excess of GH may cause:

- Gigantism and
- Acromegaly.

1. Gigantism

It is a clinical condition resulting from the hypersecretion of GH in growing children before the closure of epiphysis of long bones.

Clinical features include (Fig. 8.2-7):

- *Abnormal height.* Excessive growth of long bones results in abnormal height (7–8 feet) with huge status and the person becomes 'giant'.
- *Large hands and feet.*
- *Coarse facial features,* i.e. thick lips, macroglossia and thickness.
- *Bilateral gynaecomastia* (enlargement of breasts in male may be associated).
- *Loss of libido/impotence* may be there.

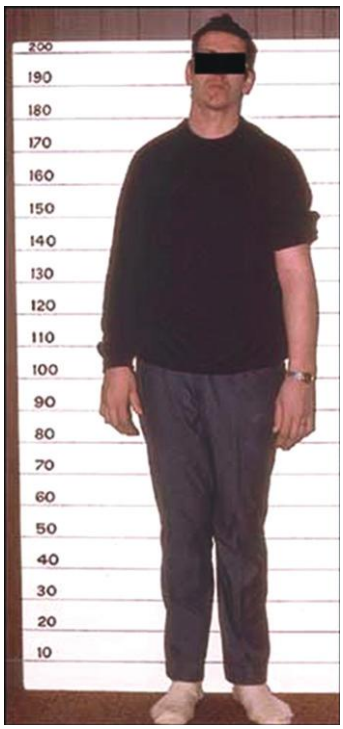


Fig. 8.2-7 Photograph showing tall stature in a patient with gigantism.

- *Hyperglycaemia* caused by GH leads to excess insulin secretion. Overactivity of beta cells of pancreas ultimately leads to degeneration of these cells and deficiency of insulin resulting in diabetes mellitus.

Features due to tumour mass include:
- Headache,
- Visual field defects,
- Cranial nerve palsies, and
- Enlargement of pituitary fossa with destruction of clinoid processes may be detected on radiograph of skull.

2. *Acromegaly*

It is a clinical condition that occurs due to excess of GH in adults (after epiphyseal closure of long bones has occurred) and causes excessive growth in those areas where cartilage persists.

Clinical features are (Fig. 8.2-8):

- *Acromegalic face,* which is characterized by thick lips, macroglossia, broad and thick nose, prominent eyebrows, thickened skin, and coarse facial features.
- *Prognathism,* i.e. protrusion of the lower jaw due to elongation and widening of mandible associated with increased spacing of the teeth.
- *Acral part abnormalities* include large spade-like hands, thick wide fingers, large feet with an increase in size of the shoes. Height is normal, build is stout and stocky.
- *Kyphosis* may occur due to improper vertebral growth.

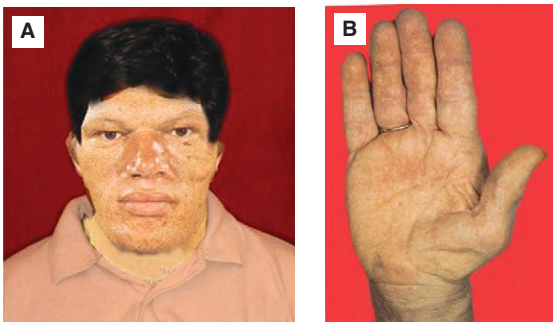


Fig. 8.2-8 Clinical features of acromegaly: A, note coarse facial features, broad thick nose, prognathism, prominent eyebrows and thickened skin and B, large spade-like hand with short, thick wide fingers.

- *Excessive growth of internal organs,* i.e. cardiomegaly, hepatomegaly, splenomegaly and renomegaly may be associated.
- *Increased sympathetic activity* may cause increased sweating and hypertension.

Hyposecretion of GH

Deficiency of GH in childhood leads to stunted growth or *dwarfism.*

Deficiency of GH in adulthood results in *mild anaemia,* which is refractory to usual treatment with haematinics like iron.

- Reduction in muscle mass and
- Hypoglycaemia may also occur.

Dwarfism. Short stature or dwarfism may be due to endocrinal or non-endocrinal causes.

Endocrinal causes of dwarfism are:
- Growth hormone deficiency (pituitary dwarf),
- Panhypopituitarism,
- Hypothyroid dwarf (see page 558) and
- Cushing's syndrome (see page 593).

Non-endocrinal causes of dwarfism include:
- Familial dwarfism,
- Achondroplasia,
- Nutritional (malnutrition or malabsorption),
- Chromosomal abnormalities, e.g. Turner's syndrome and
- Psychological dwarfism (Kasper–Hauser's syndrome).

Growth-hormone-related dwarfism

1. Pituitary dwarfism occurs due to deficiency of GH in early childhood.

Characteristic features. Deficiency of GH causes retardation of growth in all parts of the body proportionately. Consequently, a pituitary dwarf with a chronological age of 20 years has the body structure like that of a normal child of

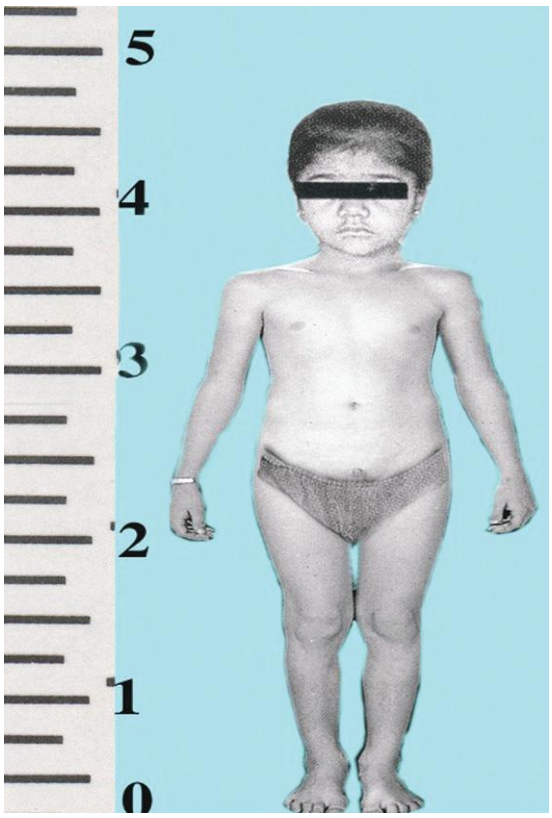


Fig. 8.2-9 Clinical features of a pituitary dwarf.

7–10 years of age. Thus a pituitary dwarf has following features (Fig. 8.2-9):

- Shortness of stature,
- Normal mental activity,
- Plumpness (fatness),
- Immature faces,
- Delicate extremities and
- Sexual maturity does not occur when associated with the gonadotropin deficiency.

2. ***African pygmies.*** In this condition, short stature is due to lack of GH receptors in the tissues; though both GH and somatomedin levels are normal. In addition, their plasma IGF-I levels fail to increase at the time of puberty.
3. ***Laron dwarfism.*** In this condition, there is congenital abnormality of the GH receptors, so it is also called *growth hormone insensitivity syndrome*. The plasma concentration of GH-binding protein decreases and IGF-I is not secreted in sufficient amount.

POSTERIOR PITUITARY HORMONES

The two important hormones released from posterior pituitary are:

- Antidiuretic hormone and
- Oxytocin.

ANTIDIURETIC HORMONE

Antidiuretic hormone, as the name indicates, prevents diuresis and is chiefly concerned with the conservation of body water. Since it also causes vasoconstriction, it is also called vasopressin or more precisely arginine vasopressin.

STRUCTURE, SYNTHESIS, STORAGE, RELEASE, TRANSPORT AND METABOLISM

Structure, synthesis, storage and release of ADH and oxytocin (OTC) being similar are discussed together.

Structure. ADH and OTC both are homologous neurohormones, polypeptide in nature, containing nine amino acids each, but the amino acids in position 3 and 8 differ in the two hormones.

Synthesis. ADH as well as OTC are synthesized in the cell bodies of magnocellular neurons of both paraventricular and supraoptic nuclei of the hypothalamus. However, supraoptic nucleus predominately contains ADH forming neurons while paraventricular nucleus contains mainly the OTC synthesizing neurons.

- Like all other peptides, they are formed on rough endoplasmic reticulum.
- The ADH and OTC are synthesized as large molecules, the pro-hormones known as *prepropressophysin* and *preprooxyphysin*, respectively.
- In prepropressophysin, the ADH molecules is associated with neurophysin II and in preprooxyphysin, the oxytocin molecule is associated with neurophysin-I.

Storage. The axons of ADH and OTC synthesizing neurons end in the posterior pituitary gland as terminal swelling. The secretory granules containing hormone precursors, known as *Herring bodies*, are transported down the axons by axoplasmic flow to the nerve endings in the posterior pituitary.

Secretion. Antidiuretic hormone and OTC are released when a nerve impulse is transmitted from the cell body in the hypothalamus down the axon, where it depolarizes the neurosecretory vesicles.

Transport. The hormone and other secreted products separately enter the closely adjacent capillary. The hormone then reaches the target cells and by circulatory interconnections to the anterior pituitary also.

Other sources of ADH and OTC. In addition to the above described magnocellular neurons of the hypothalamus, the ADH and OTC are also present in the:

- Endings of neurons which project on to brainstem and spinal cord,
- Gonads,

- Adrenal cortex and
- Thymus.

Biological half-life of ADH is 16–20 min after its release into the circulation.

Metabolism. The circulating vasopressin is rapidly inactivated in the liver and kidney.

VASOPRESSIN RECEPTORS

Three types of vasopressin receptors are recognized:

- *V_1–A receptors.* These are involved in the vasoconstrictor effect of ADH.
- *V_1–B receptors.* These are involved in the action of ADH on the anterior pituitary.
- *V_2 receptors.* These are involved in the action of ADH on kidney.

ACTIONS OF ADH

1. Action on kidney. The main role of ADH is regulation of water balance in the body by acting on the kidney, where it decreases the excretion of free water (i.e. antidiuretic and concentrating effect on kidney).

Site of action. The ADH responsive cells line the distal convoluted tubules and collecting ducts of the renal nephron. Antidiuretic hormone increases the permeability of these cells to water.

Mechanism of action. The ADH exerts its antidiuretic effect by binding to a specific plasma membrane receptor (known as V_2 receptor) on the capillary (basal) side of the cell, where it activates adenylyl cyclase. The increase in intracellular cAMP activates a protein (kinase) on the opposite (luminal apical) side of the cell. The activated protein kinase leads to rapid insertion of protein water channels (known as *aquaporins*) in the plasma membranes of the principal cells of the collecting ducts. Through these protein water channels (i.e. aquaporin) water rapidly moves from the tubular lumen into the collecting duct cells.

- Aquaporins are of different types: aquaporin-1, 2 and 3 are found in the kidneys; aquaporin-4 is found in the brain and aquaporin-5 is found in the salivary and lacrimal glands, and in respiratory tract.

2. Vasoconstrictor effect. Antidiuretic hormone in large doses cause vasoconstriction and leads to rise in blood pressure. Haemorrhage is a potent stimulus to ADH secretion.

3. Action on anterior pituitary. Antidiuretic hormone travels to the anterior pituitary via the portal veins and combines with the V_1–B receptors (also called V_3 receptors) and causes increased ACTH secretion from the corticotrophs.

4. Action on the liver. In the liver, ADH causes *glycogenolysis* by combining with the V_1–A receptor.

5. Action on the brain. V_1–A receptors are also found in brain, where ADH acts as a neurotransmitter and is involved in memory, regulation of temperature, regulation of blood pressure, circadian rhythms and brain development.

REGULATION OF ADH SECRETION

The main factors which regulate the ADH secretion are:

1. Effective osmotic pressure of plasma or plasma osmolality

Plasma osmolality in normal individuals is maintained very close to 285 mOsm/kg by ADH. In other words, change in plasma osmolality is a very potent regulator of ADH secretion. Thus, water deprivation (which increases plasma osmolality) stimulates ADH secretion. On the other hand, a water load (which decreases plasma osmolality) decreases ADH secretion.

Mechanism of action. Changes in plasma osmolality affect the osmoreceptors.

- *Osmoreceptors* refer to a group of neurons located in the anterior hypothalamus in the region of circumventricular organs and organum vasculosum of lamina terminalis. These cells are distinct from the cells which produce ADH.
- The rise in plasma osmolality even by 1–2% results in the shrinkage of osmoreceptors causing an increased rate of discharge and reflexly increased ADH secretion and thus maintained plasma osmolality (Fig. 8.2-10).

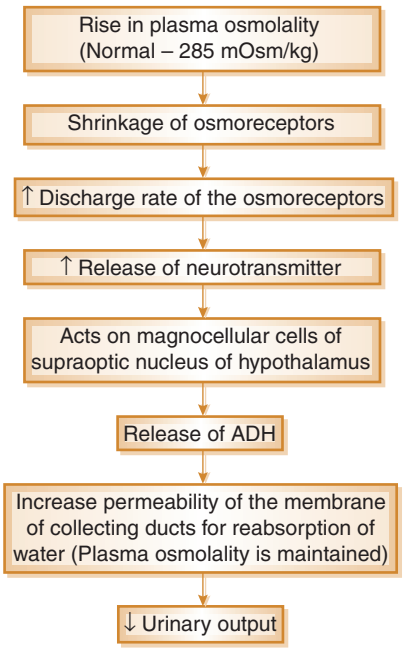


Fig. 8.2-10 Mechanism of regulation of ADH by plasma osmolality.

- The substances like Na^+, mannitol and sucrose are potent stimulators for ADH secretion. The hyperglycaemia and uraemia are comparatively less potent stimulators for ADH releases, but in uncontrolled diabetes mellitus (hyperglycaemia associated with insulin deficiency) glucose acts as an effective stimulus for ADH release.

2. Changes in blood volume

Changes in the circulating blood volume, central blood volume, cardiac output and blood pressure affect the secretion of ADH. Antidiuretic hormone secretion is increased when extracellular fluid (ECF) volume is low and decreased when ECF volume is high.

Mechanism of action. The changes in blood volume regulate ADH secretion by following mechanisms:

(i) Pressure receptor mechanism. Variations in the discharge from the pressure receptors in the circulatory system are inversely related to the rate of ADH secretion. There are two types of pressure receptors in the circulatory system: the low and high pressure receptors.

- *Low-pressure receptors* are located in the great veins, right and left atria and pulmonary vessels. They monitor the fullness of vascular system, and thus mainly respond to volume changes (hence also called volume receptors).
- *High-pressure receptors* are located in the carotid sinuses and aortic arch (i.e. high pressure portion of the vascular system) and reported to respond to pressure changes and so are also called baroreceptors.
- *Afferent impulses* from these receptors are carried by the ninth and tenth cranial nerves to their respective nuclei in the medulla. From the medulla, the impulses are carried by way of the mid brain via adrenergic neurotransmitters to the supraoptic nuclei of the hypothalamus. Normally, the pressure receptors tonically inhibit ADH secretion by modulating an inhibitory flow of adrenergic impulses from the medulla to the hypothalamus. A decrease in pressure increases ADH secretion by reducing the flow of neural impulses from the pressure receptors to the brainstem. The reduced neural input from these receptors relieves the source of tonic inhibition on the hypothalamic cells that secrete ADH.

(ii) Renin–angiotensin mechanism. Hypovolaemia also stimulates renin–angiotensin mechanism, which reinforces the release of ADH in response to hypovolaemia and hypotension by acting on the circumventricular organs.

(iii) Atrial natriuretic peptide (ANP) mechanism also reinforces the release of ADH in response to the hypovolaemia. When circulating volume is increased, ANP is released by the cardiac myocytes, which acts on the hypothalamus to inhibit ADH release.

3. Other factors affecting ADH secretion

Factors, other than the two major stimuli (i.e. hypovolaemia and plasma osmolality), which affect ADH secretion are:

- *Stress* of pain, chronic emotional stress and surgical procedures cause increase in ADH secretion leading to reduction in urine formation under these conditions.
- *Adrenaline* decreases the ADH. Hence one experiences an increased frequency of micturition during acute emotional stresses such as interviews, or examinations.
- *Alcohol* reduces ADH secretion and thus leads to diuresis.
- *Age.* Elderly individuals secrete more ADH than do younger individuals.
- *Cortisol and thyroid hormones* release ADH.
- *Some other factors* that increase ADH secretion include nausea, vomiting, standing posture and cytokines.

Summary of factors regulating ADH secretion is given in Fig. 8.2-11.

ABNORMALITIES OF ADH SECRETION

Abnormalities of ADH secretion include:

- Syndrome of inappropriate hypersecretion of ADH.
- Diabetes insipidus.

Syndrome of inappropriate hypersecretion of ADH

Syndrome of inappropriate hypersecretion of antidiuretic hormone (SIADH) refers to a condition in which ADH secretion is increased despite the presence of hypo-osmolality. Excessive ADH secretion leads to water intoxication, i.e. overhydration and because of this SIADH is also called *dilution syndrome.*

Causes. The causes of SIADH are:

I. *Excessive secretion of ADH from posterior pituitary* may occur during surgical stress because of pain and hypovolaemia.

II. *Excessive secretion of ADH from some ectopic source,* e.g.

- In cerebral disease (cerebral salt wasting),
- In pulmonary diseases (pulmonary salt wasting), such as bronchogenic carcinoma.

III. *Excess of ADH due to its decreased metabolic degradation* may occur in patients with liver cirrhosis and cardiac failure. In such cases half-life of ADH is prolonged.

Characteristic features

1. Water retention. Excessive ADH leads to water retention causing expansion of blood volume and ECF volume

2. Hypernatriuria. The expansion in ECF volume reduces aldosterone secretion causing hypernatriuria, i.e. excessive urinary excretion of Na^+.

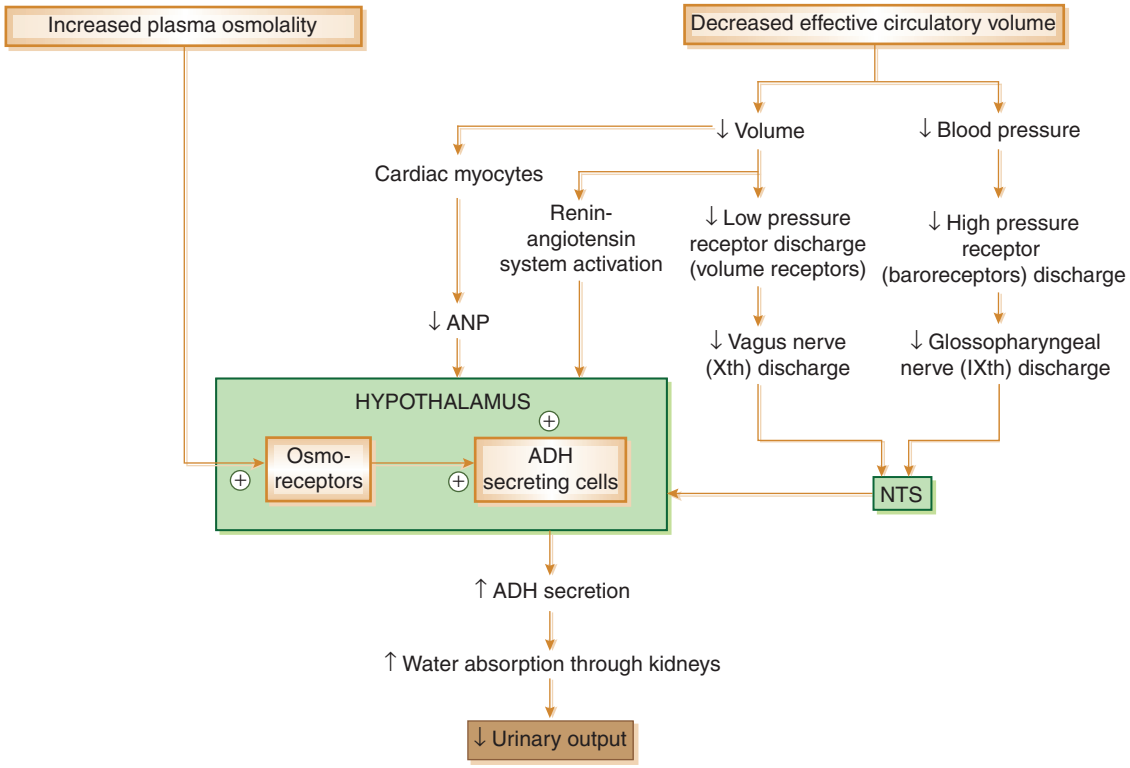


Fig. 8.2-11 Summary of factors regulating ADH secretion.

3. Hyponatraemia, i.e. decrease in plasma Na^+ level is caused by ADH by its dilutional effect (water retention) as well as by causing hypernatriuria.

4. Oedema. Water retention and excessive urinary excretion of Na^+ causes hypo-osmolality of blood (since ADH secretion is increased despite hypo-osmolality of blood that is why the condition is called syndrome of inappropriate hypersecretion of (SIADH). The low plasma osmolality causes shift of water from the plasma into the interstitial spaces producing the so-called oedema.

Treatment. SIADH can be treated with drugs like demeclocycline that blocks the effect of ADH on kidneys.

Diabetes insipidus

Diabetes insipidus refers to a clinical condition of polyuria that occurs either due to deficiency of ADH (vasopressin) release or failure of renal response to ADH.

Depending on the causes, diabetes insipidus is of two types:

A. Central (Cranial) or neurogenic diabetes insipidus. It occurs mainly due to failure of ADH secretion. The conditions include:

- Congenital
- Neoplasia of hypothalamus or Pituitary
- Surgery of pituitary or hypothalamus
- Vascular lesions

B. Nephrogenic diabetes insipidus. It occurs due to the failure of renal tubular response to ADH. The main cause is defect in vasopressin receptors (V_2 receptors) due to mutation of gene.

Characteristic features. The diabetes insipidus is characterized by decreased renal absorption of water leading to following features:

1. Polyurea, i.e. passage of large amount of urine, up to 3–20 L of low specific gravity is the most important single feature of diabetes insipidus.

2. Polydipsia. Polyuria is followed by obligatory polydipsia (drinking of large amount of water). It occurs due to the stimulation of thirst mechanism. In fact, polydipsia is an important mechanism which helps to maintain water balance with near normal plasma Na^+ level in patients with diabetes insipidus.

3. Dehydration may occur in severe cases. Its signs and symptoms include dry tongue, dry mouth, fall of blood pressure and loss of consciousness may be seen in acute severe cases.

Treatment. The treatment of diabetes insipidus is given below.

1. *Central or neurogenic diabetes insipidus* can be treated by:

- Hormonal therapy, i.e. ADH or desmopressin administration and
- Non-hormonal therapy, i.e. the drugs which increased ADH secretion, such as chlorpropamide (an oral hypoglycaemic agent), clofibrate and carbamazepine.

2. Nephrogenic diabetes insipidus is treated by the diuretics, e.g. hydrochlorothiazide. These inhibit Na^+ reabsorption in thick segment of loop of Henle and so urine osmolality does not fall below 300 mOsm/kg.

OXYTOCIN

STRUCTURE, SYNTHESIS, STORAGE AND RELEASE OF OXYTOCIN

Structure, synthesis, storage and release of oxytocin have been described along with ADH (see page 546).

ACTIONS OF OXYTOCIN

Mechanism of action

In human, oxytocin acts mainly on the uterus and breasts. Oxytocin acts through a G protein-coupled serpentine receptor and increases intracellular Ca^{2+} levels.

Action on breast. Oxytocin causes contraction of myoepithelial cells, thus plays an important role in milk ejection. For details see page 677.

Action on uterus. Oxytocin causes contraction of uterine smooth muscles, thus plays an important role during parturition (labour). For details see page 672. Oxytocin also acts on non-pregnant uterus and facilitates the transport of sperm in the female genital tract.

In males, the circulating levels of oxytocin increases during ejaculation, which causes increased contraction of smooth muscles of vas deferens and helps in propelling the sperms towards urethra.

CONTROL OF OXYTOCIN SECRETION

Stimuli which increase oxytocin release are:

- Oxytocin secretion increases on cholinergic stimulation
- Suckling stimulates oxytocin release (suckling reflex). For details see page 677.
- Genital tract simulation during coitus and labour increases oxytocin release (see page 672).

Factors which decrease oxytocin release are:

- Emotional stress
- Sympathetic stimulation, and
- Drugs such as ethanol and enkephaline

Chapter 8.3

Thyroid Gland

FUNCTIONAL ANATOMY

GROSS ANATOMY

Thyroid gland is the largest endocrine gland in the body (weighing about 15–25 g in adults). It consists of two lobes joined together by a narrow isthmus and is located on either side of the trachea just below the larynx. It receives high blood supply (400–600 mL/100 g/min).

HISTOLOGICAL STRUCTURE (FIG. 8.3-1)

Histologically, each *lobe* of thyroid gland is divided into various *lobules* by fibrous tissue septa. Each lobule is made up of an aggregation of several *follicles.* Each follicle is lined *by follicular cells.*

Follicular cells. These vary in shape with the degree of glandular activity. Normally (at an average level of activity), the cells are *cuboidal* and the colloid in the follicles is moderate in amount. During high degree of activity, the cells become columnar and flat when inactive (Fig. 8.3-1B). These cells secrete thyroid hormones.

Parafollicular cells or *C cells* are scattered between follicular cells and basement membrane (Fig. 8.3-1A), and secrete calcitonin, which is described in Chapter 8.4.

Colloid. This is a homogeneous material that fills the cavity of each follicle. When stimulated, the follicles are depleted of colloid; and when unstimulated, the follicles accumulate colloid. The major constituent of the colloid is thyroglobulin, a glycoprotein with a molecular weight of 660,000.

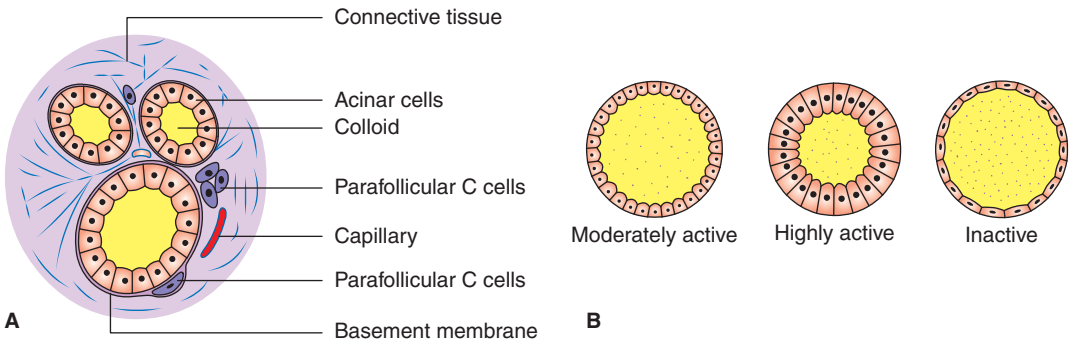


Fig. 8.3-1 Histological structure of thyroid gland (A) and variations in the follicular cell size with activity (B).

THYROID HORMONES

INTRODUCTION

The two principal thyroid hormones include thyroxine (T_4) and triiodothyronine (T_3).

- *Thyroxine* or T_4 (3,5,3′,5′-tetraiodothyronine) constitutes 90% of thyroid output.
- *Triiodothyronine* or T_3 (3,5,3′-triiodothyronine) constitutes 10% of thyroid output; however, it is responsible for most of the tissue actions of thyroid hormone.
- *Reverse triiodothyronine* (3,3′,5′-triiodothyronine) or reverse T_3 or rT_3 is a biologically inactive thyronine, which forms less than 1% of thyroid output.
- *Calcitonin* is a hormone secreted by the parafollicular cells of thyroid gland. It is concerned with calcium homeostasis and is discussed in Chapter 8.4.

Iodine metabolism

Dietary intake. Iodine is essential for the synthesis of thyroid hormones. It is ingested in the form of iodides. Sources of iodine are sea fish (richest), bread, milk and vegetables. Iodine is added to the table salt to prevent iodine deficiency.

- Daily average intake of iodine is 500 µg.
- Daily requirement of iodine is 100–200 µg.

Fate of dietary iodide. Most 80% (i.e. 400 µg/day) of the iodides absorbed from the gastrointestinal tract (GIT) are selectively removed from circulation by cells of the thyroid gland.

Plasma iodide level is 0.15–0.3 µg%.

Thyroid iodide. Thyroid gland contains 5–8 mg of iodide, i.e. about 95% of total iodine content of the body. Thus the thyroid serves as a store of iodine. Of the total thyroid iodide, only 5% is present within the cells of the follicular epithelium. The remaining 95% is present in the follicular lumen, stored in the colloid as thyroglobulin. In the colloid, two-thirds of the total iodine is present in the form of biologically inactive iodotyrosines and one-third is in the form of biologically active thyronine (T_4 and T_3).

BIOSYNTHESIS AND STORAGE OF THYROID HORMONES

Thyroxine (T_4) and triiodothyronine (T_3) are synthesized from tyrosine and iodide by the enzyme complex, peroxidase. The steps involved in the synthesis of thyroid hormones are (Fig. 8.3-2):

1. Iodine trapping. The first step in the synthesis of thyroid hormones is *uptake of iodide* by the thyroid gland, which occurs against the chemical (about 30:1) and electrical

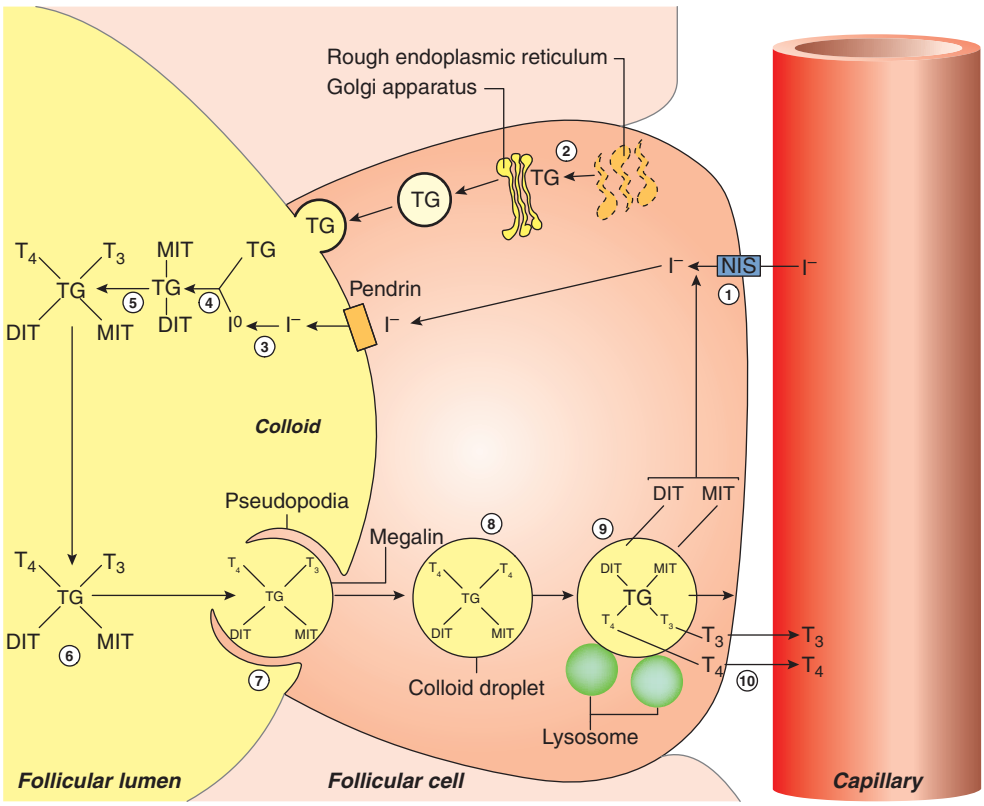


Fig. 8.3-2 Steps in synthesis and release of thyroid hormones: 1, iodine trapping; 2, synthesis of thyroglobulin; 3, oxidation of iodine; 4, organification of thyroglobulin; 5, coupling reaction; 6, storage of thyroid hormone in colloid; 7, take up of colloid by epithelial cell by endocytosis; 8, colloid vesicle; 9, release of T_3 and T_4 after proteolysis; and 10, diffusion of T_3 and T_4 into the capillary.

gradients. Therefore I^- is absorbed into the thyroid cell by iodide pump or Na/I^- symporter. It is an energy requiring process and is linked to the ATPase-dependent Na^+–K^+ pump.

- *Thyroid stimulating hormone (TSH) controls* the iodide uptake. The thyroid: plasma-free iodide ratio which is normally maintained to 30, may exceed 100 when the gland is maximally activated by TSH.
- *Antithyroid agents,* such as thiocyanate and perchlorate inhibit iodide transport.

2. Synthesis and secretion of thyroglobulin. Thyroglobulin is a large glycoprotein that is synthesized on the rough endoplasmic reticulum of the thyroid epithelial cells as peptide units of molecular weight 330,000. These units move to the apical plasma membrane and release into the lumen of follicle (Fig. 8.3-2).

- Each molecule of thyroglobulin contains about 140 tyrosine residues, which can serve as a substrate for iodine for the formation of thyroid hormones. It is also the storage site of the two hormones within the thyroid gland.

3. Oxidation of iodide. Once within the gland, iodide rapidly moves to the apical surface of the epithelial cells. From these, it is transported into the lumen of the follicles by a transporter called *pendrin.* The iodide (1) is then immediately oxidised to iodine (1°) by the enzyme *peroxidase* present near the apical border of the epithelial cells.

Note. Thyroid is the only tissue that can oxidise iodide to iodine. Thyroid-stimulating hormone promotes this reaction while antithyroid drugs (thiourea, thiouracil, methimazole) inhibit.

4. Organification of thyroglobulin refers to the iodination of tyrosine residues present in the thyroglobulin molecule. This reaction occurs at the apical membrane of the cell as soon as the thyroglobulin molecule is released by the secretory granules by exocytosis and requires thyroid peroxidase. Tyrosine (of thyroglobulin) is first iodinated at position 3 to form monoiodotyrosine (MIT) and then at position 5 to form diiodotyrosine (DIT).

5. Coupling reaction. Two molecules of DIT couple to form thyroxine (T_4). One molecule of MIT, when coupled with one molecule of DIT, triiodothyronine (T_3) is produced (Fig. 8.3-3). The enzyme peroxidase is required during coupling.

6. Storage. Once thyroglobulin has been iodinated it is stored in the lumen of the follicle as colloid for several months. It is estimated that the stored thyroid hormones can meet the body requirement for 1–3 months.

Note. The thyroid gland differs from other endocrine glands for storage of hormone. The secretion is stored in the cavity of follicle rather than in the secretory cells.

SECRETION, TRANSPORT AND METABOLISM OF THYROID HORMONES

Hormone secretion

Secretion of the thyroid hormone from the colloid stored in the lumen of follicle involves following steps (Fig. 8.3-2):

Endocytosis. The colloid containing iodinated thyroglobulin is retrieved from the lumen of the follicle by the epithelial cells through endocytosis. This process is facilitated by the TG receptor *megalin* located on the apical membrane. The colloid enters the cytoplasm in the form of *colloid droplets* by pinocytosis with the formation of reabsorption lacunae, which move through the cytoplasm toward the basal membrane (Fig. 8.3-2).

Proteolysis. The colloid droplets fuse with the lysosome vesicles containing proteolytic enzymes. The proteases digest the thyroglobulin molecule releasing T_4, T_3, DIT, MIT and other amino acid constituent into the cytoplasm of the epithelial cells.

- T_4 and T_3 diffuse through the basal border of epithelial cells into the blood stream via adjacent rich capillary plexus.
- Monoiodotyrosine and DIT are rapidly deiodinated within the follicular cells by the enzyme *deiodinase.* In this way, iodide is retrieved for recycling along with the tyrosine into T_4 and T_3 synthesis (Fig. 8.3-2).

Transport of T_4 and T_3

Secreted T_4 and T_3 circulate in the blood stream in two forms: bound and free.

1. Bound form. Most of the circulating T_4 (99.95%) and T_3 (99.5%) is bound to specific binding protein.

- *Thyroxine binding globulin* is the major binding protein which binds about 70% of T_4 and T_3.

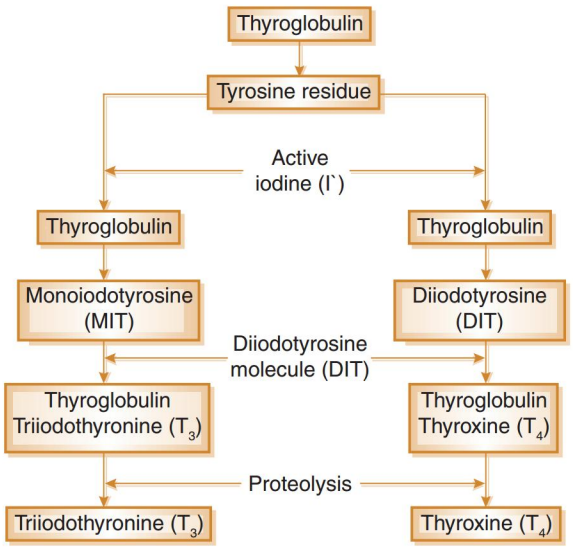


Fig. 8.3-3 Coupling reaction to form T_3 and T_4.

- *Thyroxine binding prealbumin* binds about 15–20% of T_4.
- *Thyroxine binding albumin*, binds about 10% of the T_4.

2. Free form. Only about 0.05% of T_4 and 0.5% T_3 circulate unbound (free form) in the plasma. These free, unbound hormones represent the biologically active hormone.

Note. It is important to note that most of the circulating T_3 is not of thyroid origin but diffuses from the tissues which convert T_4 into T_3 as described below.

Metabolism and excretion of thyroid hormones

The major pathways of peripheral metabolism of circulating thyroid hormone include deiodination, deamination (decarboxylation) and conjugation with glucuronic acid.

1. Deiodination. In the peripheral tissues, T_4 is deiodinated as:

- About 40% of T_4 is deiodinated into T_3 (3,5,3′-triiodothyronine) by the enzyme 5′-deiodinase.
- Remaining 60% of T_4 is deiodinated to reverse T_3, i.e. rT_3 (3,3′-5′-triiodothyronine) by 5-deiodinase. Reverse T_3 is physiologically inert.

Note. Since T_4 is deiodinated to T_3, which produces physiological effects, so T_4 itself may be metabiologically inert and hence called a *prohormone.*

T_3 and *rT_3* are deiodinated to DIT (T_2) and MIT (T_1). DIT (T_2) is further deiodinated to MIT (T_1).

2. Decarboxylation. Very small amount of T_4 and T_3 are metabolised by decarboxylation to form tetraiodothyroacetic acid (TETRAC) and triiodothyroacetic acid (TRIAC).

3. Conjugation. Approximately 15% of thyroid hormones (T_4 and T_3) are conjugated in the liver to form glucuronides and sulphates. The conjugate then is secreted via the bile duct into the intestine. In normal individuals, metabolites of T_4 and T_3 are excreted mainly in the faeces with a small amount appearing in the urine.

REGULATION OF THYROID HORMONE SECRETION

The secretion of thyroid hormones is regulated by:

- Negative feedback mechanism through hypothalamus–anterior pituitary–thyroid gland axis and
- Autoregulation of thyroid gland.

A. Regulation through negative feedback mechanism

The negative feedback mechanism operating through hypothalamus–anterior pituitary–thyroid gland axis (Fig. 8.3-4) plays the essential role in controlling secretion of thyroid hormones by:

Thyroid-stimulating hormone

Thyroid-stimulating hormone is a glycoprotein having molecular weight: 30,000. Average plasma level is 2.3 µIU/mL (range 0.2–0.5 µIU/mL):

Action of TSH. TSH exerts following effects on the thyroid gland:

1. *Increases the secretion of thyroid hormones* by accelerating all the steps in biosynthesis.
2. *Increases the number* (hyperplasia) and size (hypertrophy) *of the follicular epithelial cells.*
3. *Increases the vascularity of the thyroid gland.*

Regulation of TSH production. The production of TSH is regulated by the following.

1. Feedback control by plasma T_4 and T_3. Day-to-day secretion of TSH depends upon the negative feedback control exerted by the plasma levels of free T_4 and T_3 (Fig. 8.3-4):

- A fall in T_4 and T_3 levels stimulates TSH secretion from the anterior pituitary, while
- A rise in T_4 and T_3 levels inhibit TSH secretion.

Since there exists an established inverse relationship between the plasma levels of thyroid hormones and TSH, therefore measurement of the plasma TSH is a reliable test for assessing the status of thyroid gland.

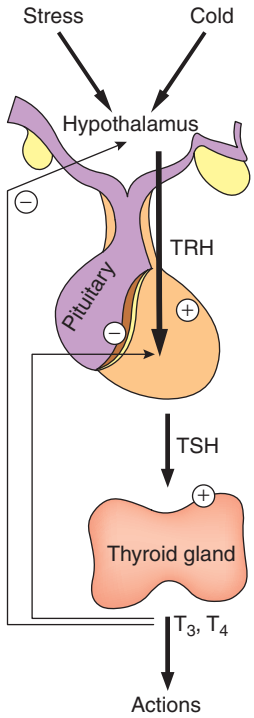


Fig. 8.3-4 Regulation of thyroid hormone secretion by negative feedback control mechanism through hypothalamus–anterior pituitary–thyroid gland axis.

2. Hypothalamic control of TSH. Hypothalamus adjusts TSH secretion under certain special circumstances such as exposure to cold, warmth, stress, anxiety, excitement etc. Hypothalamus exerts its effect by secreting thyrotropin-releasing hormone (TRH).

Thyrotropin-releasing hormone

Thyrotropin-releasing hormone is a tripeptide secreted by the arcuate nucleus of hypothalamus and stored in median eminence from where it is released into the hypothalamo-hypophyseal portal vessels to reach the anterior pituitary. Thyrotropin-releasing hormone acts on the basophils (thyrotrophs) in the anterior pituitary and controls the release of TSH.

Control of TRH. Secretion of TRH by the hypothalamus is controlled by:

- *Nervous stimuli* like emotion, stress, exposure to cold etc. and also by
- *Negative feedback control* exerted by plasma T_3 and T_4 levels on the hypothalamus (Fig. 8.3-4).

B. Autoregulation of thyroid gland

The secretions of thyroid gland are regulated by food iodine contents. If there is deficiency of iodine content in the diet then the *iodine trapping* mechanism of the follicular cells becomes super efficient and vice versa is also true, i.e. when there is excess of iodine content in the food then iodine trapping becomes less efficient and organification of excess amount of iodine does not occur. In this way, iodine availability for thyroxine synthesis remains constant and this phenomenon is called *autoregulation of thyroid gland.*

MECHANISM OF ACTIONS OF THYROID HORMONE

The thyroid hormones do not have any discrete target organ. They affect cellular activity of almost all the tissues of the body. T_3 acts by its effect on the gene expression on the target cell. Overall scheme of the thyroid hormone effects is described (Fig. 8.3-5).

For details of steps involved in gene expression see page 530.

This ultimately results in:

- *Increased synthesis of enzymes and specific structural or functional proteins.* This mechanism can explain the anabolic action and other metabolic action of thyroxine.
- *Increased synthesis of Na^+–K^+–ATPase.* It explains the calorigenic action of thyroxine. Increased metabolic rate has been attributed to increased energy consumption associated with increased Na^+ transport.
- *Increase in the number and activity of mitochondria* in the cells of the body. These increase the rate of ATP synthesis. Extremely high concentration of thyroid hormones causes uncoupling of oxidative phosphorylation process. As a result a large amount of heat is produced, but little ATP.

Resting oxygen use in humans ranges from about 150 mL/min in hypothyroid state to about 400 mL/min in the hyperthyroid state (normal 225–250 mL/min).

The magnitude of calorigenic action of thyroxine partly depends on the level of circulating catecholamines.

Increased metabolic rate is associated with increased utilization of many hormones, vitamins and certain drugs. Therefore, patients with hyperthyroidism require a larger vitamin intake.

ACTIONS OF THYROID HORMONE

1. Effects on growth and tissue development

Thyroid hormones are important for normal body growth and development.

(i) Role in normal body growth and skeletal maturation. Thyroid hormones exert their effect directly by increasing protein synthesis and enzymes; and indirectly by increasing production of growth hormone and somatomedins. Some important effects are on:

- Bone development,
- Teeth development,

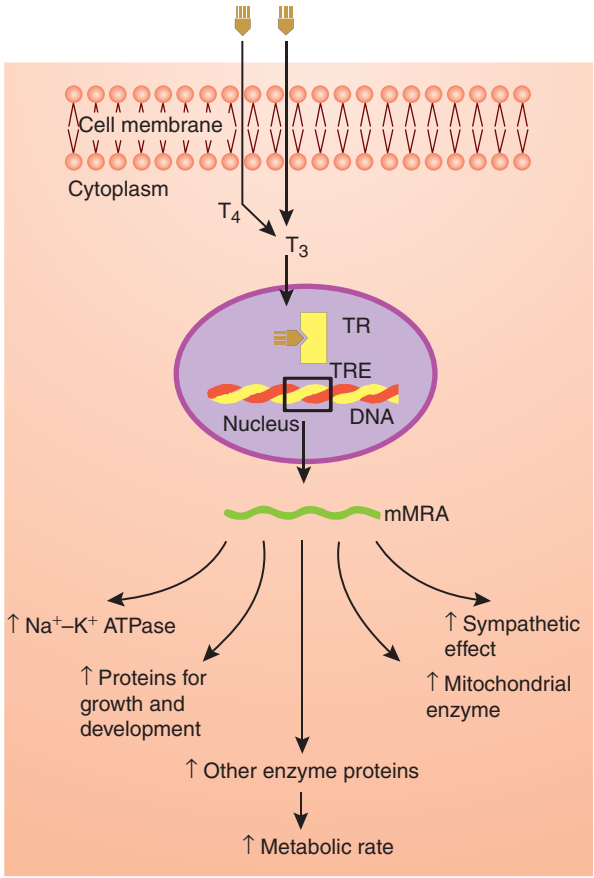


Fig. 8.3-5 Mechanism of intracellular actions of thyroid hormone.

- Normal cycle of growth and maturation and
- Subcutaneous tissues.

(ii) Role in tissue differentiation and maturation.

(iii) Role in development of nervous tissue. T_3 seems to be necessary for proper axonal and dendritic development as well as normal myelination in the nervous system. This is the reason of mental retardation being a striking feature in a child with congenital hypothyroidism. In such children, the disorder must be detected at the earliest and replacement hormonal therapy should be started, otherwise the mental retardation becomes irreversible.

2. Effect on the metabolic rate in general

The thyroid hormone in general stimulates the metabolic activities and increases the basal rate of oxygen consumption and heat production in most tissues of the body except the brain, retina, gonads, lungs and spleen.

3. Effects on metabolism

(i) Effect on carbohydrate metabolism. T_4 and T_3 lead on to an overall increase in enzymes causing:

- Increased glucose absorption from the GIT and
- Acceleration in almost all aspects of glucose metabolism, i.e. rapid uptake of glucose by the cells, enhanced glycolysis, enhanced gluconeogenesis and increased insulin secretion and its effects on the carbohydrate metabolism.

(ii) Effect on fat metabolism. Thyroid hormones cause:

- Mobilization of fat from the adipose tissue. Increase in the levels of fatty acids and enhanced oxidation of free fatty acids by cells.
- Decrease in the quantity of cholesterol, phospholipids and triglycerides in plasma, plasma cholesterol level is lowered due to increased excretion in bile.

Hypothyroidism is associated with elevated plasma cholesterol levels, which can be reversed by thyroid hormone administration.

(iii) Effect on protein metabolism. *In physiological amounts, the thyroid hormones* function as anabolic hormones. That is, they cause an increase in RNA and protein synthesis leading to positive nitrogen balance.

In high concentrations, thyroid hormones have catabolic effect leading to negative nitrogen balance. Therefore, muscle weakness and creatininuria are characteristic features of a hyperthyroid patient.

(iv) Metabolic effects through other hormones. T_4 and T_3 potentiate the respective stimulatory effects of epinephrine, norepinephrine, glucagon, cortisol and growth hormone on gluconeogenesis, lipolysis, ketogenesis and proteolysis of the labile protein pool.

(v) Effect on vitamin metabolism. Thyroid hormones increase the quantity of enzymes. Vitamins are the essential parts of some of the enzymes and coenzymes. Therefore, thyroid hormones cause an increased need for vitamins leading to relative vitamin deficiency in hyperthyroidism.

IMPORTANT NOTE

T_4 is essential for conversion of carotene to vitamin A. In hypothyroidism, this reaction is very slow and carotene accumulation in the blood and tissues (carotenaemia) gives a yellow colour to the skin. Carotenaemia can be clinically differentiated from jaundice by the fact that sclera of the eyeballs are not affected in the former condition.

(vi) Effect on water and electrolyte balance. Thyroid hormones play role in the regulation of water and electrolyte balance. This fact is clear from the observation that impairment of thyroid function is associated with retention of water and electrolytes, which can be reversed by hormonal administration.

4. Respiratory effects

Thyroid hormones stimulate O_2 utilization by following effects:

(i) Increase in the resting respiratory rate, minute ventilation and ventilatory responses to hypercapnia and hypoxia. These actions maintain a normal pO_2 when O_2 utilization is increased and a normal pCO_2 when CO_2 production is increased.

(ii) Increase in oxygen carrying capacity of blood by slightly increasing the red blood cell mass.

5. Cardiovascular effects

Thyroid hormone increases cardiac output, ensuring sufficient oxygen delivery in the tissues. In general, the thyroid hormones have the following effects on the cardiovascular system:

(i) Tachycardia, i.e. increased heart rate (at rest, even during sleep) is an important physical sign, which is used by clinicians in assessing the function of thyroid gland.

(ii) Force of cardiac contraction is increased by moderate increase in the thyroid hormone. The cardiac inotropic effects are via adrenergic stimulation. Myocardial calcium uptake and adenylyl cyclase activity are increased and enhance contractile force.

(iii) Cardiac output is increased as a result of increased blood volume, increased heart rate and increased force of contraction.

(iv) Effect on blood pressure. *Systolic blood pressure* is increased due to increased strength and rate of heart beat; whereas, *diastolic blood pressure* is decreased due to

peripheral vasodilatation. This results into an *increased pulse pressure,* but the mean arterial pressure is usually unchanged.

(v) Vasodilatation and increased blood flow to tissues occurs by two mechanisms:

- *Indirect mechanism*. Thyroid hormones cause rapid utilization of O_2 and increased production of heat and CO_2. These effects cause vasodilatation and an increase in blood flow in most of the tissues especially skin, muscle and heart. Cutaneous vasodilatation is particularly a prominent feature, which helps in dissipation of excessive heat produced.
- *Direct mechanism*. Thyroid hormones directly decrease systemic vascular resistance by dilating arterioles in the peripheral circulation.

6. Effects on nervous system

(a) Effect on development of nervous system

Thyroid hormones play an essential role in the development of nervous system. Critical period for the development of nervous system is up to 1 year of life.

(b) Effect on functioning of nervous tissue in adults

T_4 enhances wakefulness, alertness, responsiveness to various stimuli, auditory sense, awareness of hunger, memory and learning capacity. Normal emotional tone also depends on proper thyroid hormone availability.

Thyroid hormone increases the speed and amplitude of peripheral nerve reflexes.

Thyroid hormone versus sympathetic nervous system or catecholamine activity is compared in Table 8.3-1.

Table 8.3-1 Actions of catecholamine versus thyroxine (T_4)

Catecholamines	Thyroxine
1. The actions of catecholamines (epinephrine and norepinephrine) on basal metabolic rate (BMR), CNS activity and on heart (heart rate and force of contraction) are rapid and short lived.	1. The actions of thyroxine (T_4) are same on BMR, CNS and heart, but they are very slow and of long duration.
2. Catecholamines can't increase BMR in absence of thyroxine.	2. Thyroxine increases BMR. Presence of catecholamine potentiates this action of thyroxine.
3. Catecholamines stimulate reticular activating system.	3. Thyroxine also performs same function, but after sympathectomy, activity of reticular activating system get depressed.

7. Effects on gastrointestinal tract

Effects of thyroid hormones on GIT include:

- Increase in appetite and therefore increase in food intake
- Increase in rate of secretion of digestive juices
- Increase in motility of GIT. Excess of thyroid hormone often causes diarrhoea.

8. Effects on reproductive system

In both women and men, thyroid hormone plays an important permissive role in the regulation of reproductive functions.

In males, lack of thyroid hormones causes complete loss of libido and excess of hormones causes impotence.

In females, lack of thyroid has varying effects:

- Menorrhagia and polymenorrhagia,
- Irregular periods or even amenorrhoea occurs in some women.

9. Effects on other endocrine glands

Thyroid hormones also have significant effects on other parts of the endocrine system.

- *Pituitary* production of growth hormone is increased, whereas that of prolactin is decreased.
- *Adrenocortical* secretion of cortisol, as well as metabolic clearance of this hormone, is stimulated but plasma-free cortisol levels remain normal.
- *Oestrogens and androgens* ratio, in males, is increased. It accounts for occurrence of breast engorgement in males in hyperthyroidism.
- *Parathyroid hormone and 1,25-$(OH)_2$-vitamin D* are decreased as a compensatory consequence of the effects of thyroid hormone on bone resorption.

10. Effects on kidney

Renal plasma flow, glomerular filtration rate and tubular transport maximum for a number of substances are also increased by thyroid hormone.

APPLIED ASPECTS OF THYROID HORMONES

ABNORMALITIES OF THYROID GLAND

- Hyperthyroidism and
- Hypothyroidism.

HYPERTHYROIDISM

Hyperthyroidism refers to increased secretion of thyroid hormones. Its common causes are:

- *Graves' disease* (described below) and
- *Toxic nodular goitre.*

Graves' disease

Graves' disease or toxic goitre or thyrotoxicosis is the most common cause of hyperthyroidism (Fig. 8.3-6).

Aetiology

It is an autoimmune disease characterized by the development of thyroid-stimulating antibodies (TSAb) against the TSH receptors, also called long acting thyroid stimulator. These antibodies bind to TSH receptors and mimic TSH action on thyroid growth and hormone synthesis. The entire thyroid gland undergoes hyperplasia as a result of autoimmune stimulation.

Symptoms and signs

1. General features include:

- Marked increase in basal metabolic rate (BMR),
- Weight loss, despite an increased intake of food and
- Increased heat production causes discomfort in warm environments, excessive sweating and a greater intake of water.

2. Goitre. Goitre refers to the swelling of thyroid gland. Graves' disease is characterized by diffuse goitre, while single or more nodules indicate toxic nodular goitre.

3. Cardiovascular features are:

- Increased pulse rate or sinus tachycardia and
- Arrhythmias (atrial fibrillation is commonest).

4. Neuromuscular features are: nervousness, irritability, restlessness, psychosis, tremors of hand, muscular weakness and exaggerated tendon reflexes.

5. Gastrointestinal features are diarrhoea or steatorrhoea and vomiting.

6. Dermatological features are perspiration (increased sweating or hyperhidrosis), loss of hair and redness of palm.

7. Reproductive features are impotence in males and oligomenorrhoea or amenorrhoea, abortions and infertility in females.

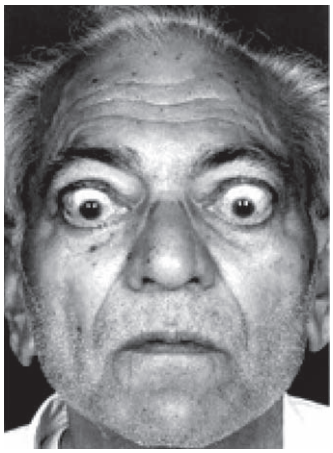


Fig. 8.3-6 Graves' disease.

8. Ophthalmological signs are lid retraction producing staring look and lid lag and exophthalmos, i.e. bulging out of eyeball.

Investigations

- Both T_3 and T_4 plasma levels are elevated.
- TSH is low or may become undetectable.
- ^{131}I Uptake is increased, i.e. >35% at 5 h.
- TRs antibodies may be increased >7 U/1, (N=<7U/l).
- Serum cholesterol is less.
- ECG shows tachycardia and arrhythmia.
- Ultrasonography of thyroid gland shows diffuse goitre.

HYPOTHYROIDISM

Hypothyroidism is a clinical syndrome caused by low levels of circulating thyroid hormones.

Aetiology

Depending upon the aetiology, hypothyroidism can be primary or secondary.

Primary hypothyroidism is caused by the disorder of thyroid gland.

Secondary hypothyroidism is caused by diseases of anterior pituitary and hypothalamus.

Clinical features

Clinical features depend upon the age at which deficiency manifests, duration and severity of the disease. Two different clinical entities are:

- Infantile hypothyroidism (cretinism).
- Adult hypothyroidism (myxoedema).

1. Infantile hypothyroidism (cretinism). It occurs when thyroid deficiency occurs during first year of life and is characterized by (Fig. 8.3-7) mental retardation, marked retardation of growth, delayed milestones of development, pot belly, protruding tongue, flat nose, dry skin and sparse hairs.

Radiograph of bone shows delayed bone age.
At adolescence, hypothyroidism is characterized by short stature, poor performance at school, delayed puberty and sexual maturation. Other features of adult hypothyroidism are present to variable degree.

Treatment should be prompt otherwise mental deficiency will persist.

2. Adult hypothyroidism is also called myxoedema because of characteristic infiltration of skin by myxoedematous tissue (Fig. 8.3-8). Symptoms and signs include:

- *General features:* Tiredness and weight gain without an appreciable increase in caloric intake (due to lower than normal metabolic rate). Decreased heat production,

lower body temperature, causes intolerance to cold and decreased sweating.

- *Cardiovascular features.* Adrenergic activity is decreased causing bradycardia.
- *Neuromuscular features.* Movement, speech and thought are all slowed and lethargy, sleepiness, delayed relaxation of ankle jerks, aches and pain are common. Pressure palsy of peripheral nerves (e.g. carpal tunnel syndrome) due to entrapment in excess ground substance.
- *Dermatological features.* Dry thick skin (toad skin), sparse hair, non-pitting oedema due to infiltration by myxoedematous tissue (myxoedema).
- *Reproductive features,* menorrhagia and infertility (common) galactorrhoea and impotence (less common).
- *Gastrointestinal features.* Constipation (common) and adynamic ileus (less common).
- *Haematological feature* includes anaemia.

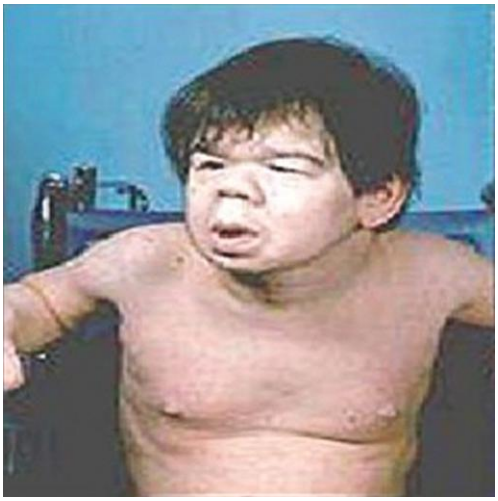


Fig. 8.3-7 Clinical features of cretinism. Note short stature, pot belly and idiotic look.

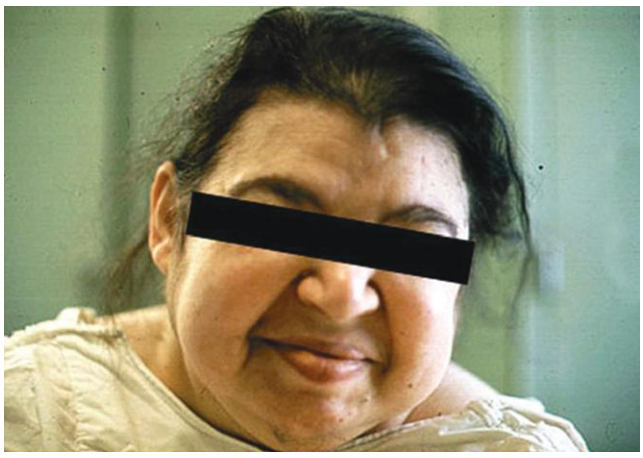


Fig. 8.3-8 Photograph of a patient with myxoedema showing puffy face, thick lips and periorbital oedema.

Investigations

- Serum T_3 and T_4 levels low
- Serum TSH levels high in primary and low in secondary hypothyroidism
- Serum cholesterol high
- Peripheral blood film macrocytic anaemia
- Photomotogram—delayed ankle jerk.

GOITRE

Goitre refers to any abnormal increase in the size of the thyroid gland. The term goitre does not denote the functional status of thyroid gland, because it may be associated with:

- *Euthyroid,* i.e. normal thyroid hormone level,
- *Hypothyroidism,* i.e. low thyroid hormone level, and
- *Hyperthyroidism,* i.e. high thyroid hormone levels; as seen in Graves' disease and toxic nodular goitre.

Goitrogenic substances (goitrogens). These are the substances that interfere with the production of thyroid hormone and cause thyroid enlargement, i.e. goitre. These include thiocyanates, nitrates and perchlorates and the drugs, such as thiourea, thiouracil, thiocarbamide, etc. Certain plant foods, such as cabbage, cauliflower, and turnip contain goitrogenic factors mostly thiocyanates.

If the goitrogen reduces thyroid hormone synthesis to subnormal levels, TSH secretion is increased chronically producing hypertrophy of thyroid gland (goitre).

Iodine-deficiency goitre or endemic goitre occurs when the daily dietary intake of iodine falls below 10 µg (normal requirement 100–200 µg/day). It decreases the synthesis and secretion of thyroid hormone leading to increased TSH levels and proliferation of thyroid gland tissue (goitre). It is mostly found in the geographic regions away from the sea coast where the water and soil are low in iodine content. Consumption of iodized salt is advocated to overcome the problem of endemic goitre. In certain cases, administration of thyroid hormone is also indicated.

THYROID FUNCTION TESTS

1. Measurement of basal metabolic rate. Theoretically, it is the physiological test of thyroid functions, since it measures the tissue response (O_2 consumption). However, because of poor sensitivity and specificity, BMR is now seldom used as a thyroid function test.

- Normal values of BMR: ±20%,
- In hyperthyroidism BMR may increase to 100% and
- In hypothyroidism it may decrease to –30 to –40%.

2. Radioactive iodine uptake (RAIU). Iodine uptake is an index for thyroid function that can be measured by using tracer dose of radioactive isotopes of iodine. Commonly

used traces are ^{123}I and ^{131}I. To perform this test, 25 curies of radioactive iodine (^{131}I) is given orally in 100 mL water and thyroid uptake is determined by placing an X-ray counter over the neck.

- Normal value of RAIU by thyroid (at 24 h) is 20–40%
- In hyperthyroidism, this value may be 60%
- In hypothyroidism, this value may be, 20%.

The analysis of radioactive iodine uptake is helpful in understanding physiology of the thyroid gland. The radioactive iodine uptake in a normal person is plotted in Fig. 8.3-9A.

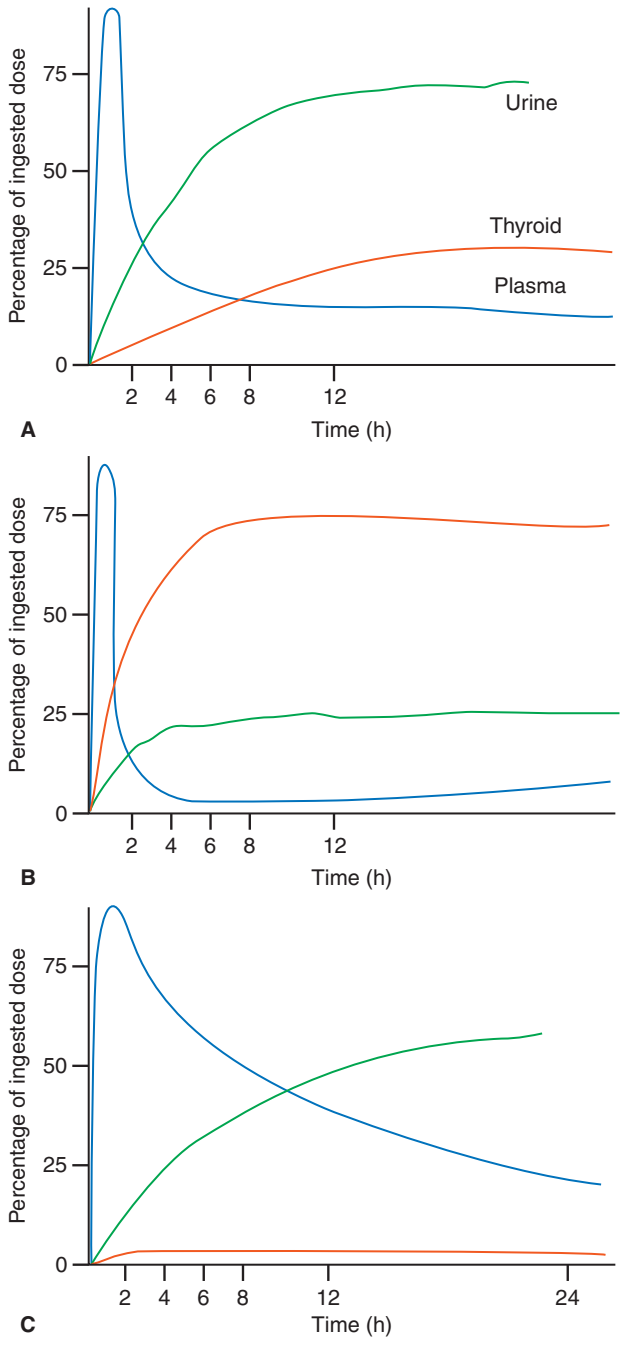


Fig 8.3-9 Distribution of radioactive iodine (^{123}I) uptake: A, in a normal (euthyroid); B, hyperthyroid and C, hypothyroid individuals.

- In hyperthyroidism, the amount of radioactivity in the thyroid gland rises sharply because iodide is rapidly incorporated into T_4 and T_3, and then start declining within 24 h (Fig. 8.3-9B).
- In hypothyroidism, the uptake is low (Fig. 8.3-9C).

APPLIED ASPECTS

A large amount of radioactive iodine destroys thyroid tissue, therefore, radioactive iodine therapy is useful in some cases of Graves' disease and thyroid carcinomas.

3. Measurement of total and free T_3 and T_4 and TSH levels in blood. These tests are considered best and are widely used for the diagnosis of various thyroid disorders. An accurate estimation of thyroid hormones can be done by radioimmunoassay or by ELISA method. These normal values and changes in hyperthyroidism and hypothyroidism are shown in Table 8.3-2.

- *TSH levels* are an important parameter of thyroid disorder, which tests the integrity of hypothalamic–pituitary–thyroid axis.

4. Ultrasonography of thyroid gland. Ultrasonography (B-scan) allows evaluation of an enlarged thyroid gland. It elucidates shape and dimension of nodules in thyroid gland.

5. Thyroid scan. A radionucleotide scan of thyroid either by ^{131}I/or ^{99}mTc is useful in demonstrating functioning thyroid tissue. It detects hot (functioning) and cold (nonfunctioning) nodule/nodules in the thyroid in cases with single or multinodular goitre.

6. Antithyroid antibodies. Detection of antithyroid antibodies is useful in diagnosing autoimmune thyroid disorder, such as Hashimoto's thyroiditis.

7. Fine-needle aspiration biopsy is carried out in patients with nodular goitre to detect any malignant process.

Table 8.3-2 Normal values of T_3, T_4, TSH and changes in hyperthyroidism and hypothyroidism

	Normal values	Hyperthyroidism	Hypothyroidism
• Total serum T_3	0.12 µg/dL	↑	↓
• Total serum T_4	8 µg/dL	↑	↓
• Free serum T_3	0.28 ng/dL	↑	↓
• Free serum T_4	2 ng/dL	↑	↓
• Serum TSH	0.2 to 5 µIU/mL	↓	↑

ANTITHYROID DRUGS

Secretions of thyroid gland can be reduced by antithyroid drugs. These drugs act by different ways, therefore, depending on the mechanism of action these drugs are of following types:

A. Drugs inhibiting iodide trapping by thyroid. These drugs act by two ways:

- *Competitive inhibition.* Monovalent ions, such as chlorate, perchlorate, thiocyanate compete with I^- for active transport into the thyroid gland.
- *Metabolic poisons* like dinitrate and cyanide

B. Drugs inhibiting oxidation of iodide and coupling. Thioureylenes (e.g. thiouracil and carbimazole) inhibit iodination of monoiodotyrosine and block coupling reaction to form T_3 and T_4.

C. Drugs inhibiting release of thyroxine (T_4 and T_3). Large dose of Iodide or Iodine decrease release of thyroid hormone, thus decreases serum concentration of T_4 and T_3. This effect is called *Wolff–Chaikoff effect.*

IMPORTANT NOTE

Hyperthyroid patients are more responsive to *iodide* as compared to normal individuals because the Wolff–Chaikoff effect is greater and more prolong due to increased iodine transport.

Chapter 8.4

Endocrinal Control of Calcium Metabolism and Bone Physiology

INTRODUCTION

The calcium, phosphorus and magnesium belong to a group of principal elements which constitute 60–80% of the body's inorganic material. The calcium and phosphorus form important structural components of bones and teeth, while calcium and magnesium are important determinants of neuromuscular excitability. Various hormones involved in the regulation of metabolism of these minerals include:

Calcitropic hormones, Parathyroid hormone (PTH), calcitonin and cholecalciferol (vitamin D_3) are primarily concerned with the regulation of calcium, phosph ate and magnesium metabolism in the body. These hormones act on three organ systems, bones, kidneys and intestinal tract to maintain calcium and phosphate levels.

Parathyroid hormone related protein (PTHrP) is the fourth local hormone that acts on the PTH receptor and is important for the skeletal development in utero.

Other hormones which also have some effect on calcium metabolism include glucocorticoids, growth hormone, oestrogens and various growth factors.

Discussion in this chapter is limited to regulatory role of PTH, calcitonin and cholecalciferol (vitamin D_3) only. An overview is presented of calcium and phosphate metabolism, as well as of the related structural and functional aspects of bone physiology.

CALCIUM, PHOSPHORUS AND MAGNESIUM METABOLISM

CALCIUM METABOLISM

PHYSIOLOGICAL AND BIOCHEMICAL FUNCTIONS

Calcium ions regulate a number of important physiologic and biochemical processes. To ensure that these processes operate normally, the plasma concentration is maintained

within a very narrow limit (9–11 mg/dL). Free, ionized calcium is the biologically active form of calcium. Important physiological and biochemical functions subserved by calcium are:

1. Development of bone and teeth. Calcium along with phosphorus is essential for the formation (of hydroxyapatite) and physical strength of the skeletal tissue. Bones which are in dynamic state serve as a reservoir of calcium.

2. Neuromuscular excitation. Calcium is essential for the transmission of nerve impulse. It interacts with troponin C to trigger muscle contraction. Calcium also activates ATPases, which increases the interaction between actin and myosin.

3. Blood coagulation. Calcium (factor IV) is involved in several reactions in the cascade of blood clotting mechanism.

4. Membrane integrity and plasma membrane transport. Permeability and transport of water and several ions across the cell membrane are influenced by calcium.

5. Mediation of intracellular action of hormones. Calcium mediates the intracellular actions of certain hormones by acting as a second messenger (e.g. epinephrine in liver glycogenolysis) and third messenger (e.g. antidiuretic hormone acts through cAMP, and then Ca^{2+}).

6. Activation of enzymes. Calcium is needed for direct activation of enzymes, such as lipase (pancreatic), ATPase and succinate dehydrogenase.

7. Release of hormones and neurotransmitters. Calcium facilitates the release of certain hormones and neurotransmitters, e.g. insulin, PTH and calcitonin.

8. Calmodulin-mediated action of calcium. Calmodulin is a calcium-binding protein. Calcium–calmodulin complex activates certain enzymes, e.g. adenylyl cyclase and calcium-dependent protein kinases.

9. Regulation of secretory processes. The microfilament and microtubule-mediated processes, such as endocytosis, exocytosis and cell motility are regulated by calcium.

10. Contact inhibition. Calcium is believed to be involved in the cell-to-cell contact and adhesion of cells in tissues. It may also be required for cell-to-cell communication.

11. Action on heart. By acting on myocardium, the calcium prolongs the systole.

CALCIUM DISTRIBUTION IN THE BODY

Calcium is the most abundant among the minerals in the body. The total content of calcium in an adult man is about 1100 g (27.5 mol). As much as 99% of it is present in the bones and teeth as hydroxyapatite. A small fraction (1%) of the calcium found outside the skeletal tissues performs a wide variety of functions.

Calcium in bones

The calcium in bones is present in two pools:

1. Pool of stable calcium is much larger (99% of total bone calcium) and is formed by the calcium present in the stable mature bones. It represents the calcium pool that is not readily exchangeable, but can be mobilized only through the action of PTH.

2. Pool (reservoir) of readily exchangeable calcium is much smaller (only 1% of the total bony content) and consists of labile (young) newly formed bone.

Calcium in plasma

Normal values of different forms of plasma calcium are:

Total plasma calcium	:	10 mg/dL (2.5 mmol/L) (Range 9–11 mg/dL)
• Diffusible calcium	:	6 mg/dL (1.5 mmol/L)
– Ionized calcium	:	5 mg/dL (1.25 mmol/L) (50% of total plasma calcium)
– Complexed to HCO_3^-, citrate etc.	:	1 mg/dL (0.25 mmol/L) (10% of total plasma calcium)
• Non-diffusible calcium	:	4 mg/dL (1 mmol/L)
– Bound to albumin	:	(i.e. 40% of total plasma calcium)

CALCIUM BALANCE

The calcium ion is fundamentally important to all the biological systems. Therefore, the concentration of calcium must be maintained within specific limits. The overall calcium homeostasis (calcium balance) or the normal daily calcium turnover is maintained by interplay of following processes (Fig. 8.4-1):

- Absorption of ingested calcium,
- Exchange of calcium between bone and extracellular fluid (ECF), and
- Excretion of calcium in the faecal matter and urine.

Absorption of calcium

Normally, at a daily intake of 1000 mg of calcium, about 35% (i.e. 350 mg) is absorbed (Fig. 8.4-1). The absorption of calcium mainly occurs in the duodenum and is regulated by 1,25-dihydroxycholecalciferol (For details see page 574).

Exchange of calcium between bone and extracellular fluid

The ECF contains about 1000 mg of calcium, which is in dynamic equilibrium with the calcium present in the bones.

Two types of exchange occur between the bone and ECF: rapid exchange and slow exchange.

Rapid exchange occurs between the ECF and the smaller (1% of the total bony content) readily exchangeable pool of

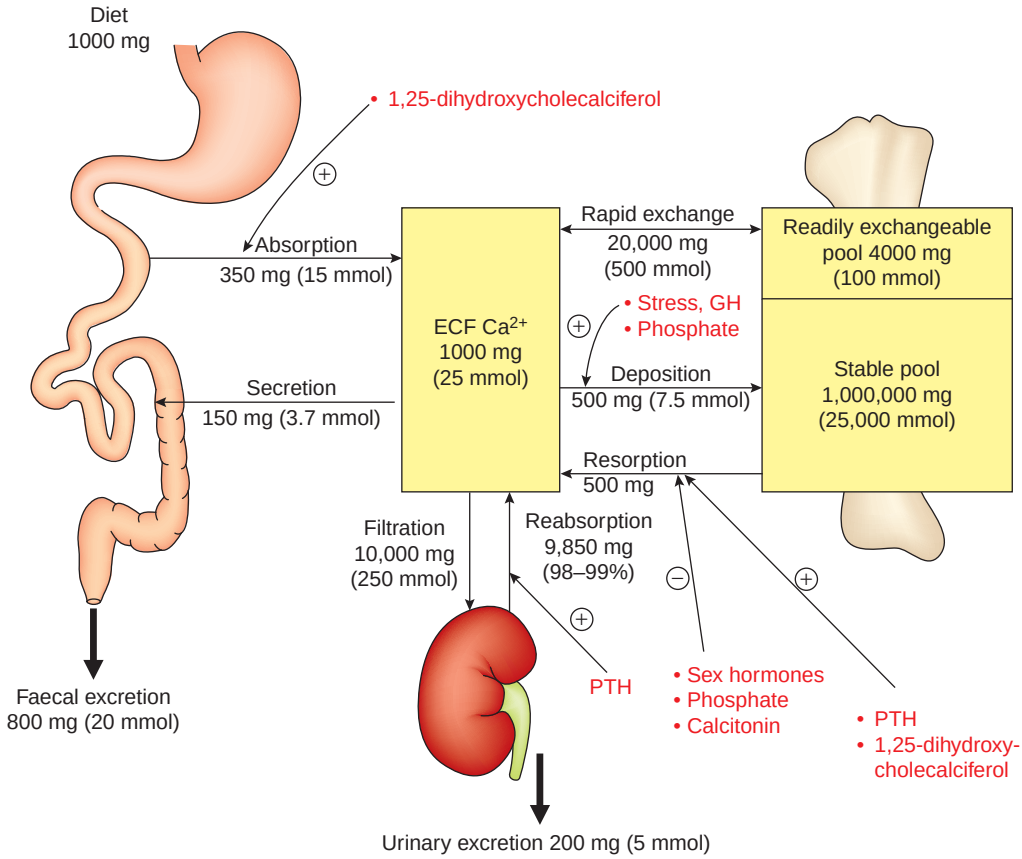


Fig. 8.4-1 Hormonal maintenance of calcium balance in an adult human ingesting 1000 mg (25 mmol) of calcium per day.

bone calcium. A large amount of calcium (about 20,000 mg) per day moves into and out of the readily exchangeable pool in the bone.

Slow exchange occurs between the ECF and larger (99% of total bone content) pool of stable calcium. This exchange is the one concerned with bone remodelling by constant interplay of bone resorption and deposition (about 500 mg/day only).

Excretion of calcium

The same amount of calcium as absorbed from the gut, i.e. about 350 mg, must ultimately be excreted to maintain balance. Excretion of calcium occurs in the faecal matter as well as in the urine (Fig. 8.4-1).

Faecal excretion of calcium. About 150 mg calcium is secreted into the intestine through bile, pancreatic juice and intestinal secretions and excreted in the stools along with the unabsorbed fraction (650 mg) from the diet. In this way, about 800 mg of calcium is excreted in the faecal matter (Fig. 8.4-1).

Urinary excretion of calcium. A large amount (about 10,000 mg) of calcium is filtered in the kidneys/day, but 98–99% of the filtered calcium is reabsorbed. Thus, in a normal healthy adult with calcium intake of 1000 mg, about 200 mg is excreted in the urine (Fig. 8.4-1). Adjustment of this small fraction of filtered calcium that is finally excreted provides a sensitive means of maintaining calcium balance.

Types of calcium balance

Three types of calcium balance exist:

1. Neutral calcium balance. It is seen in normal healthy individuals in which excretion of calcium in the urine and faeces exactly matches (equals) the daily intake of calcium (Fig. 8.4-1). There also exists an internal balance between the entry into and exit from the bone.

2. Positive calcium balance. It is seen in growing children, where the intestinal calcium absorption exceeds total excretion of calcium. The excess calcium is deposited in the growing bones, i.e. entry of calcium into bone is more than the exit.

3. Negative calcium balance. It is seen in women during pregnancy and lactation. Intestinal calcium absorption is less than the calcium excretion. The deficit comes from the maternal bones, i.e. exit of calcium out of the bone is more than the entry into the bone.

HORMONAL REGULATION OF PLASMA CALCIUM LEVEL

As mentioned earlier, maintenance of plasma calcium level within narrow range (9–11 mg/dL) is essential as it is involved in a number of important physiological and biochemical processes. Deviations of the ionized calcium from the normal range cause many disorders and can be life-threatening. The hormones regulating plasma calcium levels include:

A. Calcitropic hormones. The three primarily involved in the calcium homeostasis are:

- Parathyroid hormone (PTH),
- Active form of vitamin D (1,25-dihydroxycholecalciferol) and
- Calcitonin,

Main role of these hormones in calcium metabolism is summarized in Table 8.4-1 and Fig. 8.4-1.

B. Parathyroid hormone-related protein (PTHrP). It is a local hormone that acts on the PTH receptors and is important for the skeletal development in utero.

C. Other hormones, which have some effect on calcium metabolism include:

- *Growth hormone* has stimulatory effect on bone deposition.
- *Sex hormones* have inhibitory effect on bone resorption.
- *Glucocorticoids* have stimulatory effect on bone resorption.
- *Growth factors* have stimulatory effect on bone deposition.

Hypocalcaemia is compensated by *(Fig. 8.4-2)*

- Release of calcium from the bones,
- Increased fractional reabsorption in the kidney and
- Increased absorption from the intestine.

PHOSPHORUS METABOLISM

PHYSIOLOGICAL AND BIOCHEMICAL FUNCTIONS

The phosphate ion is also critically important to all biological systems. Important functions subserved by phosphate are:

1. Development of bone and teeth
2. Structural part of:
   - *High energy* transfer and storage compounds, such as ATP, GTP and creatine phosphate.
   - *Co-factors,* such as NAD, NADP and thiamine pyrophosphate.
   - *Second messengers,* e.g. cAMP and inositol triphosphate.
   - *Nucleic acids* (DNA, RNA), phospholipids and phosphoproteins.
3. *Activation of enzymes* by phosphorylation.
4. *Role in carbohydrate* metabolism.
5. *Phosphate buffer system* is important for the maintenance of pH in the blood as well as in the cells.

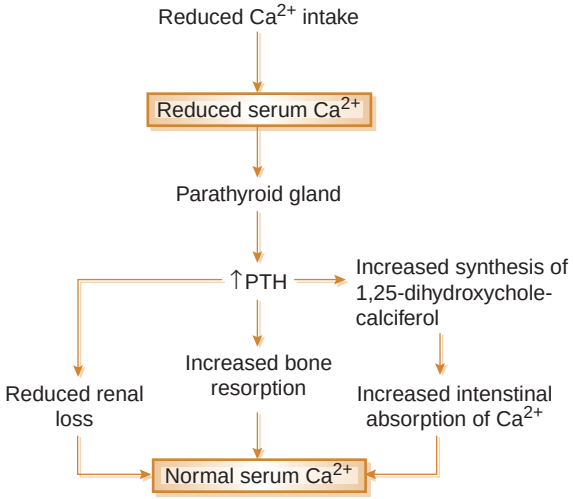


Fig. 8.4-2 Mechanism of regulation of serum calcium levels in hypocalcaemia.

Table 8.4-1 Summary of calcitropic hormones that regulate calcium balance

	PTH	1,25-dihydroxycholecalciferol	Calcitonin
Stimulus for secretion	↓ Serum Ca^{2+}	↓ Serum Ca^{2+} ↑ PTH ↓ Serum phosphate	↑ Serum Ca^{2+}
Actions on:			
• Bone	↑ Resorption	↑ Resorption	↓ Resorption
• Kidney	↓ P Reabsorption	↑ P Reabsorption	–
	↑ Ca^{2+} Reabsorption	↑ Ca^{2+} Reabsorption	–
• Intestine	↑ Ca^{2+} Absorption	↑ Ca^{2+} Absorption	–
Overall effect on:			
• Serum Ca^{2+}	↑	↑	↓
• Serum phosphate	↓	↑	–

↑ = Increase; ↓ = decrease; – = no effect.

6. *Important intracellular anion* that balances the certain cations (K^+ and Mg^{2+}) inside the cells.

DISTRIBUTION OF PHOSPHATE IN THE BODY

An adult body contains about 1 kg phosphate (P) which is distributed as:

- Bones and teeth : 80% (in combination with Ca^{2+}).
- Muscles and blood : 10% (in association with proteins, carbohydrates and lipids).
- Chemical compounds : 10% widely distributed in body.

Blood phosphate

- Plasma levels : 3–5 mg/dL
- Plasma phosphate exists in three forms:
  - Protein bound : 10%
  - Free ions : 40%
  - Complexed with cations (Ca^{2+}, Mg^{2+}, Na^+, K^+) : 50%

} 90% is filterable

PHOSPHORUS BALANCE (FIG. 8.4-3)

1. Intake and absorption. Recommended intake is about 800 mg/day. The recommended ratio of Ca^{2+}:P in adults is 1:1 and in infants 2:1. Sources of P are milk, cereals, leafy vegetables, meat and eggs. Ca^{2+} and P are distributed in the majority of natural foods.

On an average dietary intake of P in adults is about 1000 mg/day. Phosphorus is absorbed actively and maximally in duodenum. About 70–80% of P, as compared to 30–40% of Ca^{2+} is absorbed from the gut.

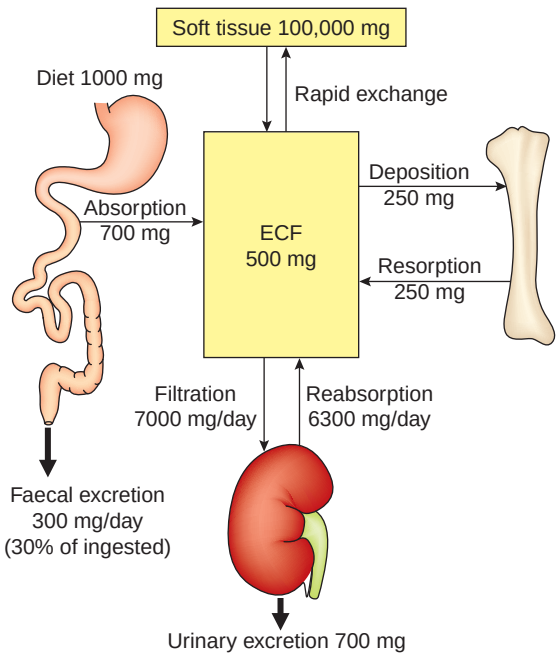


Fig. 8.4-3 Maintenance of phosphorus balance in an adult human ingesting 1000 mg of phosphorus.

Factors affecting phosphorus absorption.

- Vitamin D, PTH and growth hormone (GH) promote absorption.
- Cortisol and heavy metal ions inhibit absorption.

2. Exchange of phosphate between extracellular fluid and soft tissues. The soft tissue stores of the phosphate, such as those in the muscle mass, undergo rapid exchange with the ECF pool of phosphate (Fig. 8.4-3). This process plays an important role in the minute-to-minute regulation of the plasma phosphate concentration.

3. Exchange of phosphate between extracellular fluid and bone. About 250 mg of phosphate enters and leaves the bone from 500 mg of ECF pool in the process of bone remodelling.

Excretion of phosphate occurs in the faecal matter and urine.

(i) Faecal excretion includes 300 mg (30% of ingested) of phosphate which is not absorbed (Fig. 8.4-3).

(ii) Urinary excretion. About 7000 mg of phosphate is filtered by kidney per day. A larger fraction (90%) of the filtered phosphate is reabsorbed.

REGULATION OF SERUM PHOSPHATE LEVELS

Hypophosphataemia and hypocalcaemia due to dietary or other causes bring about different adaptative changes to normalize the plasma levels. Responses to hypocalcaemia are more immediate than to hypophosphataemia.

The hypophosphataemia is mainly compensated by reduced urinary loss and there occurs no change in dietary absorption.

MAGNESIUM METABOLISM

The divalent cation, magnesium (Mg^{2+}), is related in some respects to calcium and phosphates.

Functions subserved by magnesium are:

- *Role in formation of bone and teeth.*
- *Serves as a co-factor for several enzymes* requiring ATP, e.g. hexokinase, glucokinase, phosphofructokinase, adenylyl cyclase.
- *Required for proper neuromuscular function.* Low levels of Mg^{2+} lead to neuromuscular irritability.
- *Required for release of PTH* in response to hypocalcaemia and also for the actions of the hormone on its various target tissues.

Distribution of Mg^{2+} in the body. The body contains a total of 25 g of Mg^{2+}, which is distributed as:

- 10% in bones, in combination with calcium and phosphate,
- 50% in soft tissues and body fluids.

Plasma levels range from 1.8 to 2.4 mg/dL.

Magnesium balance. Daily requirement of magnesium is 300–500 mg. Leafy vegetables, nuts and soyabean are rich sources of magnesium. Magnesium is mainly absorbed in the distal part of the small intestine. On an average 40% (i.e. 120–200 mg) of intake is absorbed daily. Consumption of large amounts of calcium, phosphate and alcohol diminish Mg^{2+} absorption. Parathyroid hormone increases Mg^{2+} absorption. In a steady state, the same amount is excreted in the urine. Magnesium deficiency is compensated by decreased urinary excretion.

BONE PHYSIOLOGY

FUNCTIONS AND COMPOSITION OF BONE

FUNCTIONS OF BONE

Bone is a specialized tough connective tissue that forms the skeleton of the body. It subserves following functions:

1. Protective function. The framework formed by the bones protects the vital organs and soft tissues of the body, e.g. thoracic cage protects lungs and heart, and skull protects brain.

2. Mechanical functions served by the bones include:

- *Support* to body,
- *Attachment* to muscles and tendons.

3. Movements are performed at the joints by leverage effects of bones.

4. Metabolic functions of bone include their important role in homeostasis of calcium and phosphate metabolism.

5. Haemopoietic function includes the formation of blood cells in the red bone marrow.

COMPOSITION OF BONE

Bone, a special form of connective tissue, is composed of a collagenous framework (matrix) impregnated with the bone salts. The dry, fat-free bone consists of one-third organic bone matrix and two-thirds minerals (inorganic).

Bone matrix

Bone matrix, also called osteoid, consists of collagen fibres embedded in the gelatinous ground substance.

Collagen fibres are arranged in lamellae. The fibres of one lamella run parallel to each other, but those of adjoining lamellae run at varying angles to each other. Over 90% of the organic matrix is type I collagen. Collagen protein is rich in *glycine, proline* and *hydroxyproline*.

Ground substance of a lamella is continuous with that of adjoining lamellae. It is formed by the ECF and proteoglycans (which include chondroitin sulphate and hyaluronic acid). These substances are concerned with the regulation and deposition of bone salts.

Bone salts

The bone salts constitute the inorganic component of bone which is comprised primarily of calcium and phosphate in the form of *hydroxyapatite* crystals [$Ca_{10}(PO_4)_6(OH)_2$]. Each crystal measures about 400 A° units in length, 100 A° units in breadth and 10–30 A° units in thickness. Adsorbed on the surface of hydroxyapatite crystals are present small amounts of other salts such as sodium, potassium, magnesium and carbonate. The bone salts strengthen the bone matrix.

STRUCTURAL CONSIDERATIONS

STRUCTURE OF BONE

Structurally, two types of bones are known: compact or cortical bone, and trabecular or spongy or cancellous bone. In most of the bones, both compact and cancellous forms are present, but thickness of each type varies in different regions of the bone. For example, in long bones, the *epiphyseal region* contains a large amount of cancellous bone and outer thin compact bone. While in *diaphyseal regions*, the amount of compact bone is more and cancellous (spongy) bone is very thin (Fig. 8.4-4).

Structure of compact bone

The compact bone makes the outer layer of most bones and accounts for the 80% of the bone in the body. Histologically, the compact bony tissue is made up of several minute

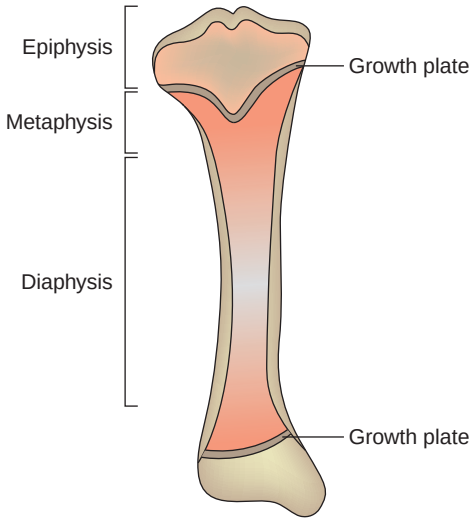


Fig. 8.4-4 Parts and gross structure of long bones as seen in longitudinal cut section.

cylindrical structures called *osteons* or *Haversian system* (Fig. 8.4-5). Each osteon is formed by several layers of collagen lamellae (Haversian lamellae) arranged concentrically around a centrally placed canal called the *Haversian canal,* which contains the blood vessels, lymph vessels and nerve fibres. In between the concentric layers of collagen tissue are present many *lacunae* (small cavities), which contain *osteocytes*. The osteocytes send long process called canaliculi all around. The canaliculi from the neighbouring osteocytes unite to form tight junctions.

The Haversian canals (and therefore the osteons) run along the longitudinal axis of long bones and branch and anastomose with each other. They also communicate with the external surface of the bone through channels that are called *canals of Volkmann*. Blood vessels and nerves pass through all these channels, so that compact bone permeated by a network of blood vessels that provide nutrition to it. The compact bone is lined externally by the periosteum and internally by the endosteum. Both periosteum and endosteum of the long bones contain osteoprogenitor cells, which can differentiate into osteoblasts or osteoclasts.

Structure of trabecular or spongy bone

The trabecular or spongy or cancellous bone is present inside the compact bone and makes up the 20% of bone in the body. It is made up of spicules or plates or trabeculae which are separated by wide spaces that are filled in by bone marrow. Nutrients diffuse from the bone ECF to trabeculae.

The trabeculae are thin and consist of irregular lamellae of bone with lacunae containing osteocytes. The trabeculae are covered by a thin layer of connective tissue called *endosteum,* which contains osteoblasts, osteoclasts and osteoprogenitor (stem) cells (Fig. 8.4-6).

CELLS OF BONE

Osteoprogenitor cells

These are stem cells of mesenchymal origin that can proliferate and convert themselves into osteoblasts whenever there is need for bone formation.

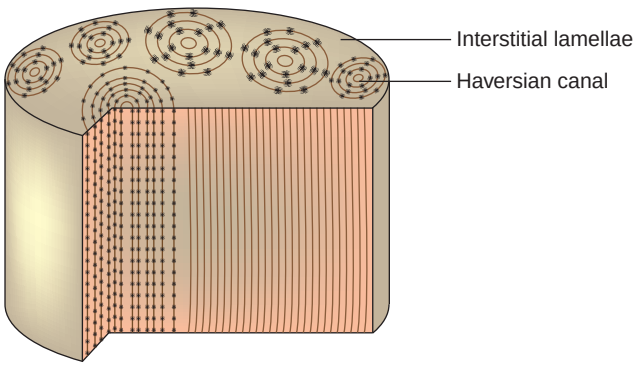


Fig. 8.4-5 Structure of compact bone.

In the fetus, the osteoprogenitor cells are numerous at sites where bone formation is to take place.

In the adults, these cells are present over the periosteum as well as the endosteum.

Osteoblasts

Bone forming cells are called osteoblasts. These are derived from the osteoprogenitor cells. Being concerned with bone formation they are situated in the outer surface of bone (Fig. 8.4-7), the marrow cavity and epiphyseal plate cells.

Functions of osteoblast cells include:

1. Role in laying down of the organic matrix of bone. Osteoblasts are responsible for the synthesis of bone matrix by secreting type I collagen and a protein called *matrix gla protein* and other proteins involved in the matrix formation.

2. Role in calcification. Enzyme alkaline phosphatase present in the cell membranes of osteoblasts plays an important role in the calcification of bone matrix. Osteoblasts are believed to shed off matrix vesicles which possibly serve as points around which formation of hydroxyapatite crystals takes place.

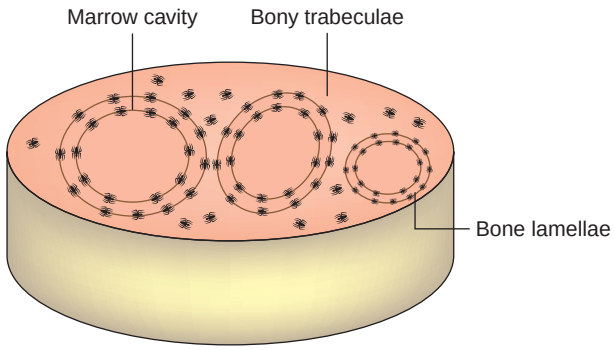


Fig. 8.4-6 Structure of trabecular bone.

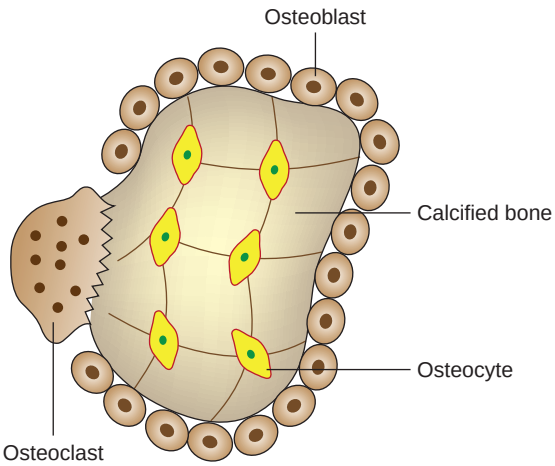


Fig. 8.4-7 Location of various bone cells.

3. Role in bone resorption. Osteoblasts may indirectly influence the resorption of bone by inhibiting or stimulating the activity of osteoclasts.

Fate of osteoblasts. After taking part into bone formation, the osteoblasts are converted into osteocytes, which are trapped inside the lacunae of calcified bone.

Osteocytes

Cells of mature (or developed) bone are called osteocytes. Osteoblasts, which during bone formation are 'imprisoned' in the lacunae between the bone lamellae.

These play an important role in maintaining the exchange of calcium between the bone and the ECF.

- Metabolic activity of osteocytes helps to maintain the bone as living tissue.
- Maintain the integrity of lacunae and canaliculi, and thus keep open the channels for diffusion of nutrients through bone.

Osteoclasts

Bone removing cells are called osteoclasts. These are giant multinucleated cells found in relation to surfaces where bone removal is taking place.

Osteoclasts are derived from the haemopoietic stem cells via monocytes. Probably, they are formed by fusion of many monocytes.

Function. Osteoclasts are responsible for bone resorption during bone remodelling. The lysosomal enzymes required for bone resorption are synthesized and released into the bone resorbing compartment of osteoclasts.

Bone lining cells

Bone lining cells are flattened cells which form a continuous epithelium-like layer on bony surfaces where active bone deposition or removal is not taking place. They are present on the periosteal surface as well as on the endosteal surface.

PHYSIOLOGICAL CONSIDERATIONS

The main physiological considerations which need emphasis are:

- Bone growth and
- Bone remodelling.

BONE GROWTH

The process of bone formation is called ossification. There are two mechanisms of bone formation: endochondral bone formation and intramembranous bone formation.

Endochondral bone formation. During fetal development, formation of most of the bones is preceded by the formation of a cartilaginous model, which is subsequently replaced by bone. This kind of ossification is called endochondral bone formation.

Intramembranous bone formation. Formation of some bones, e.g. clavicle, vault of skull and mandibles is not preceded by formation of a cartilage model, but they are formed directly in a fibrous membrane. This kind of ossification is called intramembranous bone formation.

Steps of growth of a long bone

1. Formation of a cartilage model. In the region, where a long bone is to be formed the mesenchyme first lay down a cartilaginous model of bone.

2. Ossification and calcification. The ossification is carried out by osteoblasts, which enter the central part of the cartilaginous model. This area is called *primary centre of ossification* (Fig. 8.4-8A). Gradually, bone formation extends from the primary centre towards the ends of shaft (Fig. 8.4-8B).

3. Growth in length and girth. At about the time of birth, developing bone consists of the bony diaphysis formed by

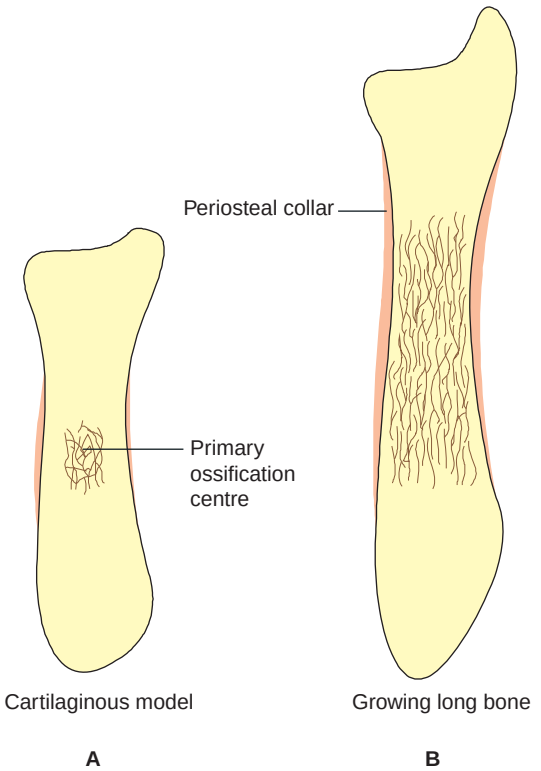


Fig. 8.4-8 Formation of a long bone: A, cartilage model with primary centre for ossification; and B, bone growth by extension of primary centre for ossification.

extension of primary centre for ossification and cartilaginous ends. At varying times after birth, *secondary centres* of endochondral ossification appear in the cartilages forming the ends of bones. These centres enlarge and convert the cartilaginous ends into bone. The portion of the bone formed from one secondary centre is called *epiphysis*. During growth, the bone of diaphysis and the bone of epiphysis are separated by a plate of actively proliferating cartilage, the *epiphyseal plate* (Fig. 8.4-9). The portion of the diaphysis adjoining the epiphyseal plate is called metaphysis. It is highly vascular and region of active bone formation. The bone increases in length as this plate lays down new bone on the end of shaft. The width of the epiphyseal plate is proportionate to the rate of growth. The width is affected by a number of hormones, but most markedly by the pituitary growth hormone and insulin-like growth factor (IGF-1). The bone increases in length as long as the epiphyseal plates remain separated from diaphysis (shaft). The growth of the bone stops when the epiphysis fuses with the diaphysis (*epiphyseal closure*). At this juncture, the cartilage cells stop proliferating, become hypertrophic and secrete vascular endothelial growth factor (VEGF), leading to vascularization and ossification.

Even after bone growth has ceased, the calcium turnover function of bone is most active in the metaphysis, which acts as a storehouse of calcium. The metaphysis does not have a bone marrow cavity and is frequently the site of infection.

BONE FORMATION

Bone formation is carried out by the active osteoblasts. Bone is continuously deposited by these cells. The process of bone formation includes two main processes:

1. Osteoid formation

The osteoblasts synthesize and lay down the type-I procollagen molecules into the adjacent extracellular space (Fig. 8.4-10A). These cells also secrete a gelatinous matrix in which the fibres get embedded. The collagen polymerizes to form collagen fibres which then swell up and can no longer be seen distinctly. The resultant mass of swollen fibres and matrix is called osteoid (Fig. 8.4-10B).

Factors affecting process of osteoid formation include protein intake and a number of growth factors, such as TGF-β (transforming growth factor), IGF-I (insulin-like growth factor), IGF-II, PDGF (platelet-derived growth factor), acidic and basic fibroblast growth factors, etc. Besides these growth factors, insulin, GH, sex hormones (oestrogens, androgen), thyroid hormones, calcitriol and calcitonin also affect the process of osteoid formation.

2. Bone matrix mineralization

Soon after formation of osteoid, the process of bone matrix mineralization starts.

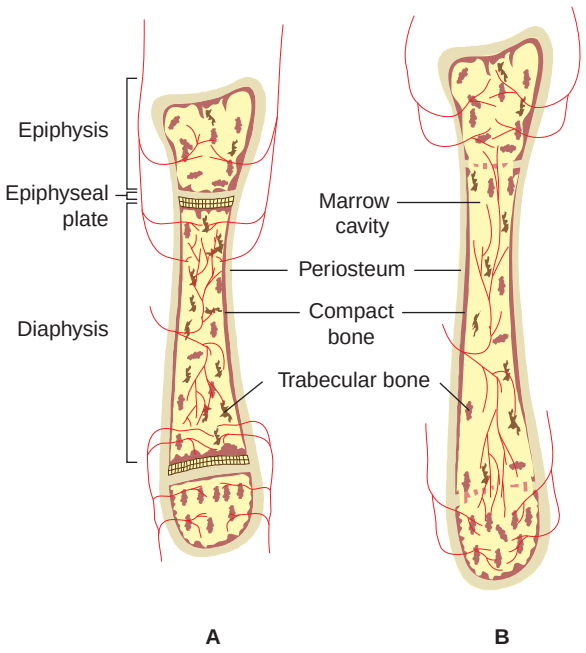


Fig. 8.4-9 Structure of a typical long bone before (A) and after (B) ossification.

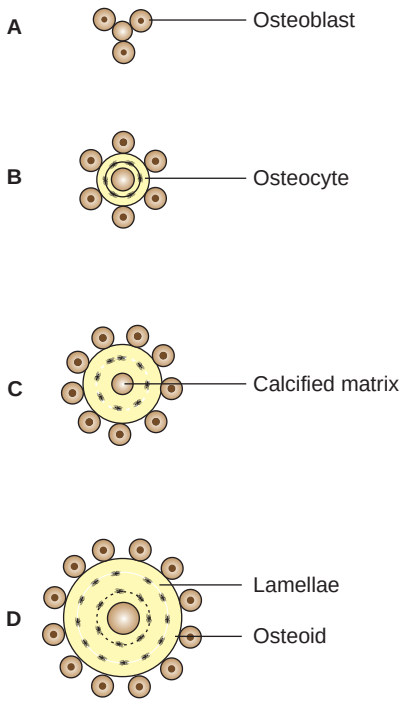


Fig. 8.4-10 Schematic depiction of process of formation of bony lamellae. For explanation see text.

Initiation of mineralization or nucleation. The bone matrix is surrounded by a metastatic solution of calcium and phosphate ions. The process of mineralization greatly depends upon the *calcium × phosphate* ion product in the ECF. This product must be above 30/dL for this process to occur.

Rapid calcification after enucleation. Once mineralization is initiated, i.e. after nucleation, most of the calcium phosphate is deposited within 6–12 h. Thereafter, hydroxide and bicarbonate ions are gradually added to the mineral mixture and mature hydroxyapatite crystals are slowly formed. After the process of mineralization of bone matrix is completed, the osteoid is converted into a bone lamella (Fig. 8.4-10C).

Formation of a trabecular bone

After the formation of one bone lamella (as described above Fig. 8.4-10A–C) another layer of osteoid is laid down by osteoblast. The osteoblasts move away from the bone lamella to line the new layer of osteoid. However, some osteoblasts are caught between the lamella and the osteoid (Fig. 8.4-10D). The osteoid is now ossified to form another lamella. The cells trapped between the two lamellae become osteocytes. In this way, a number of lamellae are laid down one over another and these lamellae together form a trabecula of bone, but many such trabeculae constitute the trabecular or cancellous bone.

Conversion of trabecular bone to compact bone

All newly formed bone is cancellous. It is converted into compact bone (Fig. 8.4-11):

- Each space between the trabeculae of cancellous bone comes to be lined by the osteoblasts (Fig. 8.4-11A and B).
- The osteoblasts lay down lamellae of bone as already described. The first lamella is formed over the inner wall of the original space and is therefore, shaped like a ring (Fig. 8.4-11C).
- Subsequently, concentric lamellae are laid down inside this ring thus forming an *osteon*. The original space becomes smaller and smaller and persists as a *Haversian canal* (Fig. 8.4-11D).

BONE RESORPTION

Bone resorption, like bone formation, is a continuous process. In bone resorption, there occurs destruction of entire matrix of bone resulting in a diminished bone mass.

Osteoclasts are the cells responsible for the bone resorption.

Process of bone resorption involves following steps:

1. Removal of unmineralized osteoid layers. Before osteoclastic resorption can begin, a thin 1–2 µm outer layer of unmineralized osteoid must be removed. This is achieved by collagenase released from lining cells. The lining cells also secrete a molecule that attracts osteoclasts to the site of new denuded bone.

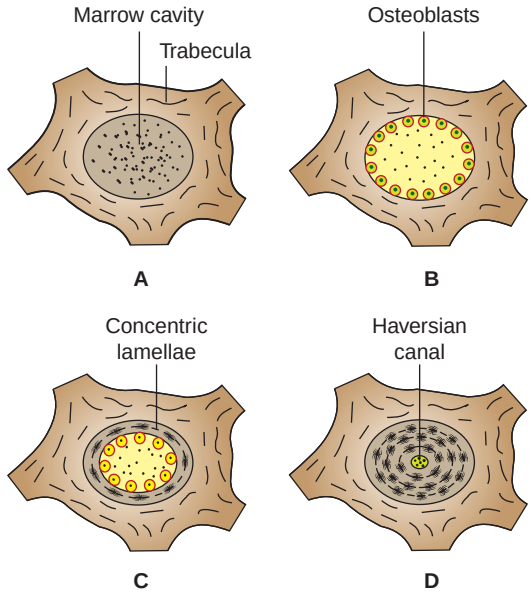


Fig. 8.4-11 Steps in the conversion of trabecular bone into compact bone.

2. Attachment of osteoclast on denuded bone surface (periosteum or endosteum) is the second step of bone resorption. This is mediated by the surface receptors called *integrins*. At the point of attachment, a ruffled border is created by infolding of the osteoclast's plasma membrane (villi formation). The part of the bone to be resorbed is called bone resorption compartment.

3. Release of proteolytic enzymes and acids. At the site of attachment, the osteoclasts release proteolytic enzymes and lysosomal enzymes and acid from the villi-like projections.

4. Digestion and dissolution of bone. The enzymes digest and dissolve organic matrix of the bone and acids cause dissolution of the bone salts. All the dissolved materials are now released into ECF, some elements enter the blood. The remaining elements are cleaned up by the macrophages and a shallow cavity is formed in the bone resorbing compartment. Urinary excretion of organic products released during resorption provides quantitative indices of the bone resorption.

Regulation of bone resorption. The bone resorption is stimulated by PTH, calcitriol, EGF (epidermal growth factor), PDGF and some other growth factors. The response is mediated through release of prostaglandins, TGF-β, and interleukin-I (IL-I) which stimulate osteoclastic activity. Thyroxine and vitamin A also increase bone resorption. Calcitonin acts on osteoclasts through its receptors to *inhibit* their activity.

BONE REMODELLING

Definition. Bone remodelling refers to a process of bone resorption followed by bone formation which keeps on occurring throughout life in a cyclic manner.

Bone remodelling unit. The bone remodelling appears to be the result of co-ordinated activity of groups of interacting osteoclast and osteoblast cells, which make up the bone remodelling unit. A single remodelling unit creates about 0.025 mm^3 of bone.

About 5% of the bone mass is being remodelled by about 2 million bone remodelling units in the human skeleton at any one time. The removal rate for bone is about 4% per year for compact bone and 20% per year for trabecular bone.

Phases of bone remodelling cycle. A bone remodelling cycle takes about 100 days and consists of two phases: the resorption phase and the succeeding formation phase.

1. Resorption phase lasts for initial 10 days. In this phase, mineralized bone is reabsorbed by osteoclasts releasing calcium and phosphate.

2. Formation phase lasts for next 90 days and is characterized by reformation of bone by osteoblasts (assimilating calcium and phosphate).

Regulation of bone remodelling. The paired activity of osteoclast and osteoblast cells in bone remodelling is well regulated. All aspects of the remodelling cycle are influenced by a large number of hormones and growth factors, as well as cytokines from immune cells. The process of bone remodelling is one example of co-ordinated function of the endocrine and immune systems.

Physiological significance of continuous bone remodelling includes:

- *Bone adjusts its strength* in proportion to the degree of bone stress. For example, in athletes, soldiers and others in whom the bone stress is more, the bones become heavy and strong.
- *Shape of bone can be rearranged* for proper support of mechanical force in accordance with the stress.
- *Old bone becomes relatively weak and brittle.* The development of new bone matrix maintains the toughness of bone.

CALCITROPIC HORMONES

PARATHYROID HORMONE

FUNCTIONAL ANATOMY OF PARATHYROID GLANDS

The parathyroid glands are two pairs of small endocrine glands closely applied to the back of the thyroid gland. Each gland is about the size of a split pea, measuring 6 × 4 × 2 mm. The total weight of four normal glands is about 140 mg.

Histological structure

The parenchyma of the parathyroid gland is made up of cells that are arranged in cords. The cells of the parathyroid glands are of two main types: chief cells and oxyphil cells.

Chief cells, also called as principal cells, are much more numerous. Chief cells secrete the PTH or parathormone.

Oxyphil cells. These cells are much larger than the chief cells and first appear at puberty and their function is still not clear.

STRUCTURE, SYNTHESIS AND SECRETION OF PTH

Structure. PTH is a single chain polypeptide, containing 84 amino acids and having molecular weight 9500.

Synthesis. PTH is synthesized from a precursor molecule called prepro-PTH, which contains 115 amino acids.

Secretion. PTH is released from the chief cells by exocytosis in response to decrease in plasma-ionized calcium concentration that is sensed by the calcium receptors in the parathyroid cells.

REGULATION OF PTH SECRETION

1. Role of plasma-ionized calcium. The secretion of PTH is mainly regulated by circulating levels of ionized calcium. The secretion of PTH is inversely related to the plasma calcium concentration. Maximum secretion occurs when plasma-ionized calcium levels fall below 3.5 mg/dL. As the plasma-ionized calcium concentration rises, the PTH secretion progressively diminished and reaches to a persistent low basal rate when ionized calcium reaches up to 5.5 mg/dL. Further, rise in plasma-ionized calcium levels do not further decrease PTH secretion (Fig. 8.4-12).

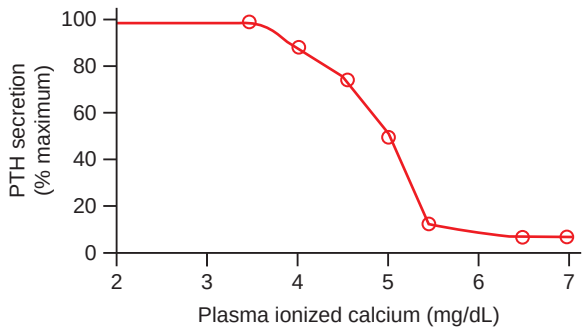


Fig. 8.4-12 The inverse relationship between parathormone (PTH) and plasma-ionized calcium.

2. Role of serum magnesium concentration

- *Mild decrease* in serum Mg^{2+} concentration stimulates PTH secretion, while
- *Severe decrease* in serum Mg^{2+} concentration inhibits PTH secretion and produces symptoms of hypoparathyroidism (e.g. hypocalcaemia).

3. Role of plasma phosphate concentration. A rise in plasma concentration of phosphate causes an immediate fall in ionized calcium concentration, which in turn stimulates PTH secretion.

4. Role of vitamin $1,25(OH)_2D_3$. It inhibits transcription of the PTH gene and decreases PTH secretion.

PLASMA LEVELS, HALF-LIFE AND DEGRADATION OF PTH

Plasma level of PTH is about l30 pg/mL (approximately 3×10^{-12} M).

Half-life of PTH in plasma is 5–8 min.

Degradation of PTH occurs rapidly in the peripheral tissues. PTH is predominantly split in the liver.

MECHANISM OF ACTION AND ACTIONS OF PTH

Mechanism of action of PTH

PTH binds to a membrane receptor proteins on the target cells (in bones, kidney and intestine) and activates adenylyl cyclase to liberate cAMP. The cAMP, in turn, increases intracellular calcium that promotes the phosphorylation of proteins (by kinases).

Actions of PTH

The prime function of PTH is to elevate plasma calcium concentration and to decrease the plasma phosphate concentration by acting on three major target organs: directly on bone and kidney, and indirectly on the gastrointestinal tract (Fig. 8.4-13).

1. Actions on the bone

Parathyroid hormone stimulates calcium and phosphate resorption from the bones, i.e. causes decalcification or demineralization of bone which occurs in two phases:

(i) Rapid phase of demineralization. This phase is also called *osteocytic osteolysis*. In this process, the calcium is transferred from the bone canalicular fluid into the osteocytes and then into the ECF. In this process, phosphate is not mobilized along with calcium.

(ii) Slow phase of demineralization. This effect requires several days of exposure to PTH. Parathyroid hormone stimulates the formation of new osteoclasts from the osteoprogenitor initiate process of bone resorption in which both calcium and phosphate are released from bone and are transferred to the ECF.

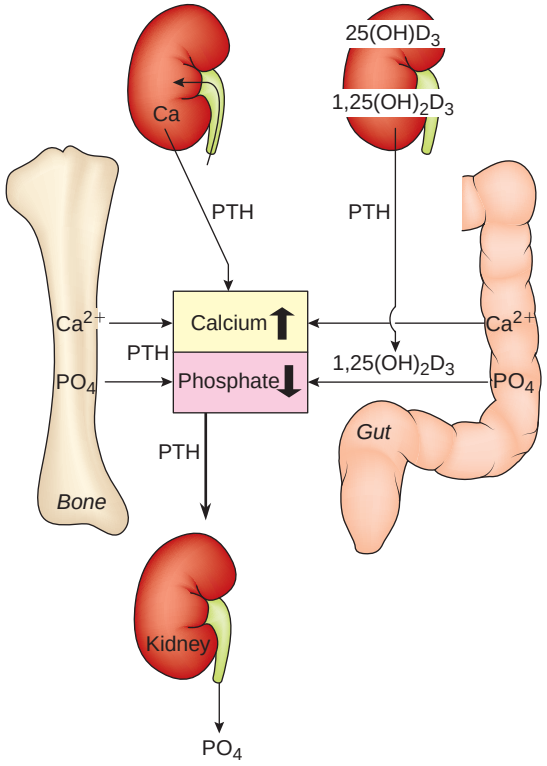


Fig. 8.4-13 Actions of PTH on bones (stimulation of calcium and phosphate resorption), kidneys (stimulation of calcium reabsorption but inhibition of phosphate reabsorption) and intestine (increase in absorption of calcium and phosphate both). PTH action leads to direct increase in calcium and decrease in serum phosphate level.

2. Actions on kidney

(i) Increase in calcium reabsorption. PTH increases the reabsorption of calcium from the ascending limb of loop of Henle and the distal tubules of kidney and helps to prevent hypocalcaemia.

(ii) Inhibition of phosphate reabsorption in the proximal tubule is the most dramatic effect of PTH on the kidney. This effect produces phosphaturia and hypophosphataemia.

(iii) Stimulation of reabsorption of Mg^{2+} by the renal tubules.

(iv) Stimulation of synthesis of 1,25-dihydroxycholecalciferol is a very important action of PTH in the kidney.

3. Actions on intestines

Parathormone greatly enhances both calcium and phosphate absorption from intestine indirectly by increasing synthesis of 1,25-dihydroxycholecalciferol in the kidney.

VITAMIN D

The term vitamin D refers to a group of closely related steroids produced by the action of ultraviolet light on certain provitamins. The active form of vitamin D, i.e. 1,25-dihydroxycholecalciferol also called as calcitriol is now considered a hormone.

FORMATION OF CALCITRIOL

Calcitriol $1,25(OH)_2D_3$ is an active form of vitamin D_3. Steps involved in its formation are summarized:

Source and synthesis of vitamin D_3

Vitamin D_3, the precursor (prohormone) of the hormone 1,25-dihydroxycholecalciferol, reaches the blood from two sources:

1. Dietary sources, include fish, fish liver oils, egg yolk. The daily requirement of vitamin D is 400 IU or 10 µg of cholecalciferol. In countries with good sunlight (like India), the recommended dietary allowance for vitamin D is 200 IU (or 5 µg cholecalciferol).

2. Cutaneous synthesis. Besides, dietary intake, vitamin D_3 is synthesized primarily in the specialized skin cells called *keratinocytes*, located in the inner layers of epidermis. The synthesis occurs by the action of ultraviolet (UV) rays on 7-dehydroxycholesterol (an intermediate in the synthesis of cholesterol). First pre vitamin D_3 is formed which is then converted spontaneously over 3 days to vitamin D_3 in a reaction that is driven by thermal energy from sunshine (Fig. 8.4-14).

Vitamin D_3 is the major storage and circulatory form of vitamin D. It is transported in the plasma bound to specific globulin called vitamin-D binding proteins (DBP).

Synthesis of hormone 1,25-dihydroxycholecalciferol ($1,25(OH)_2D_3$) from vitamin D_3 is accomplished by two steps: first step takes place in the liver and second in the kidney.

- In the liver, vitamin D_3 is converted to 25-hydroxycholecalciferol ($25(OH)D_3$) by the enzyme 25-hydroxylase (Fig. 8.4-14), which through circulation reaches the kidney (bound to DBP).
- In the kidneys, the enzyme 1α hydroxylase converts $25(OH)D_3$ to $1,25(OH)_2D_3$, i.e. 1,25-dihydroxycholecalciferol or calcitriol. In the kidneys, the less active metabolite 24,25-dihydroxycholecalciferol is also formed.

Regulation of synthesis of 1,25-dihydroxycholecalciferol. The formation of $1,25(OH)_2D_3$ in kidney is regulated as:

1. Plasma calcium levels regulate synthesis of $1,25(OH)_2D_3$ by a feedback mechanism indirectly through PTH.

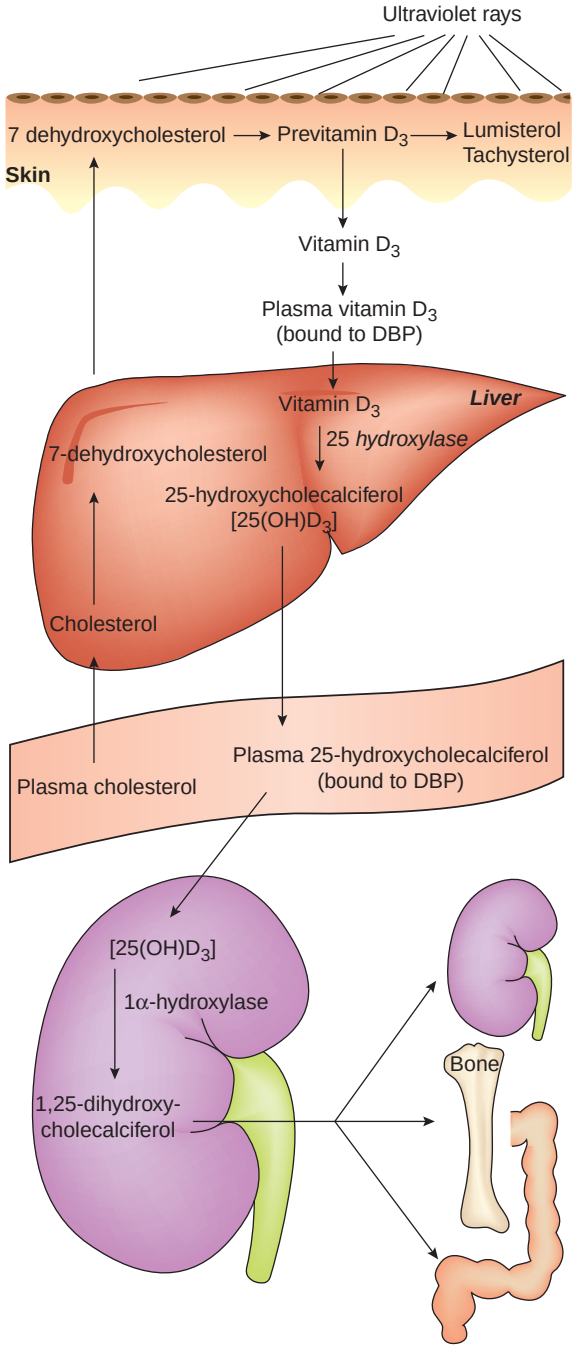


Fig. 8.4-14 Synthesis and sources of vitamin D_3 and its hydroxylation to form the hormone 1,25-dihydroxycholecalciferol. Main sites of actions of 1,25-dihydroxycholecalciferol are also shown.

- ↓ Calcium → ↑ PTH → ↑ $1,25(OH)_2D_3$
- ↑ Calcium → ↓ PTH → ↓ $1,25(OH)_2D_3$

2. Plasma phosphate level regulates the synthesis of $1,25(OH)_2D_3$ by a feedback mechanism by its direct effect on the enzyme 1,α-hydroxylase.

- ↓ Phosphate → ↑ 1,α-hydroxylase → ↑ $1,25(OH)_2D_3$ activity
- ↑ Phosphate → ↓ 1,α-hydroxylase → ↓ $1,25(OH)_2D_3$ activity

3. 1,25(OH)$_2$D$_3$ level itself has:

- A direct negative feedback effect on its formation and
- A direct action on the parathyroid gland to inhibit the production of mRNA for PTH.

↑ 1,25(OH)$_2$D$_3$ → ↓ 1,25(OH)$_2$D$_3$ formation; ↑ 24,25(OH)$_2$D$_3$ formation; ↓ PTH formation

4. *Other factors* regarding 1,25(OH)$_2$D$_3$ synthesis are:

- *Prolactin* increases 1,25(OH)$_2$D$_3$ synthesis.
- *Oestrogen* increases total circulatory 1,25(OH)$_2$D$_3$, probably due to an increase in the secretion of binding protein (DBP).
- *Hyperthyroidism* is associated with decreased circulating 1,25(OH)$_2$D$_3$ and an increased incidence of osteoporosis.
- *Metabolic acidosis* depresses the synthesis,
- *Growth hormone,* human chorionic somatomammotropin (HCS) and calcitonin stimulate the formation of 1,25(OH)$_2$D$_3$.

MECHANISM OF ACTION AND ACTIONS OF CALCITRIOL

Mechanism of action of calcitriol

Calcitriol (1,25(OH)$_2$D$_3$) acts by exerting its effect on the gene expression in the target cells by binding with the intracellular receptors. The vitamin D receptor is found both in the cytoplasm and nucleus. This mechanism has been described in detail on page 530.

Actions of calcitriol

I. Regulation of plasma levels of calcium and phosphate

Calcitriol [1,25(OH)$_2$D$_3$] is the biologically active form of vitamin D. It regulates the plasma levels of calcium and phosphate by acting at three different sites: intestine, bone and kidney.

1. Action on intestine. The major action of calcitriol is to help calcium absorption from the intestine (Fig. 8.4-15).

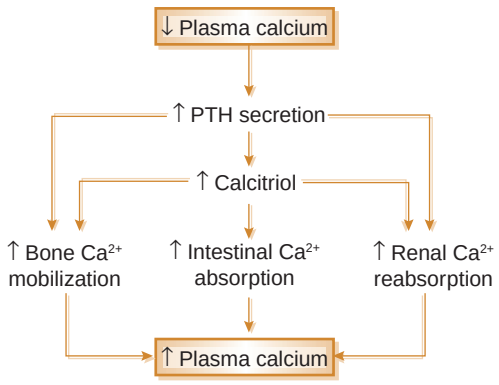


Fig. 8.4-15 Summary of actions of calcitriol in elevating plasma calcium.

It performs function by:

- Increasing permeability of brush border for Ca^{2+} absorption and
- Inducing synthesis of calbindin (Ca^{2+}-binding protein).

2. Actions on bone. Calcitriol increases bone resorption as well as bone mineralization.

- *Bone resorption*. Calcitriol helps bone resorption by PTH. The osteoclasts cause bone resorption for which PTH is also required.
- *Osteocytic osteolysis* is also increased by calcitriol.
- *Bone mineralization*. Calcitriol maintains levels of calcium and phosphate and calcium phosphate ion product in the normal range by causing bone resorption (as above). The ion product is important in the process of bone calcification. It also causes direct effect on the bone formation by increasing osteoblastic proliferation, alkaline phosphatase secretion and osteoclastin synthesis. Lack of vitamin D is associated with defective mineralization of cartilage as well as bones.

3. Action on kidneys. Calcitriol increases renal reabsorption of calcium and phosphate by increasing the number of calcium pump.

II. Other actions of calcitriol

Besides the above well known sites of action (intestine, bone and kidney) of vitamin D, the other possible actions of calcitriol in such tissues are summarized.

1. Calcium transport into the skeletal and cardiac muscles is stimulated by calcitriol. Therefore, vitamin D deficiency can result in muscle weakness and cardiac dysfunction.

2. Stimulation of differentiation of keratinocytes and inhibition of their proliferation are thought to be caused by calcitriol by its paracrine and autocrine function.

3. Stimulation of differentiation of immune cells is caused by calcitriol. Therefore an increased incidence of infections is noted in patients with deficiency of vitamin D. Calcitriol stimulates T-helper-2 cells to secrete interleukin-4 (1L-4) and TGF-β and T-helper-1 cells to decrease their production of interleukin-2, γ-interferon and tumour necrosis factor α.

4. Calcitriol appears to be involved in regulation of growth and production of growth factors.

CALCITONIN

SYNTHESIS AND STRUCTURE

Synthesis. Calcitonin is synthesized in the C-cells or parafollicular cells of the thyroid gland. These cells are of neural crest origin, which during development migrate to the developing thyroid gland.

Structure. Calcitonin is a straight-chain polypeptide with 32 amino acids. Its molecular weight is 3500.

Secretion. Calcitonin is secreted in response to rise in the plasma calcium level. The cAMP prompts exocytosis of calcitonin containing granules.

REGULATION OF SECRETION

1. Increase in plasma calcium concentration is the major regulator of calcitonin secretion. It is important to note that calcitonin is not secreted until the plasma Ca^{2+} concentration reaches to 9.5 mg/dL and that above this calcium level, plasma calcitonin is directly proportional to plasma calcium (Fig. 8.4-16).

2. Gastrointestinal hormones such as gastrin, CCK (cholecystokinin), glucagon and secretin have all been reported to stimulate calcitonin secretion, with gastrin being the most potent stimulus.

IMPORTANT NOTE

In Zollinger–Ellison syndrome, the plasma calcitonin level is high due to an increased secretion of gastrin. However, the dose of gastrin required to secrete calcitonin is very high.

3. Other factors like β-adrenergic agonist, dopamine and oestrogen also stimulate calcitonin secretion.

PLASMA LEVELS, HALF-LIFE AND DEGRADATION

Plasma levels of circulating calcitonin range from 10 to 20 pg/mL.

Half-life of calcitonin is very short, i.e. less than 10 min.

Degradation. Circulating calcitonin is heterogeneous and it is largely degraded and cleared by the kidney.

ACTIONS AND PHYSIOLOGICAL ROLE OF CALCITONIN

Actions

The major effect of calcitonin is to rapidly lower the plasma calcium level and it also decreases the plasma phosphate. These effects of calcitonin are due to its following actions:

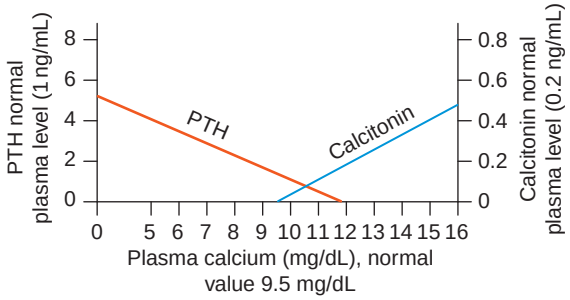


Fig. 8.4-16 Relationship of plasma calcium concentration with release of calcitonin and parathyroid hormone.

1. Action on the bone. The main action of calcitonin on the bone is to oppose the bone resorptive action of PTH. Calcitonin inhibits osteoclastic activity due to its direct action on the bone which can occur in the absence of parathyroid gland, gastrointestinal tract and kidneys.

2. Action on kidney. Calcitonin increases loss of calcium and phosphate in the urine. This effect also contributes in producing hypocalcaemia and hypophosphataemia.

Physiological significance of calcitonin

The possible physiological roles of calcitonin are:

- *In adults,* exact physiological significance of calcitonin is uncertain because bone resorption by osteoclasts leads to secondary osteoblastic activity. Therefore, effect on blood calcium concentration is transient and weak.
- *In children,* where bone turnover is high, the calcitonin may play a role in the skeletal development by promoting calcium storage in bones.
- *Post-prandial hypercalcaemia* may be prevented by calcitonin.
- *Protects the bones of mother from excess calcium loss during pregnancy* and lactation when demand for calcium to be used elsewhere dramatically increases.
- *Calcitonin could participate* in the fetal skeletal development.
- *Calcitonin may have a functional role* in the development of accelerated bone loss after the menopause.
- *Calcitonin is useful in the acute treatment of hypercalcaemia and in certain bone diseases,* in which a sustained reduction in osteoclastic resorption is therapeutically beneficial.
- *Calcitonin and calcitonin gene-related peptide (CGRP) may also have a paracrine and neurotransmitter function.*

PTH-RELATED PROTEIN AND OTHER HORMONES AFFECTING CALCIUM METABOLISM

PTH-RELATED PROTEIN

Origin and structure

Sites of origin. The PTH-related protein (PTHrP) is produced by many different tissues in the body, such as skin keratinocytes, lactating mammary epithelium, placenta and fetal parathyroid glands.

Structure. PTHrP has 140 amino acid residues, compared with 84 in PTH.

Physiological roles of PTHrP

The PTH-related protein is found in many tissues and may play a physiological role during intrauterine life and early infancy and later during development of different tissues.

1. Regulation of endochondral bone formation. PTHrP is important for the endochondral skeletal development. At puberty, the sex steroids stop the operation of PTHrP system and the epiphyseal growth plate permanently closes.

2. Role in the breast development. PTHrP is produced in large amounts in breast and is involved in the breast development and lactation. It is also secreted in large amounts in the milk.

3. Role in tooth development. PTHrP allows normal tooth development and eruption by resorbing alveolar bone.

4. Role in skin development. PTHrP acts as a growth factor for the development of skin and hair follicle.

5. Protective role in central nervous system. The PTHrP is found in the brain (cerebral cortex, hippocampus, and granular layer of the cerebellar cortex) where it protects the neurons from toxic overstimulation by glutamate receptors that activate voltage-dependent calcium channels.

OTHER HORMONES AND HUMORAL FACTORS AFFECTING CALCIUM AND BONE METABOLISM

Certain hormones, other than the calcitropic hormones described above, which also have some effect on calcium metabolism, are:

1. Growth hormone. This increases calcium excretion in urine, but it also increases intestinal absorption of calcium, and this effect seems to be greater than the effect on excretion, with a resultant positive calcium balance. Growth hormone also generates IGF-I, which stimulates protein synthesis in bone.

2. Glucocorticoids. These inhibit bone formation and increase bone resorption by several actions, resulting in osteoporosis.

3. Thyroid hormones. Normal plasma levels of thyroxine are essential for the proper skeletal development.

4. Insulin. This is required for bone formation and there is significant bone loss in untreated diabetics.

5. Oestrogens. These prevent osteoporosis, probably by direct effects on the osteoblasts. Therefore, incidence of osteoporosis in females increases after menopause.

APPLIED ASPECTS

Some of the important applied aspects with respect to endocrinal control of calcium metabolism and bone physiology are:

- Hyperparathyroidism and hypercalcaemia,
- Hypoparathyroidism and hypocalcaemia and
- Metabolic bone diseases.

HYPERPARATHYROIDISM AND HYPERCALCAEMIA

HYPERPARATHYROIDISM

Hyperparathyroidism is a clinical condition characterized by excessive secretion of PTH. It is of two types: primary and secondary.

Primary hyperparathyroidism

Aetiology. Primary hyperparathyroidism occurs due to excessive secretion of PTH by single autonomous parathyroid adenoma (most common).

Clinicobiochemical features

- *Typical manifestations* are hypercalcaemia, hypophosphataemia, hypercalciuria and renal calculi (kidney stones).
- *Hypercalcaemia* may produce muscle weakness, lethargy and constipation. Since calcium can stimulate release of gastrin there may occur hyperchlorhydria and peptic ulceration. Hypercalcaemia may also cause hypertension, cardiac arrhythmias and ECG changes.

Secondary hyperparathyroidism

In this condition, excessive PTH secretion occur secondary to persistent hypocalcaemia, which causes continued stimulation of parathyroid gland.

Aetiology. Secondary hyperparathyroidism is typically seen in slowly developing renal failure.

Clinicobiochemical features. The main characteristic feature of secondary hyperparathyroidism is involvement of bones. Bone pains, fractures and deformity may result. Alkaline phosphatase and osteocalcin levels are elevated.

HYPERCALCAEMIA

Causes

Causes of hypercalcaemia depending on the levels of PTH can be divided into two groups:

1. *Conditions associated with hypercalcaemia and raised PTH levels are described above.*

2. *Conditions associated with hypercalcaemia and low or undetectable PTH levels are:*

- Hypercalcaemia of malignancy,
- Multiple myeloma,
- Familial hypercalcaemia,
- Hyperthyroidism,

Hypercalcaemia of malignancy

Hypercalcaemia is not uncommon in malignancy. Tumours produce hypercalcaemia by two mechanisms:

- *Local osteolytic hypercalcaemia* is seen in 20% of the patients which have bone metastasis.
- *Humoral hypercalcaemia of malignancy* is seen in 80% of the patients who do not have bone metastasis. Hypercalcaemia in these patients is caused by raised levels of PTHrP. The gene encoding for PTHrP is different from the one coding for PTH.

Familial hypercalcaemia

Familial hypercalcaemia occurs due to mutations in the gene for Ca^{2+} receptor.

HYPOPARATHYROIDISM AND HYPOCALCAEMIA

HYPOPARATHYROIDISM

Hypoparathyroidism refers to a clinical condition characterized by low level of plasma calcium either due to deficient production of PTH or its unresponsiveness.

Hypoparathyroidism can be classified into two main groups:

- True hypoparathyroidism and
- Pseudohypoparathyroidism.

A. True hypoparathyroidism

In true hypoparathyroidism there is deficient production of PTH due to heritable or acquired causes.

Post-ablative or post-operative hypoparathyroidism. This is the most common cause of hypoparathyroidism is either damage to glands or their blood supply or their inadvertent removal during thyroidectomy operation. The incidence is 1% of all the thyroidectomies.

B. Pseudohypoparathyroidism

This is a congenital condition, in which PTH production is normal but the target tissues are resistant to its effects. The defect may lie in parathyroid receptors or there may be post-receptor defect. The clinical and biochemical features are similar to hypoparathyroidism, but PTH levels are elevated (since hypocalcaemia produces more production of PTH).

Characteristic features of hypoparathyroidism

Characteristic features of hypoparathyroidism are:

Hypocalcaemia. Total serum calcium may be decreased to 4–8 mg/dL and the ionized calcium to 3 mg/dL. A 50% fall in the levels of ionized calcium leads to a clinical condition called *tetany* (described below).

Hyperphosphataemia, i.e. an increase in serum inorganic phosphate levels to 6–16 mg/dL.

TETANY

Tetany refers to a clinical condition resulting from increased neuromuscular excitability.

Causes

Causes of tetany include:

1. Hypocalcaemia. Extracellular calcium plays an important role in membrane integrity and excitability. Thus when concentration of ionic calcium is reduced to <50% of normal in ECF, cell membrane of neurons becomes more permeable resulting in a series of action potentials. Thus hypocalcaemia is the most common cause of increased neuromuscular irritability leading to tetany.

2. Hypomagnesaemia also causes tetany, because magnesium ions are also associated with neuromuscular irritability.

3. Alkalosis, which reduces ionic calcium, can also produce tetany.

Clinical features

Symptoms and signs depend upon the age of the patients. The following symptoms may be seen:

- *Carpopedal spasm.* The hands in carpopedal spasm adapt a peculiar posture in which there occurs flexion at metacarpophalangeal joints, extension at interphalangeal and there is apposition of thumb (Fig. 8.4-17). This peculiar posture of hand is called *obstetric hand.* Pedal spasm is less frequent. In it the toes are plantarflexed and feet are drawn up.
- *Laryngeal stridor* (loud sound) results from spasm of laryngeal muscles. It may produce asphyxia.

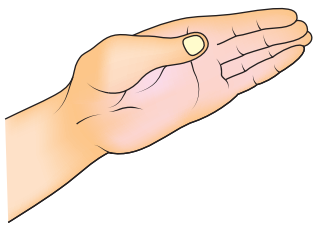


Fig. 8.4-17 Carpal spasm in a patient with tetany.

- *Paraesthesias,* i.e. tingling sensations in the peripheral parts of limbs or around the mouth is common feature.
- *Trousseau's sign* (pronounced as 'Troosoz's sign'). Occluding the blood supply to a limb for about 3 min by inflation of a sphygmomanometer cuff (above the systolic blood pressure level) produces characteristic carpal spasm.
- *Chvostek's sign* refers to the twitching of facial muscles produced by tapping the facial nerve at the angle of jaw. This occurs due to increased excitability of nerves to mechanical stimulation.

IMPORTANT NOTE

Latent tetany. In latent or subclinical tetany, the above described typical symptoms and signs of tetany are absent, but can be unmasked by following provocative tests:

- *Trousseau's sign and*
- *Chvostek's sign*

Management

Management of tetany includes an intravenous injection of 20 mL of 10% calcium gluconate is given to correct hypocalcaemia and relieve tetany.

METABOLIC BONE DISEASES

The term metabolic bone disease is used for the bone diseases, such as rickets, osteomalacia and osteoporosis.

Rickets (occurring in children) and osteomalacia (occurring in adults) are metabolic bone diseases produced due to deficiency of vitamin D in which there is defective calcification of bone matrix.

RICKETS

As mentioned above, in rickets mineralization of organic bone matrix in growing children is defective. It may also involve cartilaginous matrix of growing end plates of bones.

Causes and types of rickets

Depending upon the cause, rickets is of following types:

A. Vitamin D deficiency rickets is the most common variety. It may occur:

1. Nutritional rickets is caused by dietary deficiency of vitamin D, either due to poor intake or poor absorption due to high phytate in diet. Vegetarian diet is a poor source of vitamin D.

2. Deficient synthesis through skin due to inadequate exposure to the sun light in smoggy cities is particularly known to cause rickets. This disorder usually manifests between the age of 6 months to 2 years, when the growth of bones is very rapid and the baby is fed only on milk and is kept mostly indoor.

B. Vitamin D-resistant rickets is of rare occurrence. In it, rickets occurs without deficiency of vitamin D. It is caused by inactivating mutations of the gene for renal hydroxylase resulting in non-formation of 1,25-dihydroxycholecalciferol from vitamin D_3 but a normal response to 1,25-dihydroxycholecalciferol is seen.

Clinical features

- *Clinical features of rickets* seen in children are:
  - *Craniotabes* refers to the small rounded areas in the membranous bones of skull which yield under pressure of finger.
  - *Widening of wrist* occurs due to epiphyseal widening of lower end of radius bone.
  - *Collapse of chest wall* occurs due to flattening of sides of thorax with prominent sternum.
  - *Rickety rosary* refers to the beading of costochondral junction of ribs.
  - *Frontal bossing* and posterior flattening of skull.
  - *Bowing of legs* or knock-knee occurs when child starts walking.
  - *Kyphosis and pelvic deformities* are also known.
- *Management.* Adequate supply of calcium and vitamin D should be ensured. Therapeutic dose of vitamin D varies from 25 to 125 mg (1000–5000 IU) daily for 6–8 weeks followed by 200–400 IU daily.

OSTEOMALACIA

Osteomalacia can be thought of an adult counterpart of rickets. It is characterized by defective mineralization of the adult bones in which epiphyseal growth plates are already closed.

Clinical features

Clinical features of osteomalacia are almost similar to rickets.

1. Skeletal abnormalities are dominant features and include:

- Diffuse skeletal pain and bony tenderness are common complaints. Pain may vary from mild backache to severe pain around hip.
- Muscle weakness is also common. A waddling gait may be present due to proximal muscle weakness.

2. Tetany may occur in few cases with carpopedal spasm.

Treatment

Treatment is similar to rickets.

OSTEOPOROSIS

Osteoporosis is characterized by a reduction of bone mass per unit volume with normal ratio of bone matrix and minerals, i.e. there occurs loss of both bone matrix and mineral component. Senile osteoporosis is common disease and has become a major public health problem of elderly.

In males, bone loss is usually less significant than the females. Further, after menopause, women initially have more rapid bone loss because of additional factor of oestrogen deficiency.

Factors contributing to development of osteoporosis include:

- *Immobilization.* Osteoporosis of immobilization occurs if bones are not subjected to the stress of walking.
- *Weightlessness.* Weight bearing is essential for maintenance of bone mass. Weightlessness in space travel produces significant osteoporosis within few months.
- *Hyperthyroidism* is associated with more increased osteoclastic activity vis-à-vis bone formation activity, resulting in bone loss.
- *Hyperthyroidism* and Cushing's syndrome are also associated with increased osteoclastic activity and secondary osteoporosis.
- *Bone secondaries* are associated with local release of certain osteoclastic factors which cause osteoporosis.
- *Chronic renal failure* is associated with osteomalacia, but in severe cases osteoporotic bone lesions due to tertiary hyperparathyroidism may also be present.

Characteristic features of osteoporosis

1. Bone density is reduced. In radiographs, the affected bones show clear glass appearance (e.g. ground glass appearance seen in osteomalacia). In severe cases, excessive bone resorption may lead to cyst formation (osteitis fibrosa cystica).

2. Incidence of fractures is increased, particularly, fractures of the distal forearm (Colles' fracture), vertebral bodies and hips are more common in osteoporosis.

Treatment

Treatment of osteoporosis should include:

1. *Calcium intake,* particularly from natural sources, such as milk should be increased.
2. *Moderate exercise* may be useful in preventing or slowing the progress of osteoporosis.
3. *Oestrogen treatment* is effective in arresting the rapidly developing osteoporosis in women after menopause.

OSTEOPETROSIS

Osteopetrosis is a rare condition in which bone density increases due to defective bone resorption. The osteoclast activity becomes defective due to lack of a protein (encoded by gene c-fos), therefore due to unopposed activity of osteoblasts bone density increases. The steady increase in bone density leads to:

- Neurological defects due to narrowing of foramina of bones through which nerves pass.
- Haematological defects due to narrowing (crowding) of bone marrow cavity.

Chapter 8.5

Adrenal Glands

FUNCTIONAL ANATOMY

GENERAL CONSIDERATIONS

- There are two adrenal glands, situated one on either side, at the upper pole of kidney, hence also called 'suprarenal gland' (Fig. 8.5-1).
- Normally, each gland weighs about 5 g and consists of two parts, the adrenal cortex and the medulla.

HISTOLOGICAL STRUCTURE

Adrenal cortex

The adrenal gland is covered by a connective tissue capsule from which septa extend into the gland substance. The mature human adrenal cortex consists of three distinct layers or zones of cells (Fig. 8.5-2):

1. Zona glomerulosa, constituting outer one-fifth of cortex, is a small zone present under the capsule. It consists of cells that secrete *aldosterone* and *corticosterone.*

2. Zona fasciculata is the widest zone forming middle three-fifths of the cortex. It is made up of cells that are arranged in two cell thick straight columns. Sinusoids intervene between the columns.

3. Zona reticularis forms the inner one-fifth of the cortex. It is made up of a network of compactly arranged cords of cells, (hence the name zona reticularis).

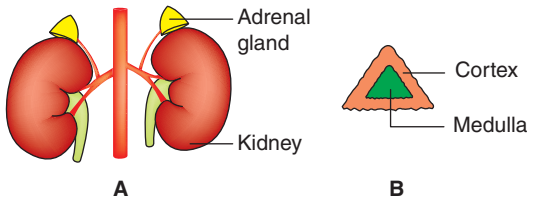


Fig. 8.5-1 A, Location and B, divisions of adrenal glands.

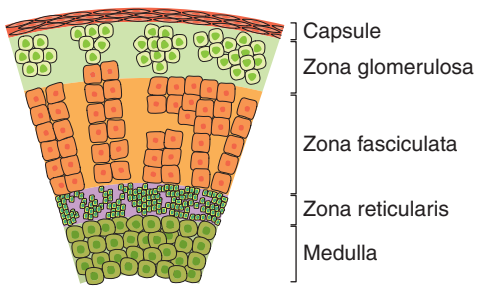


Fig. 8.5-2 Schematic histological structure of adrenal gland.

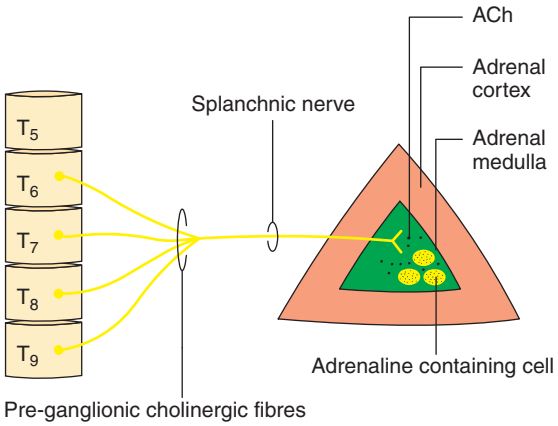


Fig. 8.5-3 Pre-ganglionic sympathetic fibres synapsing directly on the chromaffin cells in the adrenal medulla.

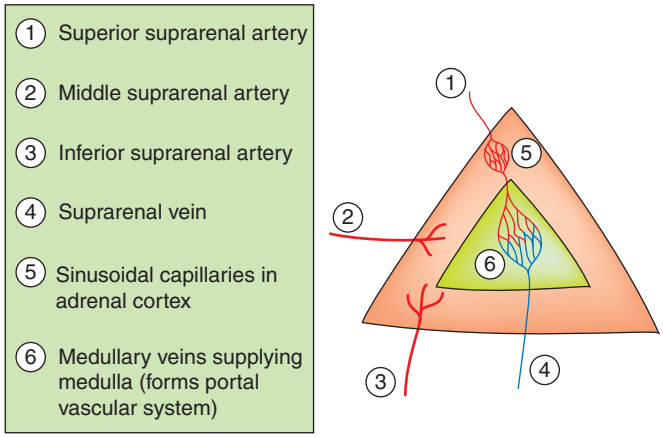


Fig. 8.5-4 Schematic diagram to show arterial blood supply and venous drainage of adrenal gland. The portal vascular system which exposes the medulla to high concentration constitutes a functional connection between the cortex and the medulla.

Zona fasciculata and zona reticularis constitute a single functional unit where mainly cortisol (and some corticosterone) and androgen (dehydroepiandrosterone, i.e. DHEA) are synthesized.

Adrenal medulla

Histologically, it is made up of chromaffin cells, innervated by preganglionic sympathetic neurons.

Chromaffin cells. The cells forming adrenal medulla show yellow granules in their cytoplasm (i.e. *chromaffin reaction*) and hence called chromaffin cells.

- Functionally, these cells are considered to be modified post-ganglionic neurons which do not have axons.
- *Catecholamines are stored in the chromaffin granules.* In addition to the catecholamines, the chromaffin granules also contain proteins, lipids and adenine nucleotides (mainly ATP).

Nerve endings, present in the adrenal medulla, are the *cholinergic preganglionic* sympathetic fibres that synapse directly on the chromaffin cells. These fibres traverse the splanchnic nerve and are myelinated emanating mainly from the lower thoracic segments (T_5 and T_9) of the ipsilateral intermediolateral grey column of the spinal cord (Fig. 8.5-3).

BLOOD SUPPLY (FIG. 8.5-4)

Arterial blood supply. The adrenal glands have one of the body's highest rates of blood flow per gram of tissue. The arterial blood to the gland reaches the outer capsule from the superior suprarenal artery (a branch of the inferior phrenic artery), middle suprarenal artery (a branch of the abdominal aorta) and inferior suprarenal artery (a branch of the renal artery). The arterial blood enters sinusoidal capillaries in the cortex and then drains into the medullary veins, which supply blood to medulla and thus form *a portal system.* This arrangement of portal circulation exposes the medulla to relatively high concentrations of corticosteroids from the cortex.

Venous drainage. The venous blood drains via single central vein. The right suprarenal vein drains into the inferior vena cava and left suprarenal vein into the left renal vein.

HORMONES OF ADRENAL CORTEX

Hormones secreted by the adrenal cortex, called *corticosteroids,* can be grouped as:

- *Glucocorticoids,* which include *cortisol* and *corticosterone,* have widespread effect on glucose and protein metabolism.
- *Mineralocorticoids.* Aldosterone is the chief mineralocorticoid. It regulates sodium balance and extracellular fluid (ECF) volume in the body.
- *Adrenal sex steroids.* These include DHEA and its sulphate ester.

GLUCOCORTICOIDS

SYNTHESIS

The glucocorticoids are synthesized largely by the cells forming zona fasciculata with a small contribution by the cells of zona reticularis of adrenal cortex. The steps involved in the synthesis of glucocorticoids are summarized in Fig. 8.5-5 and their intracellular localization is depicted in Fig. 8.5-6.

Uptake of cholesterol. The corticosteroid hormones are synthesized from cholesterol which is actively taken up by the adrenal cells from low-density lipoprotein of the blood. After entry into the cell, most of the cholesterol is stored as cholesterol ester. Under basal conditions, corticosteroids

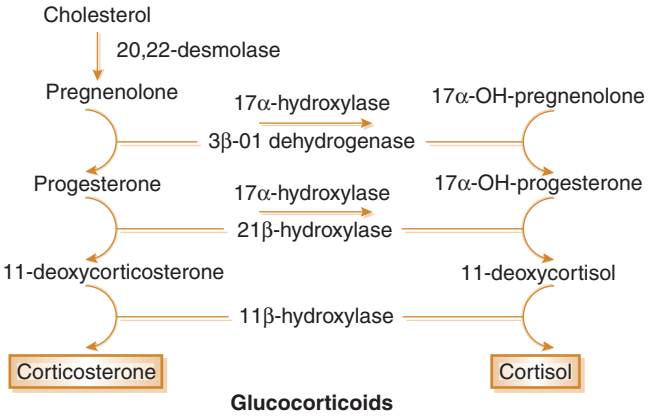


Fig. 8.5-5 Steps involved in the synthesis of glucocorticoids in the zona fasciculata and zona reticularis.

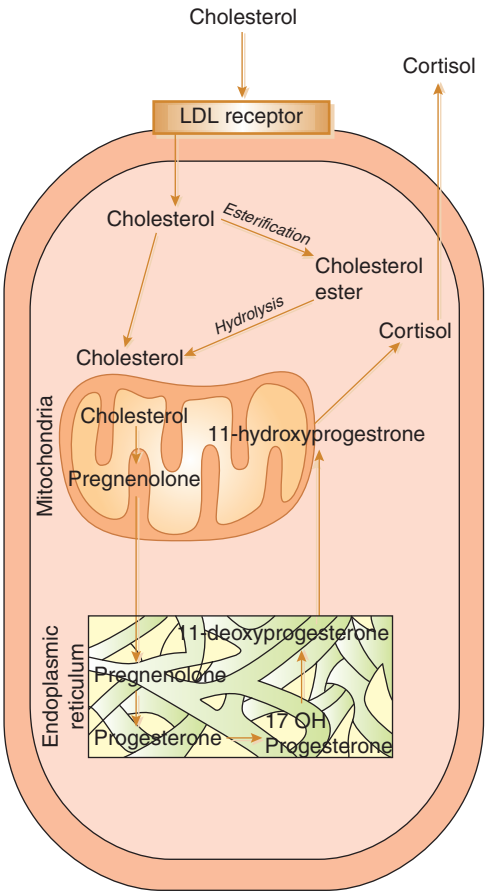


Fig. 8.5-6 Intracellular sites involved in the synthesis of cortisol.

are synthesized from free plasma cholesterol but when the production is stimulated by adrenocorticotropic hormone (ACTH), the stored esterified cholesterol becomes the most important precursor, which is hydrolysed by the cholesterol esterase. The free cholesterol enters the mitochondria.

Five oxidative CYP (formerly P 450) enzymes (Table 8.5-1) act at various ring carbons of cholesterol to form corticosteroid hormones. As shown in Fig. 8.5-7, the basic steroid structure ring (cyclopentoperhydrophenanthrene nucleus) consists of four rings designated as A, B, C and D. The individual carbon atoms comprising the steroid ring are numbered 1–21 (Fig. 8.5-7). Substituent groups in derivative steroid molecules are designated by the number of carbon ring atom to which they are attached.

Side-chain cleavage of cholesterol. This is caused by the mitochondrial enzyme 20, 22 desmolase also known as side-chain cleavage enzyme (P-450 scc). As a result, the cholesterol is converted to pregnenolone, which is a common precursor of all the steroid hormones.

Conversion of pregnenolone to 11-deoxycortisol and 11-deoxycorticosterone. This occurs by the hydroxylation reactions. These reactions follow subsequently after the formation of pregnenolone and progesterone and occur within the *endoplasmic reticulum* (Fig. 8.5-6).

Conversion of 11-deoxycortisol to cortisol. The 11-deoxycortisol is transferred back into the mitochondria (Fig. 8.5-5). This step is very efficient in humans, 95% of the 11-deoxycortisol formed is converted into cortisol. The cortisol so formed rapidly diffuses out of the cell.

Conversion of 11-deoxycorticosterone to corticosterone. Under normal circumstance, little corticosterone is formed and cortisol is the dominant glucocorticoid in humans. The 11-deoxycorticosterone is converted to corticosterone by the enzyme 11-β-hydroxylase.

PLASMA LEVELS, TRANSPORT, METABOLISM AND EXCRETION OF GLUCOCORTICOIDS

Plasma levels

Plasma levels of glucocorticoids and other corticosteroids are shown in Table 8.5-2. The plasma levels of total cortisol show diurnal fluctuation and range from 10 to 25 µg/dL with an average of 14 µg/dL. The rate of secretion of cortisol, which is about 15 mg/day under normal condition, may increase to 300–400 mg/day under conditions of severe stress.

Transport

Cortisol. In the plasma cortisol circulates in two forms: bound (90%) and free (10%).

Bound form. Most of the plasma cortisol is bound to specific corticosteroid binding α_2-globulin, which is a glycoprotein and is also called *transcortin.* A small amount (15%) is bound to albumin.

Free form of cortisol constitutes only 5–10% of the total plasma cortisol. However, it is the free form which is responsible for the physiological actions of the hormone including the feedback regulation of ACTH.

Table 8.5-1 Nomenclature and location of enzymes involved in synthesis of glucocorticoids

Trivial name	Code		Location	Action
	Past	Current		
20,22-Desmolase	$P450_{SCC}$	CYP-11-A1	Mitochondria	Cleaves the side chain between carbons 20 and 22 of cholesterol
3β-01-Dehydrogenase	3β-HSD	3β-HSD	Endoplasmic reticulum microsome	Catalyses conversion of pregnenolone to progesterone
17α-Hydroxylase	$P450_{C17}$	CYP-17	Endoplasmic reticulum microsome	Catalyses the hydroxylation of C-17
21-Hydroxylase	$P450_{C21}$	CYP-21-A2	Endoplasmic reticulum microsome	Catalyses the hydroxylation of C-21
11β-Hydroxylase	$P450_{C11}$	CYP-11-B1	Mitochondria	Catalyses the hydroxylation at C-11

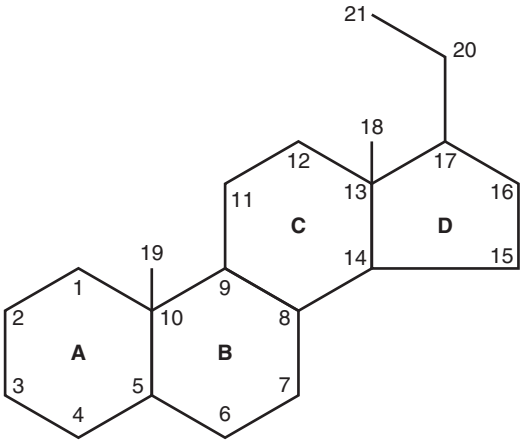


Fig. 8.5-7 Basic structure of steroid ring. Note, the four rings are designated A, B, C and D.

Inter-relationship of free and bound form. At the normal levels (average 14 µg/dL), the free form is less than 10% of total plasma cortisol. When the total plasma cortisol increases beyond 20 µg/dL, the binding sites on transcortin are saturated and there occurs some increase in albumin binding but main increase is in the free form.

IMPORTANT NOTE

Transcortin levels in the third trimester of pregnancy are twice that in non-pregnant state. The increased total cortisol is, however, not associated with the symptoms of excess cortisol as the levels of (physiologically active) free form of cortisol are normal.

Transcortin levels in plasma are decreased in—cirrhosis liver (decreased synthesis), nephrosis (more loss in urine), and in multiple myeloma. Decreased levels of plasma transcortin are associated with decreased total plasma cortisol levels but no symptom of cortisol insufficiency, as the levels of free form are normal.

Corticosterone. Like cortisol, it also exists mainly in bound form, but to a lesser degree. That is why, its half-life is slightly shorter than cortisol.

Table 8.5-2 Average 8 AM plasma levels and secretion rate of corticosteroids in adult humans

Corticosteroid	Plasma concentration (µg/dL)	Secretion rate (mg/day)
Cortisol	14.0	15
Corticosterone	1.0	03
Aldosterone	0.009	0.15
Dehydroepiandrosterone (DHEA) sulphate	115	15

Metabolism and excretion

Corticosteroids are degraded in the liver and conjugated with glucuronic acid.

Major pathway of cortisol and cortisone metabolism is shown in Fig. 8.5-8. The reduced metabolites of cortisol and cortisone (the cortol and cortolone) are conjugated and excreted in the urine as cortol glucuronide and cortolone glucuronide, respectively.

Minor pathway of cortisol metabolism includes its conversion to 17-ketosteroid derivatives, which are conjugated to sulphates and are rapidly excreted in urine. Approximately 10% of the secreted cortisol is metabolized by this pathway (Fig. 8.5-8).

Excretion. There is enterohepatic circulation of glucocorticoids. About 70% of the conjugated steroids are excreted in the urine and about 20% in the faeces.

A total of about 22 mg of glucocorticoid derivatives are excreted in urine per day.

IMPORTANT NOTE

The rate of inactivation and conjugation of glucocorticoids is decreased in liver diseases and during surgery or other stresses. Therefore, in such conditions, the plasma-free cortisol level rises higher than it does with maximum ACTH stimulation in the absence of stress.

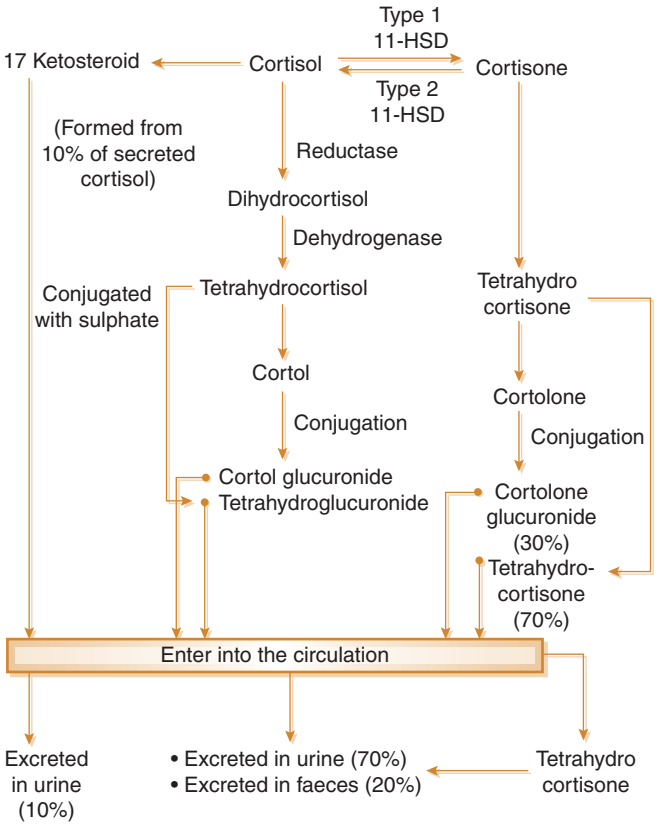


Fig. 8.5-8 Pathways of metabolism and excretion of glucocorticoids.

MECHANISM OF ACTION OF GLUCOCORTICOIDS

Like other steroid hormones, the glucocorticoids act through effect on gene expression by binding with specific intracellular receptors called glucocorticoid receptor (GR). For details of mechanism of action, see page 530.

ACTIONS OF GLUCOCORTICOIDS

Glucocorticoids are essential for survival.

I. Metabolic effects of glucocorticoids

Cortisol has major effects on glucose, protein, and fat metabolism. The different metabolic effects of glucocorticoids are:

1. Effects on carbohydrate metabolism. Glucocorticoids exert an anti-insulin effect, which leads to hyperglycaemia by following actions (Fig. 8.5-9):

(i) Increased gluconeogenesis. Glucocorticoids increase the rate of glucose production from non-carbohydrate sources by as much as 6–10 folds by following mechanisms:

- Accelerating the synthesis of hepatic enzymes (e.g. glucose-6-phosphatase) involved in gluconeogenesis.
- Providing more amino acids to the liver for gluconeogenesis by their catabolic effect on muscle protein and inhibitory effect on protein synthesis.

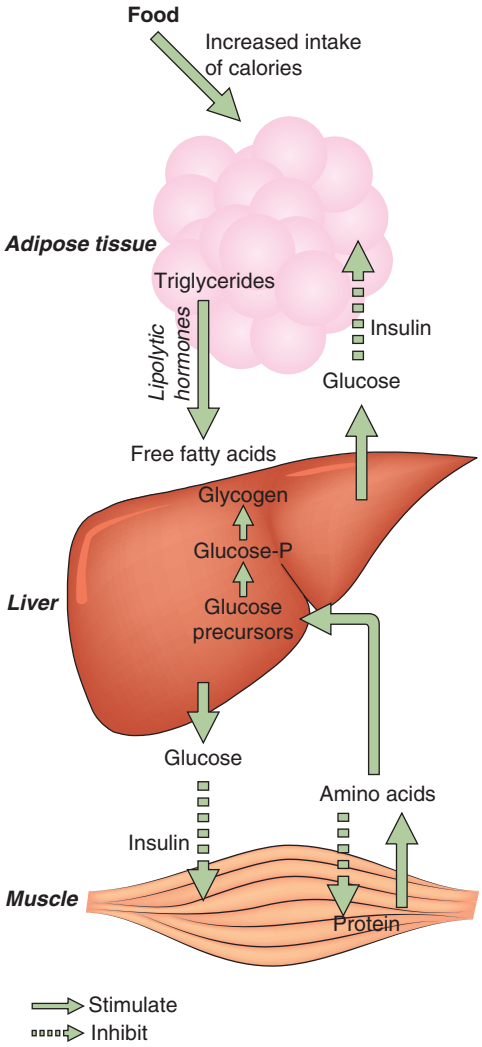


Fig. 8.5-9 Metabolic effects of cortisol. ***Note:*** Cortisol increases intake of calories by increasing appetite; facilitates release of amino acids from the muscle and their use for gluconeogenesis in liver and storage as glycogen; inhibits peripheral utilization of glucose and synthesis of proteins from amino acids and facilitates release of free fatty acids from the adipose tissues.

- Providing more glycerol to liver for gluconeogenesis by increasing lipolysis.

(ii) Decreased utilization of glucose in peripheral tissues. Glucocorticoids inhibit glucose uptake by peripheral tissues like the muscle, skin and connective tissue, lymphoid tissue, bone and adipose tissue. The heart, brain, liver and erythrocytes are spared from this action.

Note. It is important to note that normally increased glucose synthesis is associated with glycogen breakdown. But glucocorticoids promote gluconeogenesis as well as the storage of carbohydrates as hepatic glycogen. Glycogen synthesis is increased by increasing glycogen synthetase.

2. Effects on protein metabolism exerted by glucocorticoids are (Fig. 8.5-9):

- *Catabolic effect.* Cortisol enhances the release of amino acids by proteolysis in the skeletal muscle and other extrahepatic tissues.
- *Antianabolic effect.* It is the ability of the glucocorticoids to inhibit the de novo synthesis of protein.

Note. In the liver, the glucocorticoids increase the synthesis of enzymes involved in the production of hepatic proteins, plasma proteins and glycogen.

3. Effects on fat metabolism are complex and include:
(i) Lipolytic effects. Although, cortisol itself has only a slight lipolytic activity, its presence is necessary for epinephrine, growth hormone and other lipolytic substances to stimulate hydrolysis of stored triglycerides at maximal rates.

Fatty acid synthesis is inhibited in the liver by cortisol, an effect which not observed in the adipose tissue.

Note. It is important to note that in diabetic patients, cortisol increases plasma lipid level and increases ketone bodies formation and makes diabetes worse. But in normal subjects, insulin secretion is increased by raised blood glucose level, and the insulin decreases lipase activity and counterbalances hyperglycaemia.

(ii) Lipogenic role. Glucocorticoids increase differentiation of adipose tissue cells and stimulate lipogenesis.

IMPORTANT NOTE

Lipogenic effect varies in different regions of the body. Therefore, in cortisol excess there occurs a selective accumulation of fat in the abdomen, trunk and above (trunked obesity) sparing the extremities, which become thin due to loss of muscle mass. The deposition of fat in face is called 'moonface', and that in the suprascapular region is referred to as '*buffalo hump*' or 'dowagers hump' (features Cushing's syndrome).

Glucocorticoids increase appetite, thus food intake by inducing Neuropeptide-Y(NPY) and Leptin synthesis which act on appetite centre located in hypothalamus. For details see page 743.

4. Effects on electrolyte and water metabolism. Glucocorticoids control distribution of body water and electrolytes by their opposing actions:

(i) Retention of sodium and water by aldosterone-like activity, the glucocorticoids, increase sodium and chloride retention and potassium excretion by the kidney.

(ii) Promotion of diuresis. Glucocorticoids promote diuresis by increasing the inactivation of ADH by liver and by antagonising the action of ADH at the level of distal convoluted tubules of the kidney.

II. Physiological actions on various organs and systems

In addition to the metabolic effects noted above, the glucocorticoids affect various organs and systems throughout the body (Fig. 8.5-10):

1. Effects on muscle
(i) Contractility and work performance of skeletal and cardiac muscle are maintained by the cortisol.

(ii) Decrease in muscle mass and strength is caused by an excess of cortisol. This occurs due to the decrease in muscle protein synthesis and increase in muscle catabolism.

2. Effects on bone
(i) Increased bone resorption by increasing activity of osteoclasts and collagenase enzyme

(ii) Inhibition of bone formation by:

- Decreasing collagen synthesis,
- Inhibiting formation of mature osteoblasts from the undifferentiated cells,
- Increasing rate of apoptosis of osteoblasts and osteocytes,
- Impeding calcium absorption from the intestinal tract by antagonizing the action of $1,25(OH)_2$ vitamin D_3 and inhibiting its synthesis and
- Increasing calcium excretion in urine, as glomerular filtration rate (GFR) increases.

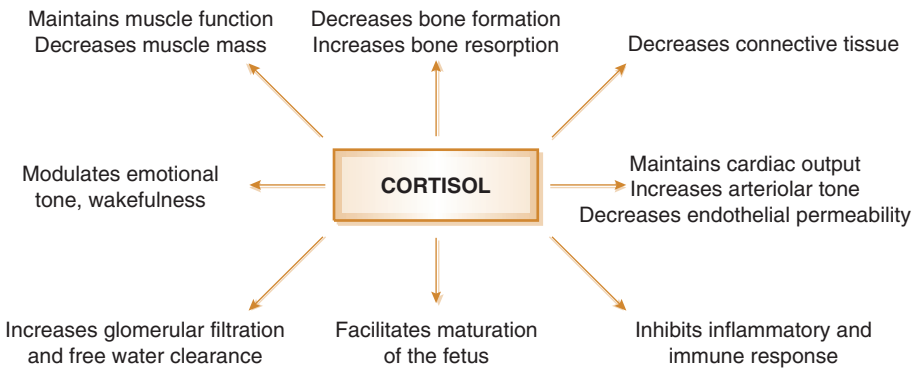


Fig. 8.5-10 Effects of glucocorticoids on various tissues, organs and systems of the body.

Note. Because of the above effects on bone, *osteoporosis* results in skeletal deformity.

3. Effects on connective tissue. Cortisol decreases collagen synthesis producing thereby:

- Thinning of the skin and
- Thinning of walls of capillaries, which leads to their easy rupture and to intracutaneous haemorrhage.

4. Effects on vascular system. Cortisol is essential for maintaining normal blood pressure by:

- Sustaining myocardial performance
- Enhancing the vasopressure effect (responsiveness of arterioles to constrictive effect) of catecholamines (especially norepinephrine) and angiotensin II
- Decreasing production of vasodilator prostaglandins
- Maintaining normal blood volume by decreasing the permeability of the vascular endothelium.

5. Effects on kidney are:

- Increase in GFR by increasing glomerular plasma flow,
- Rapid excretion of water load and
- Increase in calcium and phosphate excretion by decreasing their reabsorption in the proximal tubules.

In the absence of cortisol-free water clearance is diminished, because the activity of ADH is not antagonized (also see effect on water and electrolyte metabolism).

6. Effects on central nervous system. GR are present in various parts of the brain, especially in the limbic system. Through these receptors, the glucocorticoids modulate excitability behaviour and mood.

7. Effects on gastrointestinal tract. The glucocorticoids increase gastric acid secretion and decrease proliferation of gastric mucosal cells. These effects can lead to peptic ulceration following long-term use of cortisol.

8. Effects on blood cells and lymphatic organs. The excess of glucocorticoids lead to:

- *Eosinopenia* and *basopenia* due to their increased destruction and increased sequestration in lungs and spleen. The changes in eosinophilic count have been used as an index of change in ACTH secretion.
- *Lymphopenia*, i.e. decrease in the number of lymphocytes is caused by inhibiting their proliferation and increasing their destruction in the circulation. This leads to decreased size of lymph nodes, thymus, spleen and other lymphoid tissues.
- *Neutrophilia*, i.e. increase in neutrophil count occurs due to their increased release from bone marrow and decreased migration into tissues from the vascular spaces.
- *Polycythaemia*, i.e. increased red blood cell (RBC) count occurs due to the stimulation of erythropoiesis.
- *Thrombocytosis*, i.e. increased platelet count.

III. Anti-inflammatory and antiallergic effects

These effects are not produced by the glucocorticoids which are normally secreted physiologically, but are produced by their large doses when administered therapeutically and are thus called the pharmacological actions of glucocorticoids.

1. Anti-inflammatory effects of glucocorticoids are produced by following actions (Fig. 8.5-11).

- Cortisol inhibits the activity of phospholipase A_2,
- Cortisol stabilizes the lysosomal membrane,
- Cortisol inhibits migration of circulating leucocytes to the site of inflammation,
- Cortisol inhibits leukotriene and
- Cortisol decreases collagen formation.

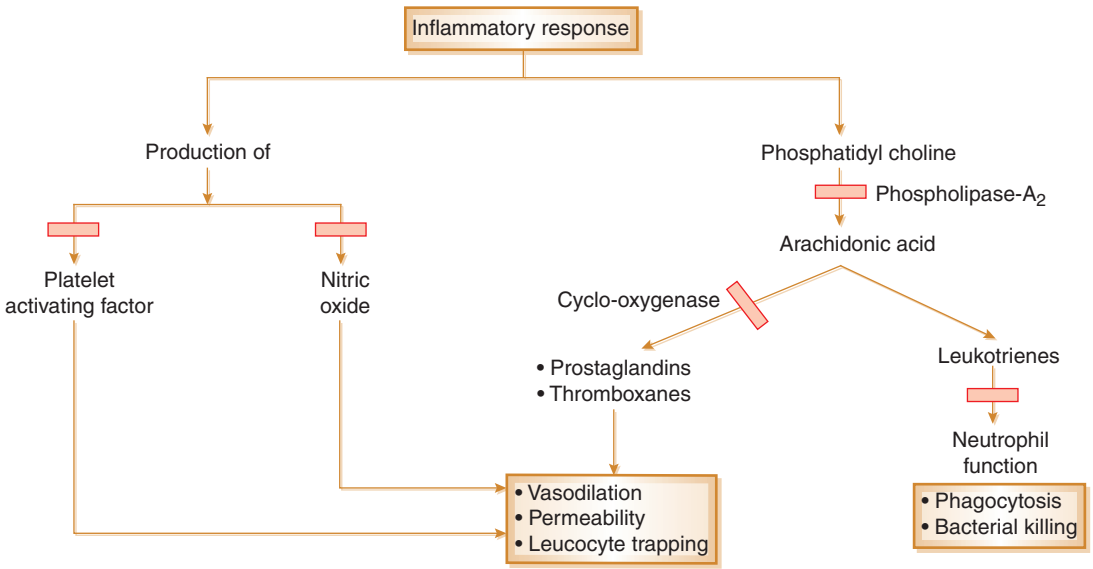


Fig. 8.5-11 Sites (▭) of action of cortisol in inhibiting inflammatory response.

IMPORTANT NOTE

Glucocorticoids must be used undercover of antibiotics in patients with bacterial infection, otherwise signs and symptoms get masked and serious and fatal complications may occur due to delay in diagnosis and start of proper treatment (antimicrobial drugs).

2. Anti-immunity effect. Cortisol inhibits both cellular and humoral immunity by decreasing the proliferation of T cells (involved in cellular immunity) and B cells (involved in humoral immunity) response (Fig. 8.5-12).

Note. glucocorticoids are used as an immunosuppressor in recipient of an organ (transplant) to prevent graft rejection.

3. Antiallergic effect. Cortisol reduces the number of circulating basophils and protects against the release of secretory products of granulocytes, mast cells, and macrophages, which have vesicles containing serotonin, histamine and hydrolases.

IV. Role of glucocorticoids in fetal life

- *Maturation of central nervous system (CNS), retina and skin* is facilitated by the cortisol in utero
- *Maturation of lungs.* The pulmonary surfactant lowers the surface tension in pulmonary alveoli and thus permits proper inflation of lungs immediately after birth.
- *Maturation of gastrointestinal tract.* The digestive enzyme capacity of the intestinal mucosa changes from a fetal pattern to a mature adult pattern under the influence of cortisol. This maturation process allows the newborn to digest disaccharides present in the milk.

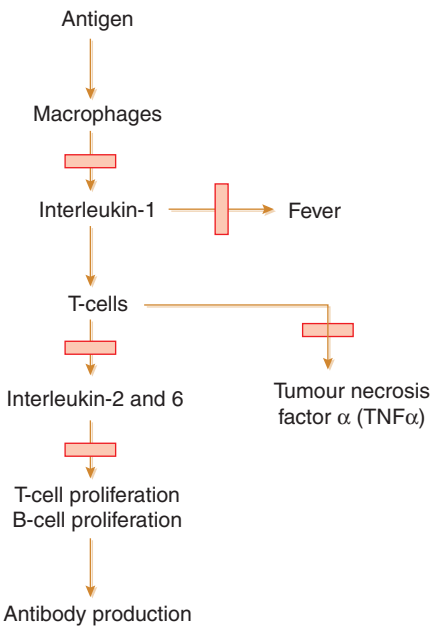


Fig. 8.5-12 Sites (▭) of action of cortisol inhibiting immune response.

V. Role of glucocorticoids in stress

Various stresses, e.g. trauma, cold, illness, starvation are associated with activation of the hypothalamic–hypophyseal–adrenal axis. Increased secretion of glucocorticoids is one of the various mechanisms involved in adaptation to various stresses (see page 599).

REGULATION OF GLUCOCORTICOID SECRETION

The glucocorticoid secretion is regulated by hypothalamic anterior pituitary–adrenal cortex axis, which exerts its effect through (Fig. 8.5-13):

- Corticotropin-releasing hormone (CRH),
- Adrenocorticotropic hormone (ACTH) and
- Glucocorticoids negative feedback effect.

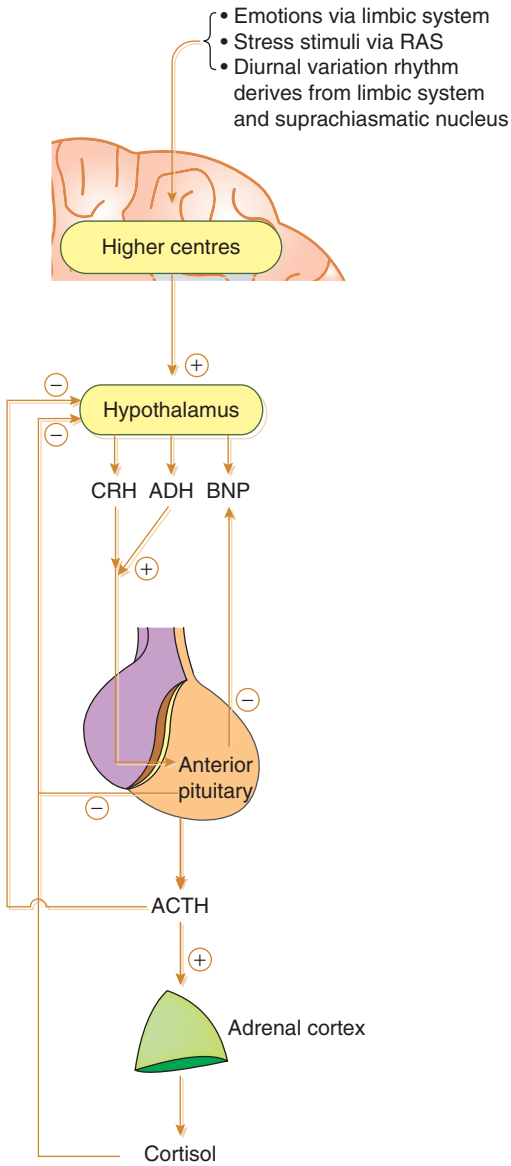


Fig. 8.5-13 Hypothalamic–anterior pituitary–adrenal cortex axis and negative feedback mechanism controlling glucocorticoid secretion.

Note. Since CRH and ACTH are related to regulation of glucocorticoid release, they are described in detail here rather than along with other hormones of hypothalamus and anterior pituitary, respectively.

1. Role of corticotropin-releasing hormone

Corticotropin-releasing hormone is secreted by the small cells of paraventricular nucleus of hypothalamus. Corticotropin-releasing hormone is a polypeptide with 41 amino acids. The CRH reaches the anterior pituitary through hypothalamic–hypophyseal portal system and acts by cAMP mechanism through the CRH receptors.

Control of CRH secretion

(i) Stressful stimuli (e.g. pain, anaesthesia, surgery, haemorrhage etc.), which ultimately increase cortisol secretion within minutes to as much as 20-fold primarily act on the hypothalamus to increase the secretion of CRH.

(ii) Circadian rhythm. The CRH secretion, and thus ACTH and cortisol secretion show circadian rhythm (Fig. 8.5-14).

(iii) ACTH acts on the hypothalamus to reduce the secretion of ACTH by short-loop negative feedback mechanism.

(iv) Glucocorticoids exert a long-loop negative feedback effect on CRH secretion (Fig. 8.5-13).

Actions of CRH

(i) Stimulation of synthesis and release of ACTH by acting on the *corticotrophs.*

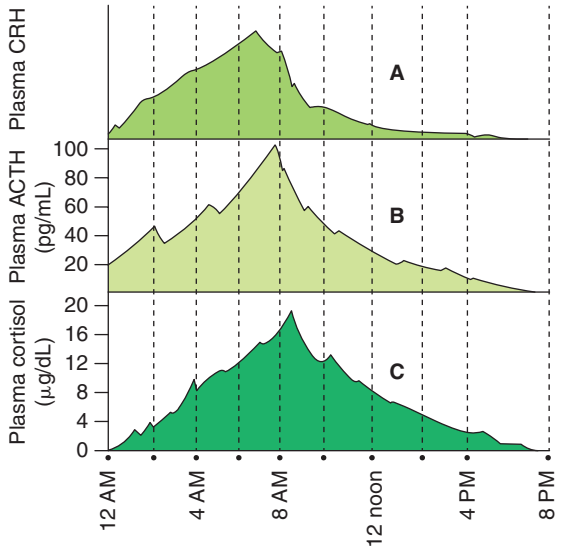


Fig. 8.5-14 Diurnal variations and pulsatility in secretion of CRH (A), ACTH (B) and cortisol (C). Note the ACTH peak follows CRH peak and cortisol peak follows ACTH peak by 10 min.

(ii) Other actions of CRH related to or independent to ACTH include:

- Central arousal,
- Increase in blood pressure,
- Diminution of reproductive function by decreasing synthesis of gonadotropin-releasing hormone (GnRH) and gonadotropins,
- Decrease in feeding activity and growth, and
- Stimulation of release of cytokines in immune cells.

2. Role of adrenocorticotropic hormone

Adrenocorticotropic hormone is secreted by *corticotrophs* and is a straight chain peptide containing 39 amino acids with a molecular weight of 4500.

Synthesis. ACTH, along with other peptides co-secreted in plasma, are derived from the single precursor, the pro-opiomelanocortin consisting of 241 amino acids.

Mechanism of action. ACTH acts by combining with the ACTH receptors present on the surface of the adrenal cortical tissue cells. The receptor–hormone complex in the presence of calcium activates the cAMP, which is principal second messenger.

Actions of ACTH

(i) Actions on adrenal cortex. ACTH is primarily concerned with growth and functions of the adrenal cortex:

- It promotes conversion of cholesterol to pregnenolone, which is the precursor of synthesis of all the hormones of adrenal cortex.
- It stimulates the secretion of glucocorticoids and the adrenal androgens.

(ii) Extra adrenal actions of ACTH occur only with very high levels, which are seen in abnormal conditions.

Regulation of ACTH secretion

(i) Hypothalamic control on ACTH secretion is mainly exerted through CRH (see above). The hypothalamic control is responsible for following characteristics of ACTH secretion:

- *Diurnal variation* in the levels of ACTH (and thus of cortisol) is due to variation in CRH release. As shown in Fig. 8.5-14, a large peak in the levels of ACTH and cortisol occurs in the morning (6–8 AM) during awakening (plasma level of ACTH range between 20 and 100 pg/mL with an average of 50 pg/mL). Thereafter, the average level decreases markedly (5 pg/mL), just before or after the subject falls asleep. In night workers, the rhythm is reversed. The biological clock responsible for diurnal variation in CRH, ACTH and cortisol levels is located

either in the limbic system and suprachiasmatic nucleus of the hypothalamus.

- *Pulsatile release of ACTH is* also due to pulsatile release of CRH. Up to three pulses per hour and each pulse lasts about 20 min. Cortisol pulses follow the ACTH pulses.
- *Release of ACTH in response to stress* is mediated by CRH release (see above) and ADH release. During stress, ADH significantly augments the effect of CRH.
- *Brain natriuretic peptide,* an analogue of atrial natriuretic hormone inhibits the ACTH release and the ACTH responses to CRH stimulation in humans.

(ii) Negative feedback inhibition of ACTH is caused by (Fig. 8.5-13):

- Plasma cortisol levels (long-loop negative feedback) and
- Plasma ACTH levels (short-loop negative feedback).

3. Negative feedback control of glucocorticoid secretion

Chronically, elevated plasma levels of free cortisol (and not the total cortisol, i.e. free plus bound form) exert a direct negative feedback action on its own secretion. This effect is exerted at two levels (Fig. 8.5-13):

- On hypothalamus to decrease formation of CRH and
- On anterior pituitary to decrease formation of ACTH.

APPLIED ASPECTS

Two important clinical applications of the negative feedback control of glucocorticoid secretion which need to be considered are:

(i) *Therapeutic administration of exogenous glucocorticoids* for a long period suppresses the secretion of CRH and ACTH by negative feedback mechanism. However, the absence is not noticed because the pharmacologic doses of exogenous glucocorticoids continue to perform the physiological functions of glucocorticoids as well. But, when the exogenous glucocorticoids are stopped suddenly, the hypothalamus, pituitary and adrenal cortex cannot recover equally suddenly leading to *acute adrenal deficiency* characterized by sudden fall in the blood pressure. The individual may even die of *adrenal crisis*. Therefore, to avoid this complication, exogenous steroids should never be stopped suddenly but their doses should be tapered slowly over a long period.

(ii) *Dexamethasone suppression test*. This test is based on the ability of dexamethasone (a potent synthetic glucocorticoid) to inhibit ACTH secretion. When the hypothalamic–pituitary–adrenocortical axis is normal, the administration of dexamethasone inhibits the secretion of ACTH and cortisol.

MINERALOCORTICOIDS

The mineralocorticoids include:

- Aldosterone. It is the chief mineralocorticoid,
- Deoxycorticosterone (DOC) and
- 18-hydroxy-deoxycorticosterone (18-OH-DOC) is secreted in small amount and has some mineralocorticoid activity.

Since aldosterone is the major mineralocorticoid so discussion in this section is limited to it.

SYNTHESIS

Aldosterone, the chief mineralocorticoid, is synthesized exclusively by the *zona glomerulosa cells.* The steps involved in the synthesis summarized in Fig. 8.5-15, include:

Uptake of cholesterol for formation of corticosterone. This involves the same steps as in the synthesis of glucocorticoids in zona fasciculata (for details see page 582).

Formation of aldosterone. Some of the corticosterone is hydroxylated and converted to an aldehyde by aldosterone synthase, a mitochondrial P_{450} mixed oxygenase to yield aldosterone, which is rapidly released. The 18-hydroxycorticosterone is not a direct intermediate but a by-product of this enzyme reaction (Fig. 8.5-15).

PLASMA LEVELS, TRANSPORT, METABOLISM AND EXCRETION

Plasma levels. Depending upon the dietary intake of sodium, the aldosterone secretion ranges from 50 µg/day (with dietary sodium intake of 150 mEq) to 250 µg/day (with dietary sodium intake of 10 mEq). Plasma levels of aldosterone show diurnal variation with a highest concentration at 8 am (0.009 µg/dL) and lowest at 11 PM.

Transport. In the plasma, 40% aldosterone circulates in free form and 60% in bound form. Aldosterone is weakly bound to the specific aldosterone-binding globulin to transcortin and to albumin.

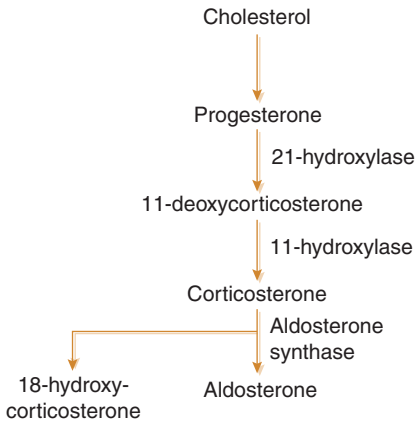


Fig. 8.5-15 Steps involved in the synthesis of aldosterone.

Metabolism and excretion. Ninety percent of aldosterone like the glucocorticoids is degraded in the liver and is reduced to *tetrahydroaldosterone,* the major metabolite that is excreted in the urine as aldosterone-3-glucuronide conjugate. A smaller amount of aldosterone is conjugated in the kidney and excreted in urine as 18-glucuronide. The aldosterone-18-glucuronide is most commonly measured in the urine for diagnostic purposes. Values of this metabolite in subjects with a normal sodium diet range from 5 to 20 µg/day.

ACTIONS OF ALDOSTERONE

A. Primary actions of aldosterone

1. Effects on renal tubules

Aldosterone acts on late distal tubules and collecting ducts of kidney and causes following effects:

(i) *Sodium reabsorption* from the tubular fluid into the renal tubular epithelial cells.

(ii) *Potassium excretion.* In the kidney, the active reabsorption of Na^+ occurs in exchange of K^+ and H^+. Thus, aldosterone not only causes reabsorption of Na^+, but excretion of K^+ as well by renal tubular epithelial cells.

(iii) *H^+ excretion.* Aldosterone also enhances the tubular secretion of H^+ as Na^+ is reabsorbed.

(iv) *Ammonium and magnesium* excretion is also increased by aldosterone.

Mechanism of action. Aldosterone acts by promoting specific protein synthesis. The *aldosterone-induced protein (AIP) synthesis* increases Na^+ reabsorption by following effects:

- *Membrane permeability* of tubular cells is increased by AIP increasing passive Na^+ absorption along the electrical and concentration gradients.
- *Increase in number of thiazide-sensitive NaCl co-transporters* is caused by aldosterone in the apical membrane of the distal convoluted tubular cells. The NaCl co-transporters increase the inflow of Na^+ from the tubular urine to the renal cells.
- *Increase in the content of Na^+–K^+–ATPase* at the basal (capillary) surface of the renal tubular cells is caused by aldosterone, which pumps the sodium out and then back into the plasma.
- *Increase in the Krebs' cycle enzyme activity* in the mitochondria caused by aldosterone helps to generate energy required for extrusion of Na^+ into the interstitial fluid and capillary blood.
- *Increase in phospholipase activity* caused by aldosterone in the cytosol of cells leads to increased synthesis of fatty acids, which are used in membrane generation.

2. Effects on sweat glands, salivary glands and colon

Sweat glands and salivary glands produce primary secretions, which contain a large amount of sodium chloride. The sodium chloride is absorbed as the secretion passes through the ducts, and in turn K^+ and HCO_3^- are excreted. Thus aldosterone decreases the loss of Na^+ and Cl^- in sweat and salivary secretion.

Colon. The aldosterone stimulates sodium reabsorption from the colon while enhancing potassium excretion in the faeces.

B. Secondary effects of aldosterone

The secondary effects include:

1. Effects on plasma potassium concentration

(i) Hypokalaemia, i.e. decrease in plasma K^+ levels may occur in aldosterone excess due to increased urinary excretion of K^+. When plasma K^+ levels fall below 2.5 mEq/L, the hypokalaemia may produce.

(ii) Hyperkalaemia may occur in aldosterone deficiency.

2. Effects on plasma sodium levels

(i) Hypernatraemia, which may occur in excessive aldosterone, may lead to:

- Increase in ECF volume. Absorption of Na^+ from renal tubules causes simultaneous osmotic absorption of water. This increases the ECF volume.
- Hypertension may occur due to Na^+ and water retention.

(ii) Hyponatraemia, which may occur due to excess Na^+ loss in aldosterone insufficiency.

Aldosterone escape

As mentioned above in hyperaldosteronism or when aldosterone is administered for several days to normal individuals, the increased sodium and water absorption by renal tubules leads to increase in ECF volume and hypertension. However, after 10–15% increase in ECF, there occur *pressure diuresis,* i.e. excretion of sodium and water in urine is increased in spite of continued action of aldosterone. This phenomenon is called *aldosterone escape.* This is probably due to the increased secretion of *atrial natriuretic peptide* (ANP).

IMPORTANT NOTE

Normally, glucocorticoids do not have mineralocorticoid like action on kidney and other tissues due to presence of an enzyme *11β hydroxysteroid dehydrogenase type-2.* This enzyme converts glucocorticoids to 11-oxy derivative that prevents their binding to mineralocorticoid receptors. Deficiency of enzyme 11β hydroxysteroid dehydrogenase type-2 results in a syndrome called *apparent mineralocorticoid excess.*

REGULATION OF ALDOSTERONE SECRETION

Aldosterone secretion is controlled by following factors (Fig. 8.5-16):

1. Renin–angiotensin system

The secretion of aldosterone is influenced by changes in the circulating fluid volume, which are sensed in the kidney. The signals arising from the kidney increase aldosterone secretion when ECF volume is decreased and vice versa.

Conditions associated with decreased ECF are:
- Sodium deprivation (e.g. dietary restriction),
- Haemorrhage,
- Upright posture for several hours and
- Acute diuresis.

Steps involved in the secretion of aldosterone by renin–angiotensin system are:

- *Decrease in ECF volume* leads to decrease in the renal arterial blood flow and pressure.
- *Decrease in renal perfusion pressure* causes the juxtaglomerular cells of the afferent arterioles to secrete renin.
- *Renin* catalyzes the conversion of angiotensinogen (alpha 2-globulin substrate present in the plasma) to angiotensin I.

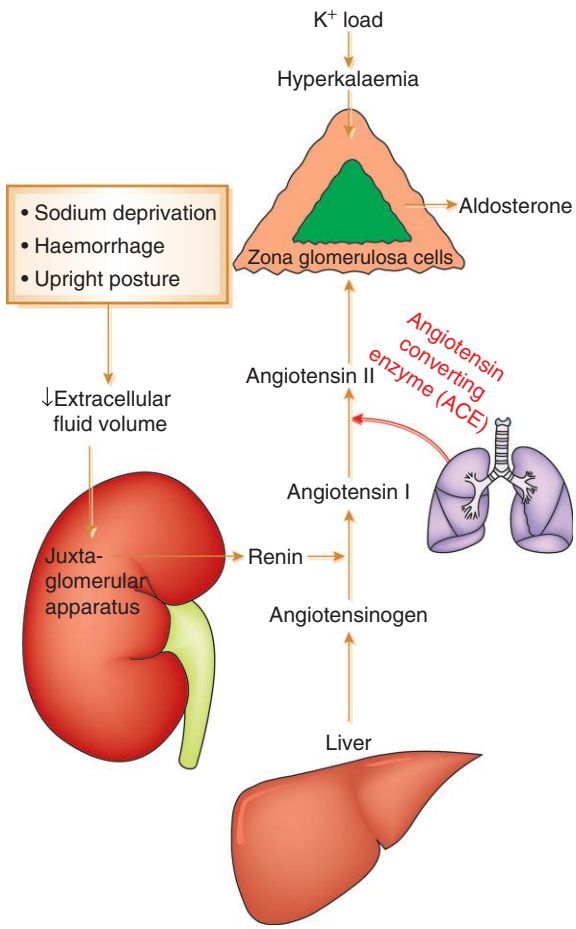


Fig. 8.5-16 Regulation of aldosterone secretion.

- *Angiotensin I* is converted into angiotensin II by the action of angiotensin converting enzyme present in the endothelium of blood vessels, especially in the lungs.
- *Angiotensin II* binds to specific plasma membrane receptors in adrenal's zona glomerulosa cells and increases the secretion of aldosterone.

Note. In addition to increasing secretion of aldosterone, angiotensin II also exerts other effects in controlling ECF volume and blood pressure (see page 417).

Aldosterone secretion, may be increased to 4–8 fold by renin–angiotensin system.

Factors affecting aldosterone secretion by the renin–angiotensin system are:

- *Sympathetic neural activity* enhances renin release in response to hypovolaemia
- *Local prostaglandins* also stimulate renin release, therefore antiprostaglandin drugs can reduce aldosterone response.
- *Atrial natriuretic peptide* (ANP) reinforces the effects of the renin–angiotensin system on aldosterone secretion.

2. Plasma potassium concentration

There exists a vital negative feedback relationship between plasma potassium concentration and aldosterone secretion, i.e.

- *An increase in plasma concentration* by only 0.5 mEq/L immediately raises plasma aldosterone levels to 3 fold.
- *Decrease in plasma potassium concentration* in potassium depletion lowers aldosterone secretion.

Note. It is important to note that an increase in dietary potassium from 40 to 200 mEq/day increases plasma aldosterone levels to 6 fold.

3. Role of ACTH

ACTH also plays following roles in mineralocorticoid secretion:

- The direct stimulating effect of ACTH on aldosterone secretion is mild and transient.
- ACTH also stimulates secretion of deoxycortisone (18-OH-DC) from zona fasciculata, which have very mild mineralocorticoid activity.

ADRENAL SEX STEROIDS

Adrenal sex steroids include:
- *Dehydroepiandrosterone* (DHEA), its sulphate ester (DHEA-S) and *androstenedione* are the major androgenic precursor products of adrenal cortex.
- *Oestrogen and progesterone* are produced in very small amounts.

SYNTHESIS

The adrenal sex steroid precursors are synthesized in the *zona reticularis.*

The 17-hydroxylated derivatives of pregnenolone and progesterone are the starting points for synthesis of androgen precursors. The circumstances that lead to impairment of cortisol synthesis at any point beyond this step, cause accumulation of 17-hydroxy-pregnenolone and 17-hydroxy-progesterone leading to greatly increased androgen synthesis. Figure 8.5-17 depicts the steps of androgen precursor synthesis.

PLASMA LEVELS AND CONTRIBUTION TOWARDS SEX STEROIDS, METABOLISM AND EXCRETION

Plasma levels. Normal plasma level of DHEA is 150–200 µg/dL at 25 years of age in both sexes.

Contribution of adrenal glands towards sex steroids and their functions are:

During fetal life, adrenal cortex is hyperplastic and secretes a large amount of DHEA, which acts as the main precursor for synthesis of oestrogen by placenta.

In adult women, the adrenal glands supply 50–60% of the androgenic hormone requirement. DHEA-S contributes to increased muscle mass, growth of pubic and axillary hair and libido.

In adult man, since testes produce a large quantity of testosterone, the adrenal androgen precursors are of little biological importance. However, they may be partly responsible for the development of male sex organs in childhood.

Metabolism and excretion

- Androsterone and etiocholanolone, the two isomers that are formed as end result of metabolism, are excreted in the urine. These metabolites are not specific for the adrenal gland, as they also arise from the gonadal androgens.
- DHEA-S is entirely excreted directly in the urine and it is virtually adrenal specific.
- Androsterone, etiocholanolone and DHEA-S are together known as *17-ketosteroids* and constitute the major part of a urinary fraction. Their normal values range from 5 to 14 ng/day in women and 8–20 ng/day in men.
- Normally, two-thirds of the urinary 17-ketosteroids are derived from the adrenal secretions and one-third from gonadal androgen secretions.

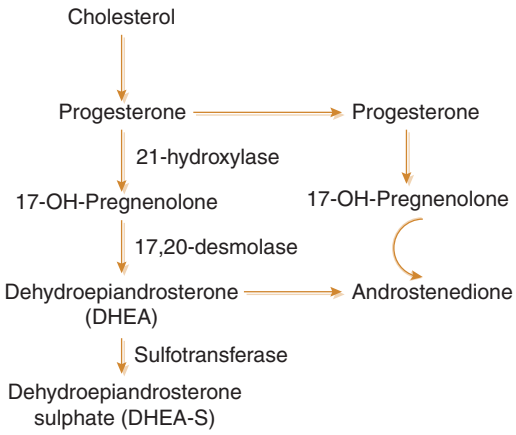


Fig. 8.5-17 Steps involved in the synthesis of androgen precursors in the zona reticularis.

APPLIED ASPECTS

The important applied aspects in relation to the adrenal cortex which need mention include:

- An integrated response to stress (see page 599),
- Hyperactivity of adrenal cortex and
- Hypoactivity of adrenal cortex (see page 595).

HYPERACTIVITY OF ADRENAL CORTEX

Disorders of hyperactivity of adrenal cortex include:

- Cushing's syndrome (hypercortisol state),
- Conn's syndrome (hyperaldosteronism) and
- Adrenogenital syndrome (excessive secretion of adrenal androgens).

1. CUSHING'S SYNDROME

Cushing's syndrome refers to the group of clinical conditions occurring due to prolonged excessive levels of glucocorticoids.

Causes

Causes of Cushing's syndrome can be divided into two groups:

I. ACTH-dependent Cushing's syndrome is more common (80% cases and occurs due to hyperplasia of adrenal cortex—secreting excessive glucocorticoids) caused by excess of ACTH.

- *Hyperactivity of pituitary* as seen in tumours of pituitary cells particularly of basophils which secrete ACTH. The resulting condition of pituitary origin is also called Cushing's disease.
- *Ectopic ACTH production* as seen in benign and malignant non-endocrine tumours, e.g. cancer of lungs or abdominal viscera.
- *Excessive ACTH secretion in hypothalamic disorders* associated with excess of CRH secretion.
- *Excessive ACTH therapy* (iatrogenic).

II. ACTH-independent Cushing's syndrome is less common (20% cases) and occurs in following conditions:

- *Adrenal origin Cushing's syndrome* is caused by glucocorticoid secreting tumours, such as adrenal adenoma and adrenal carcinoma.
- *Excessive glucocorticoid administration* (i.e. iatrogenic).

Characteristic features

Characteristics features of Cushing's syndrome are (Fig. 8.5-18):

1. Truncal or centripetal obesity. It occurs due to redistribution of body fat from extremities (which is in the abdominal wall, back and face) producing following characteristic features:

- *Buffalo hump,* due to collection of fat at upper back,
- *Moon face,* due to fat collection on the face,
- *Purple striae or cutaneous abdominal striae or livid stretch marks.* The skin and subcutaneous tissue becomes thin due to protein catabolism. The stretching of abdominal skin due to excessive subcutaneous fat deposition causes rupture of subdermal tissues producing reddish purple striae.

2. Muscle weakness and backache due to protein catabolism.

3. Sodium and water retention may cause weight gain, oedema and hypertension.

4. Hyperglycaemia occurs due to gluconeogenesis and inhibition of peripheral utilization of glucose. It may lead to glycosuria and adrenal diabetes.

5. Hirsutism and menstrual irregularity may occur due to increased adrenal androgens.

6. Susceptibility to osteoporosis and bone fracture is increased due to protein depletion and bone resorption.

7. Susceptibility to infections is increased due to immunosuppression.

8. Psychological, emotional and personality changes may occur due to CNS effects of glucocorticoids.

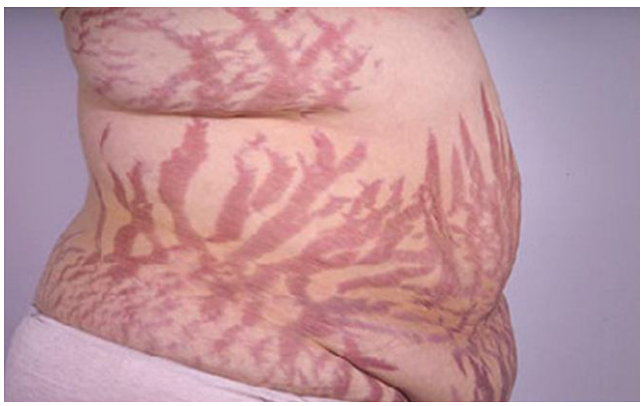


Fig. 8.5-18 Photograph of a patient with Cushing's syndrome showing truncal obesity and purple striae on the abdomen.

9. Blackening of skin may occur due to pigmentation caused by MSH-like effects of excessive ACTH.

10. Susceptibility to peptic ulceration is increased.

Tests for Cushing's syndrome

A. Tests to confirm diagnosis of Cushing's syndrome

- *Plasma cortisol level* is raised and there is loss of diurnal pattern, i.e. circadian rhythm is lost.
- *Dexamethasone suppression test* is not able to suppress plasma cortisol level.
- *Insulin-induced hypoglycaemia* which raises plasma cortisol in normal persons fails to raise it in Cushing's syndrome.
- *24-Hour urinary-free cortisol* levels are raised.

B. Tests to differentiate between ACTH-dependent and ACTH-independent Cushing's syndrome are shown in Table 8.5-3.

2. HYPERALDOSTERONISM

Hyperaldosteronism refers to over production of the hormone aldosterone, a major sodium-retaining hormone. Hyperaldosteronism may be:

1. Primary hyperaldosteronism or *Conn's disease* occurs due to tumour or hyperplasia of zona glomerulosa of adrenal cortex.

2. Secondary hyperaldosteronism occurs due to some extra adrenal cause, which stimulates renin–angiotensin–aldosterone system, e.g. nephrotic syndrome, cirrhosis of liver, congestive heart failure and toxaemia of pregnancy.

Characteristic features of hyperaldosteronism are:

- *Sodium and water retention* leading to hypertension and oedema. It is important to note that marked hypernatraemia and oedema do not occur because Na^+ excretion is soon normalized despite hypersecretion of aldosterone *(escape phenomenon,* see page 591).

Table 8.5-3 Tests to differentiate between ACTH-dependent and ACTH-independent Cushing's syndrome

Tests	ACTH dependent (Pituitary causes)	ACTH independent (Adrenal causes)
Plasma ACTH level at 8 AM	Increased	Undetectable
ACTH level following CRH stimulation	Increased	No change
Metyrapone test. Levels of 11-deoxycortisol after 24 h of administration of metyrapone	Decreased	No change

- *Hypokalaemia* may occur due to an increased potassium excretion producing muscle weakness.
- *Metabolic alkalosis* may occur due to secretion of more amount of H^+ into renal tubules. Metabolic alkalosis may produce hypocalcaemia causing tetany.

3. ADRENOGENITAL SYNDROME

As mentioned earlier, the androgen precursors secreted by the adrenal cortex are of little biological importance under normal circumstances. However, when secreted in large amounts as in *tumour of zona reticularis* of adrenal cortex, the following abnormal features may be produced:

- *In prepubertal males,* the excessive androgens produce precocious pseudopuberty.
- *In males,* the oestrogen producing cells may produce female-like secondary sexual characters, such as enlargement of breasts (gynaecomastia), atrophy of testes, loss of libido and feminine body.
- *In females,* they cause development of male secondary sexual characteristics, such as beard muscular body, breaking of voice, male type hair growth, enlargement of clitoris and amenorrhoea.

Note

- In virilized children or adult women, a large increase in urinary 17-ketosteroid excretion almost always indicates an adrenal abnormality.

HYPOACTIVITY OF ADRENAL CORTEX

Adrenocortical deficiency, depending upon the site of lesion, can be divided into two types:

I. Primary adrenocortical deficiency occurs due to involvement of adrenal cortex and is associated with high ACTH levels due to feedback mechanism. The conditions producing primary adrenocortical deficiency include:

- Addison's disease, and
- Congenital adrenal hyperplasia.

II. Secondary adrenocortical deficiency occurs due to involvement of either hypothalamus or pituitary or due to exogenous glucocorticoid administration and is associated with *low ACTH level* due to less production.

ADDISON'S DISEASE

Addison's disease occurs due to chronic deficiency of hormones secreted by all the three zones of adrenal cortex. Therefore:

1. *Glucocorticoid insufficiency* produces weight loss, malaise, anorexia, nausea, vomiting, weakness and diarrhoea. Since glucocorticoids are essential for adaptation to stress, therefore in Addison's disease exposure to any type of stress, e.g. even mild infection may be fatal.
2. *Mineralocorticoid deficiency* produces hyponatraemia, hyperkalaemia, acidosis and decreased ECF volume with hypotension.
3. *Loss of androgens* causes sparse hair in females.
4. *Increased ACTH secretion* occurs due to feedback mechanism and causes diffuse pigmentation of the skin and mucous membranes (because of its MSH like actions).

Addisonian crisis or adrenal crisis

It refers to acute adrenal insufficiency characterized by sudden collapse. The condition becomes fatal if not treated in time.

CONGENITAL ADRENAL HYPERPLASIA

Causes. Congenital adrenal hyperplasia is caused by congenital deficiency of *21-hydroxylase deficiency* and deficiency of 11-hydroxylase enzymes.

Characteristic features are virilism and excessive body growth.

In boys, it is characterized by:

- Precocious body growth leading to stocky appearance called *infant hercules.*
- Precocious sexual development with enlarged penis even at age of 4 years.

In female fetus, high plasma androgen levels cause *masculinized pattern* of development *(virilism).* Sometime the female fetus may be born with male type external genitalia. This condition is called *pseudo-hermaphroditism.*

HORMONES OF ADRENAL MEDULLA

The adrenal medulla secretes *catecholamines* which include epinephrine, norepinephrine and dopamine. About 80% of adrenal medullary catecholamine is epinephrine and rest is norepinephrine. Apart from catecholamines, the adrenal medulla also contains small amounts of dynorphins, neurotensin, encephalin, somatostatin and substance P. The functions of these adrenal peptides are not clear.

SYNTHESIS AND STORAGE OF CATECHOLAMINE HORMONES

Synthesis of catecholamines

Epinephrine and norepinephrine are synthesized in different cells. The biosynthetic pathway originates with

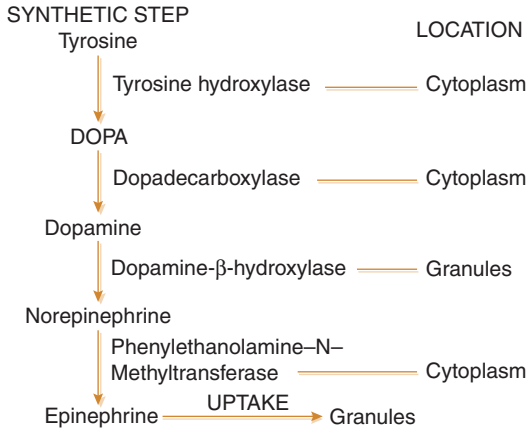


Fig. 8.5-19 Steps of catecholamine synthesis in the adrenal medulla.

L-tyrosine. Steps of catecholamines have been summarized in Fig. 8.5-19.

Storage of catecholamines in storage granules

The epinephrine formed in the cytoplasm is then taken back up by the chromaffin granules, in which it is stored as the predominant adrenomedullary hormone.

SECRETION OF CATECHOLAMINES

Nervous control of secretion

The catecholamine secretion is entirely controlled by the splanchnic nerves supplying the medulla. These nerves comprise pre-ganglionic sympathetic fibres emerging mainly from lower thoracic segments (T_5–T_9) of ipsilateral intermediolateral grey column of the spinal cord. These fibres, when stimulated, act by releasing acetylcholine close to the adrenal medullary chromaffin cells.

Physiological and psychological stimuli for release

As mentioned in the beginning, the adrenal medullary activation occurs as a part of generalized sympathetic response to any emergency situation. Therefore, this has also been called *sympathetic alarm* reaction. The various sensory stimuli associated with the rapid release of epinephrine (and probably norepinephrine) from adrenal medulla include (Fig. 8.5-20):

- Perception or even anticipation of danger or harm (anxiety),
- Pain, trauma,
- Hypovolaemia from haemorrhage or fluid loss,
- Hypotension,
- Anoxia,
- Exposure to extremes of temperature,
- Hypoglycaemia and
- Severe exercise

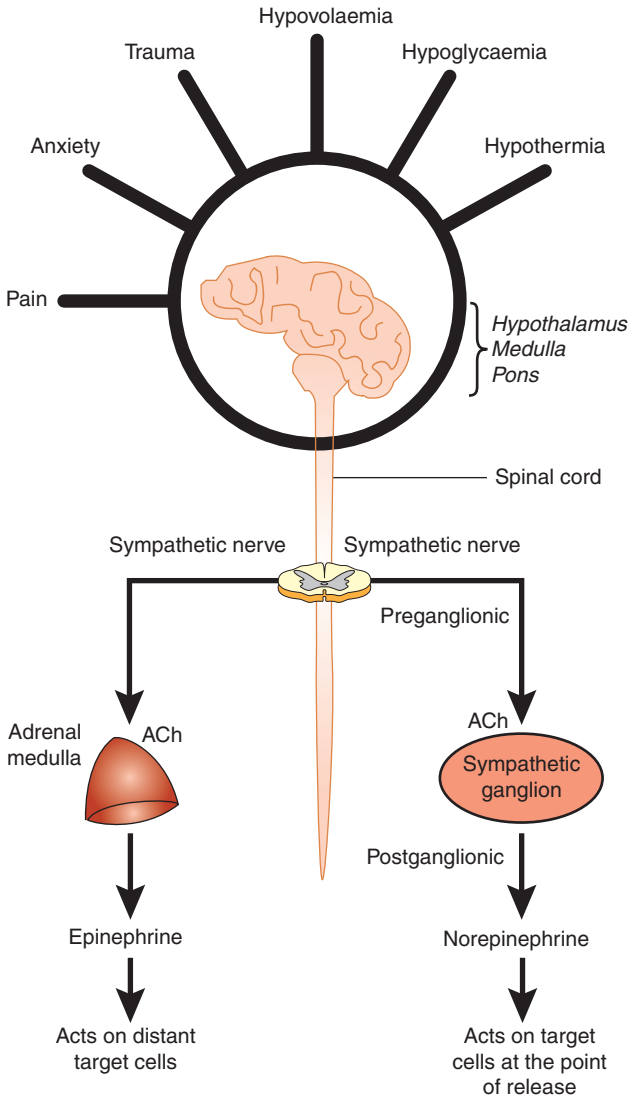


Fig. 8.5-20 Stimuli associated with secretion of catecholamines from adrenal medulla and sympathetic nervous system. ***Note.*** Adrenal medulla releases primarily epinephrine into the blood stream where it acts on distant targets. The sympathetic ganglia release norepinephrine into the synaptic cleft which act on the target cell at point of release.

Selective secretion of catecholamines in response to specific stimuli

In humans, epinephrine and norepinephrine appear to be released independently by specific stimuli:

- *Anger and aggressive states* are associated with increased norepinephrine secretion.
- *States of anxiety*, tense but passive emotional displays are associated with increased epinephrine secretion.

Stimulation of adrenal medulla also occurs independent of sympathetic system, e.g. *hypoglycaemia* activates adrenal medulla producing a marked increase in catecholamine secretion without any significant increase in sympathetic neural discharge.

CIRCULATION, METABOLISM AND EXCRETION

Circulation

- *Secreted epinephrine* and norepinephrine from the adrenal medulla is in the ratio of 4:1.
- *Basal plasma levels* (in recumbent humans) of free epinephrine are 30 pg/mL and that of free norepinephrine are 300 pg/mL.
- *Variation in plasma levels* of catecholamines according to physiological or pathological states are quite common.

However, the threshold levels at which circulating norepinephrine can produce physiological effect is about 6 times its basal levels. While the threshold level at which epinephrine produces its effects are well achieved during that physiological state. Hence, in most of the physiological and pathological conditions, increased adrenal medullary secretion results in selective epinephrine mediated effects, in spite of increased secretion of both catecholamines.

Metabolism and inactivation of circulating catecholamines

Plasma half-life of epinephrine (E) and norepinephrine (NE) is extremely short (1–3 min).

Inactivation of catecholamines released by the sympathetic nerve endings at the synaptic clefts differs from that of the catecholamines released into circulation by the adrenal medulla.

Catecholamines released at sympathetic neuroeffector junction are inactivated by:

- Active neuronal reuptake into the presynaptic nerve terminals is the most important mechanism of termination of action of NE in the junctional space.
- Dilution by diffusion out of the junctional cleft.

Circulating catecholamines (epinephrine and norepinephrine) are metabolized predominantly in the liver and kidney by the enzymes, monoamine-oxidase (MAO) and catechole-O-methyltransferase (COMT).

Steps in the metabolic disposition of catecholamines are (Fig. 8.5-21):

- Both NE and E are first oxidatively deaminated by combined action of MAO and AO (aldehyde oxidase to dihydroxymandelic acid.
- dihydroxymandelic acid is then 0-methylated by the enzyme COMT to vanillylmandelic acid (VMA).
- Alternatively, the norepinephrine and epinephrine can be first 0-methylated by COMT to produce normetanephrine and metanephrine, respectively.
- The normetanephrine and metanephrine are then oxidatively deaminated by the combined action of MAO and AO to methoxyhydroxyphenylglycol (MOPG) and then to VMA.

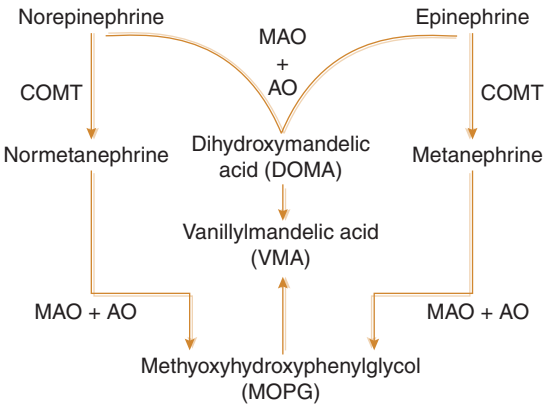


Fig. 8.5-21 Steps in the metabolic disposition of catecholamines.

The metabolites are excreted in the urine and bile as VMA and MOPG (Fig. 8.5-21).

ADRENERGIC RECEPTORS AND ACTIONS OF CATECHOLAMINES

Adrenergic receptors

The adrenergic receptors are of two types:

Alpha (α) receptors. These are further of two types (α_1 and α_2). The alpha-adrenergic receptors are sensitive to both epinephrine and norepinephrine. These receptors are associated with most of the excitatory functions of the body but have one major inhibitory function (i.e. inhibition of intestinal motility).

Beta (β) receptors. These are further of three types: β_1, β_2 and β_3. Beta-adrenergic receptors respond to epinephrine and in general are relatively insensitive to norepinephrine. These receptors are associated with most of the inhibitory functions of the body but have an important excitatory function (i.e. excitation of myocardium).

For details of adrenergic receptors see page 767.

Mechanism of action and second messenger involved are:

- α_1 receptors are coupled to the phosphatidylinositol membrane system; calcium along with protein kinase C mediates the hormone effects.
- α_2 receptors are coupled to an inhibitory G-protein; thus hormone binding decreases cAMP levels and protein kinase A activity.
- β_1, β_2 and β_3 receptors are coupled to and stimulate adenylyl cyclase; thus cAMP is the second messenger for their biological effects.

ACTIONS OF CATECHOLAMINES

I. Metabolic actions of catecholamines

Epinephrine affects metabolic functions more than norepinephrine, via alpha and beta receptors.

1. General metabolic effects of epinephrine. This includes:

- Increased O_2 consumption (by 20–40%) and increased CO_2 output.
- Raised basal metabolic rate and respiratory quotient.
- Increased heat production due to stimulation of cellular oxidative processes.

2. Effect on carbohydrate metabolism. Epinephrine produces hyperglycaemia and makes the glucose available for the brain and other tissues to meet the emergency by its following effects:

- *Glycogenolysis* is stimulated in the liver.
- *Glycogenesis* is reduced in the liver by inhibition of the enzyme glycogen synthase.
- *Gluconeogenesis,* i.e. hepatic production of glucose from lactate, amino acids and glycerol is increased.
- *Insulin secretion* is inhibited.
- *Glucagon secretion* is stimulated. This amplifies the hyperglycaemic effects of epinephrine.
- *ACTH secretion* is stimulated, which then stimulates cortisol secretion. Cortisol is a potent gluconeogenic hormone.

3. Effects on fat metabolism. Norepinephrine has more potent action on the lipid metabolism than the epinephrine which has a predominant effect on the carbohydrate metabolism.

- *Catecholamines* cause an increase in lipolysis by stimulating hormone-sensitive lipase (via beta receptor, i.e. cAMP) in adipose tissue and muscles. This results in an increase in free fatty acids in the circulation, which are effectively utilized by the heart and muscle as fuel source.

II. Physiological actions of catecholamines

1. Effects on cardiovascular system

The net effects of epinephrine and norepinephrine are (Table 8.5-4):

Epinephrine

- Increases heart rate and force of contraction via β_1 receptors resulting in an increased cardiac output and rise in systolic blood pressure (SBP).
- Causes vasoconstriction in renal, splanchnic and cutaneous vascular bed.

Table 8.5-4 Cardiovascular effects of catecholamine

Parameter	Epinephrine	Norepinephrine
• Heart rate	↑	↓
• Cardiac output	↑	↓
• Peripheral resistance	↓	↑
• Systolic blood pressure	↑	↑
• Diastolic blood pressure	↓	↑
• Mean arterial pressure	↓ or N	↑

↑ = Increase; ↓ = decrease; N = normal.

Norepinephrine

- Increases heart rate and force of contraction via β_1 receptors resulting in an increased cardiac output and rise in SBP.
- Causes vasoconstriction via α_1 receptors resulting in an increased peripheral resistance and increased diastolic blood pressure (DBP).
- As a result of increased SBP and DBP, mean blood pressure is markedly increased, which reflexly by stimulation of baroreceptors (aortic and carotid sinus) decrease heart rate, force of contraction and cardiac output.
- The net result of norepinephrine effect is decreased heart rate, decreased cardiac output, increased peripheral resistance and increased mean blood pressure. Because of this reason, norepinephrine and not epinephrine is useful in patients with shock.

2. Effects on other systems

On CNS, catecholamines via β receptors activate reticular activating system (RAS) by lowering its threshold and thus lead to arousal and alerting responses producing anxiety, apprehension and coarse tremors of extremities.

On GIT, epinephrine via β receptors causes relaxation of smooth muscles of wall of the gut decreasing its tone and motility. Via α receptors epinephrine causes contraction of sphincters of gut, the net result is production of constipation.

On urinary bladder. Epinephrine produces retention of urine by relaxing the detrusor muscles.

On skin. Catecholamines act via α receptors on pilomotor muscle producing piloerection of hair by acting on the sweat glands of palm and sole produce localized sweating called *adrenergic sweating* (e.g. generalized sweating which is cholinergic).

On skeletal muscle. During exercise, epinephrine via β_2 receptors increases blood supply (by causing vasodilation). It also increases glycogenolysis in muscle and releases glucose into circulation.

On eyes. Epinephrine causes dilation of the pupil (mydriasis) by contracting dilator pupillae (radial) muscle and via β receptors causes relaxation of the ciliary muscle producing flattening of the lens. These effects provide better far vision benefit to the endangered individual.

On respiration. Epinephrine via β_2 receptors relaxes smooth muscles of bronchioles producing bronchodilation. It also increases rate and force of respiration.

On blood. Epinephrine produces following effects:

Reduces blood coagulation time by increasing activity of factor V.

- *Increases RBC count,* haemoglobin content and packed cell volume (PCV) due to release of RBCs in circulation by causing contraction of spleen.

- *Increases plasma protein concentration* by movement of fluid out of circulation.
- *Neutrophilia* occurs due to release of sequestrated neutrophils into the circulation.

On secretion of other hormones. Catecholamines regulate secretion of a number of hormones:

- *Insulin and somatostatin* secretion is decreased via α_2 receptors (by decreasing cAMP).
- *Glucagon and pancreatic peptide* secretion is increased via β_2 receptors.
- *Thyrotropin-releasing hormone (TRH)-induced secretion of TSH* from thyrotrophs is decreased via α_2 receptors.
- *Thyroid hormone* secretion is enhanced by catecholamines under certain circumstances and peripheral conversion of T_4–T_3 is stimulated via β_2 receptors.

On renin secretion and Na^+ and K^+ movement. Catecholamines increase renin secretion by stimulation of β receptors in the kidney. The increase in renin in turn increases aldosterone secretion, which in turn enhances *sodium retention.*

III. Role of sympathoadrenal system in various physiological states

1. Role during exercise. During mild to moderate exercise mainly sympathetic nervous system is activated. However, during severe exercise the adrenal medullary secretion is also increased. For details see page 371.

2. Role during exposure to cold. Sympathoadrenal system is essential for maintenance of body temperature during exposure to cold. The epinephrine maintains body temperature by conserving body heat as well as by producing heat. For details see page 959.

3. Role during hypoglycaemia. Hypoglycaemia is a very potent stimulator of epinephrine secretion from the adrenal medulla, while it does not increase sympathetic neural activity to any significant degree. Plasma epinephrine levels may rise 10–50 fold depending upon the severity of hypoglycaemia. By its effects on metabolism described above (page 597), the epinephrine restores plasma glucose levels and glucose delivery to the central nervous system.

AN INTEGRATED RESPONSE TO STRESS

Stress, may it be emotional, physical or biological, evokes an integrated response of sympathoadrenal medullary system and hypothalamic–pituitary–adrenal cortex axis.

Steps involved in stress adaptation by an integrated response of the above system are (Fig. 8.5-22):

Perception of stress signals. Stress is perceived by many areas of the brain, from the cortex down to brainstem including limbic system and RAS.

Stimulation of hypothalamus. Major stresses activate the CRH and ADH neurons in the paraventricular nucleus and adrenergic neurons.

Activation of hypothalamic–pituitary–adrenal axis. Corticotropin-releasing hormone and antidiuretic hormone. (ADH) release stimulates ACTH release and ultimately elevates plasma cortisol levels.

Activation of sympathoadrenal medullary system. Sudden exposure to any type of stress initially produces the sympathetic alarm reaction. Stimulation of adrenergic neurons of hypothalamus ultimately leads to a release of epinephrine from adrenal medulla and norepinephrine from the sympathetic ganglia.

Integrated role of hormones released by hypothalamic–pituitary–adrenal axis and sympathoadrenal medullary system in stress adaptation. Together these hormones help in adaptation to stress by their following actions (flight or fight):

- *Increase in glucose production.* Catecholamines rapidly raise plasma glucose by activating glycogenolysis, and cortisol acts more slowly by providing amino acid substrate for gluconeogenesis. Together they shift glucose utilization towards the central nervous system away from the peripheral tissues.
- *Free fatty acid supply.* Epinephrine rapidly augments the supply of free fatty acids to heart and to the muscles, and cortisol facilitates the lipolytic role.
- *Cardiovascular adjustments.* Catecholamines and cortisol raise blood pressure and cardiac output, and they improve the delivery of substrates to tissues that are critical to the immediate defence of the organism.
- *Arousal, defensively useful behavioral activation and focused attention* result from the adrenergic stimuli to the pertinent brain centres.
- *Inhibition of activities that are not useful during stress and divert individuals, and their resources from defensive responses to danger* is an important part of adaptation to stress. For example, CRH input to the hypothalamic neurons inhibits growth hormone, gonadotropin release and sexual activity.
- *Interaction with immune system.* The hormones produced during stress interact with the immune system to produce a balance between useful local cytokine production at threatened sites and potentially dangerous systemic effects of these immune system products.

DISEASES OF ADRENAL MEDULLA

PHAEOCHROMOCYTOMA

Phaeochromocytoma is a rare benign tumour arising from the epinephrine and norepinephrine-secreting chromaffin cells of adrenal medulla.

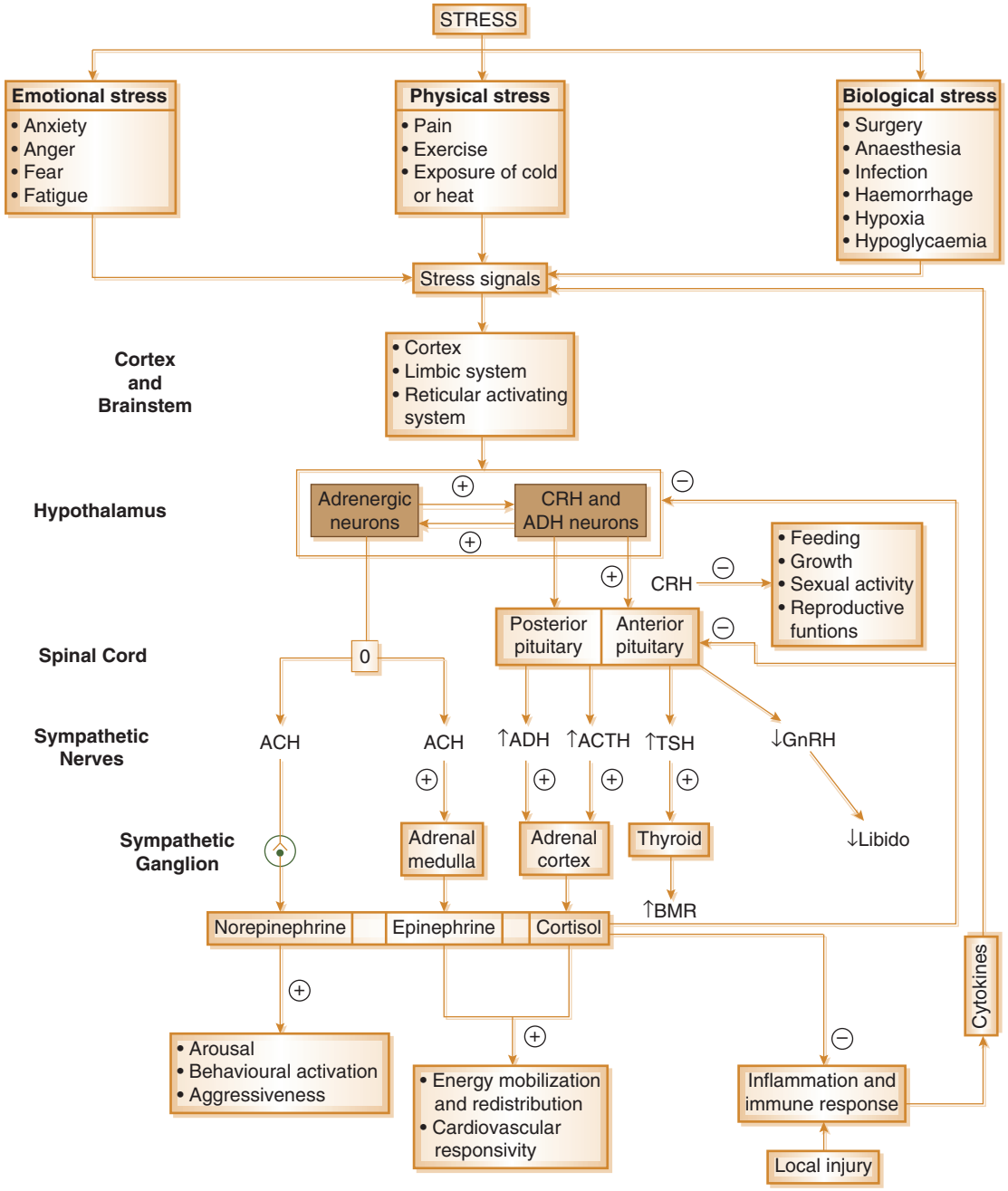


Fig. 8.5-22 Steps involved in the adaptation to stress by an integrated response of hypothalamic–pituitary–adrenal cortex axis and sympathoadrenal medullary system.

Clinical features are produced by the excess of epinephrine and norepinephrine and include:

- Episodic or non-episodic hypertension with postural drop,
- Attacks of tachycardia, palpitation, sweating, pallor, headache and chest discomfort,
- Abdominal pain, vomiting, constipation and glucose intolerance and
- Weight loss and weakness.

Tests for phaeochromocytoma include:

- *Twenty-four-hour urinary excretion* of VMA, metanephrines and catecholamines is increased.
- *Plasma epinephrine* and norepinephrine levels are elevated.
- *Phentolamine suppression test,* i.e. 2.5 mg of phentolamine does not suppress plasma catecholamine at 10 min sample.
- *CT scan and radionuclide studies* to localize any tumour.

IMPORTANT NOTE

Intravenous injection of large dose of epinephrine produces similar type of symptoms as in pheochromocytoma.

Chapter 8.6

Pancreatic and Gastrointestinal Hormones

ENDOCRINE PANCREAS

FUNCTIONAL ANATOMY

The endocrine part of the pancreas comprises numerous rounded collections of cells known *as pancreatic islets* or the *islets of Langerhans.* These are embedded within the exocrine part, and they constitute 1–1.5% of the human pancreatic mass.

Islets of Langerhans

Each islet contains four types of cells (Fig. 8.6-1):

- *Beta* (β) cells, make up 60–70% of the total cells and constitute the central core of the islet. These cells secrete insulin.
- *Alpha* (α) cells form about 20% of the total cells and constitute the outer rim of the islet. These cells secrete glucagon.
- *Delta* (δ) cells form about 10% of total cells and are intermixed. These are source of somatostatin.
- *PP cells.* These are also peripherally placed scattered amongst the α cells. These are source of pancreatic peptide.

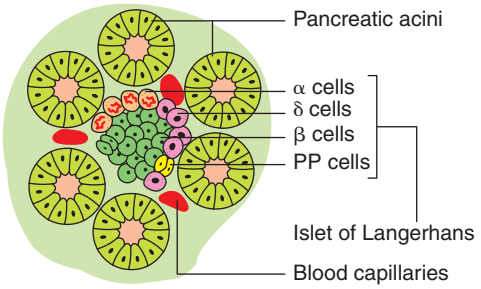


Fig. 8.6-1 Schematic histological structure of pancreas showing an islet of Langerhans surrounded by exocrine pancreatic acini.

Innervation. Pancreatic islet cells are innervated by parasympathetic and sympathetic fibres, which influence the secretory activity of α and β cells of islets.

Vascular arrangement. Small arterioles enter the core of each islet and break up into a network of capillaries with fenestrated endothelium. These capillaries then converge into venules, which carry blood to the mantle of the islet. This portal arrangement allows high concentration of insulin from β cells core so bathe the α, δ and PP cells of the respective mantles. This type of vascular pattern also suggests the possible paracrine effects of insulin on the outer islet cell types.

The interrelationship among cells of islets of Langerhans is shown in Fig. 8.6-2.

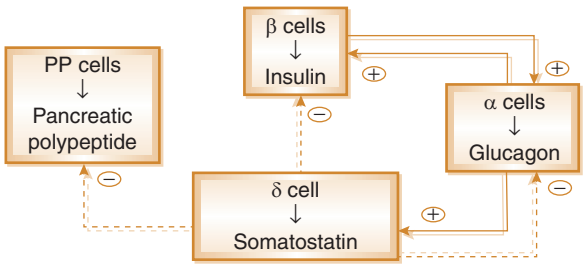


Fig. 8.6-2 Interrelationship among cells of islets of Langerhans of pancreas.

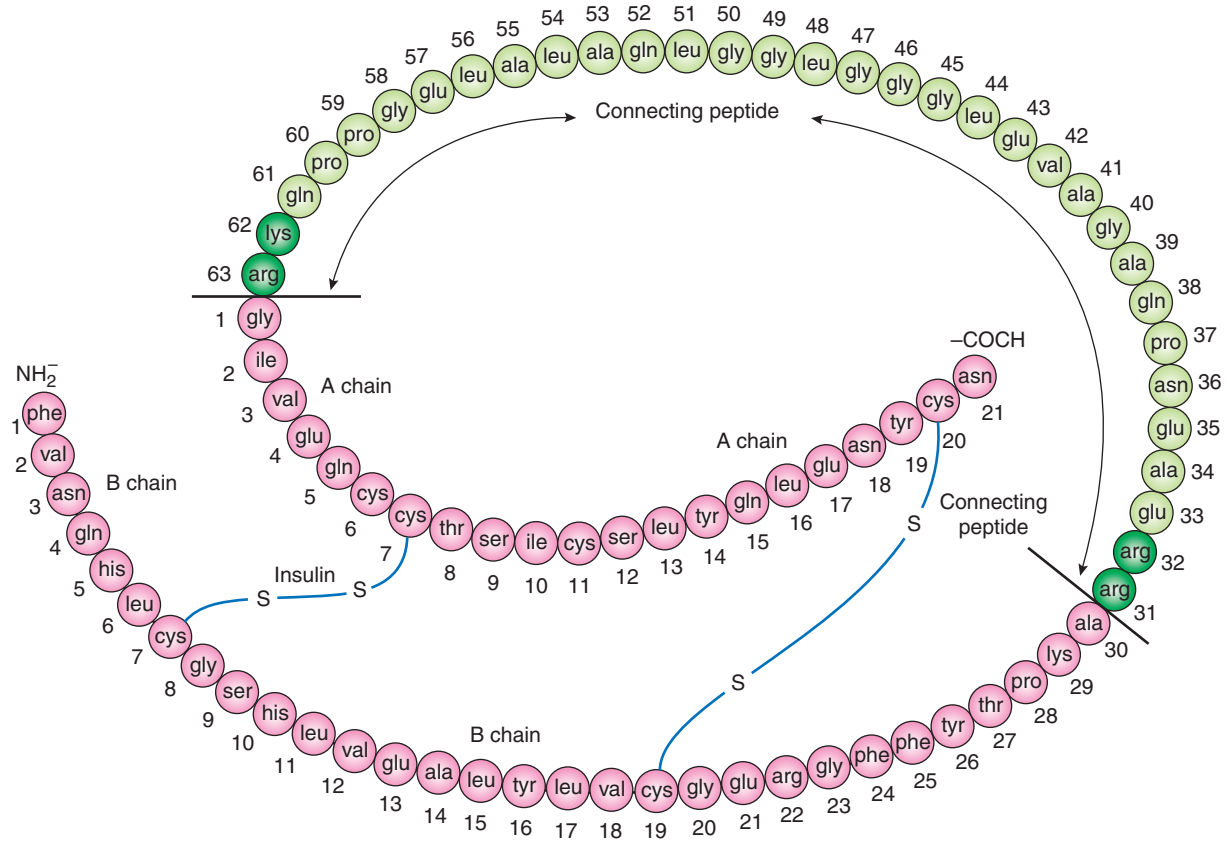


Fig. 8.6-3 Structure of an insulin molecule.

Gap junctions link beta cells to each other, alpha cells to each other and beta cells to alpha cells for rapid communication.

INSULIN

Insulin is a polypeptide hormone secreted by the β cells of islets of Langerhans of pancreas. Historically, insulin is the first hormone to be isolated, purified, crystallized and synthesized.

STRUCTURE AND BIOSYNTHESIS

Structure. The human insulin is protein containing 51 amino acids, arranged in two polypeptide chains: A (having 21 amino acids) and B (having 30 amino acids). These chains are connected to each other by two interchain disulphide linkages, connecting A_7 to B_7 and A_{20} to B_{19}. In addition, there is an intrachain disulphide link in chain (Fig. 8.6-3).

Biosynthesis of insulin. Beta cells of islets of Langerhans synthesize insulin by usual protein synthetic machinery. The steps involved are (Fig. 8.6-4):

- The insulin gene (located on chromosome 11) directs the synthesis of *preproinsulin*, an insulin precursor consisting of 108 amino acids with a molecular weight of 11,500.
- Preproinsulin is cleaved to form proinsulin having 86 amino acids and a molecular weight of 9000.

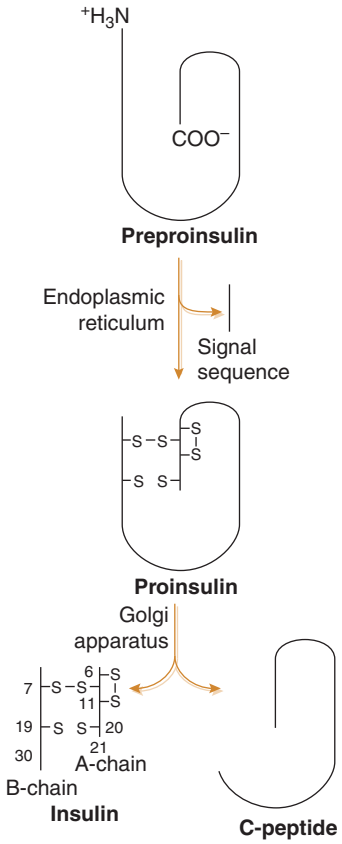


Fig. 8.6-4 Steps in the synthesis of insulin.

- As the *proinsulin molecule,* containing the A and B chain of insulin and connecting peptide (C-peptide), is guided to Golgi apparatus, disulphide linkages are established to yield the *folded proinsulin* molecule.
- Proinsulin is further cleaved in Golgi apparatus to form the active *hormone insulin* and a connecting peptide (C-peptide).
- Insulin becomes associated with zinc as the secretory granules mature.

C-peptide has no biological activity; however, its estimation in the plasma serves as an useful index for the endogenous production of insulin.

REGULATION OF INSULIN SECRETION

I. Role of exogenous nutrients

Exogenous nutrients (glucose, amino acids, free fatty acids, ketoacids and potassium) control insulin secretion by a feedback mechanism (Fig. 8.6.5). When substrate supply is abundant, insulin is secreted in response. Insulin then stimulates the use of these incoming nutrients and simultaneously inhibits the mobilization of analogous endogenous substrates. When nutrients supply is low or absent, insulin secretion is dampened and mobilization of endogenous fuels is enhanced.

II. Role of blood glucose

Insulin secretion is mainly controlled by level of blood glucose by feedback relationship.

The relationship between plasma glucose and plasma insulin is sigmoidal (Fig. 8.6-6). As shown in Fig. 8.6-6:

- Below 50 mg/dL levels of plasma glucose, virtually no insulin is secreted,
- Above 100 mg/dL levels of plasma glucose, rate of insulin secretion rises rapidly,
- At about 150 mg/dL levels of plasma glucose, a half-maximal insulin secretory response is obtained and
- At a level of about 300 mg/dL, a maximal insulin response occurs.

The rapidly increased insulin secretion above 100 mg/dL levels of plasma glucose, in turn reduces blood glucose concentration to fasting level. Reduction of blood glucose causes rapid turning off of insulin secretion (feedback relationship).

Biphasic response of insulin secretion occurs in response to continuous glucose stimulation (Fig. 8.6-7):

- *An immediate pulse of insulin* is released, within seconds of exposure to glucose that peaks at 1 min (about 10 fold rise) and then returns towards baseline in another 5–10 min.
- *A second phase of insulin secretion* begins after about 10 min of continuous stimulation. During this phase, plasma levels of insulin rise more slowly and reach a second plateau, which can be maintained for many hours in normal individuals.

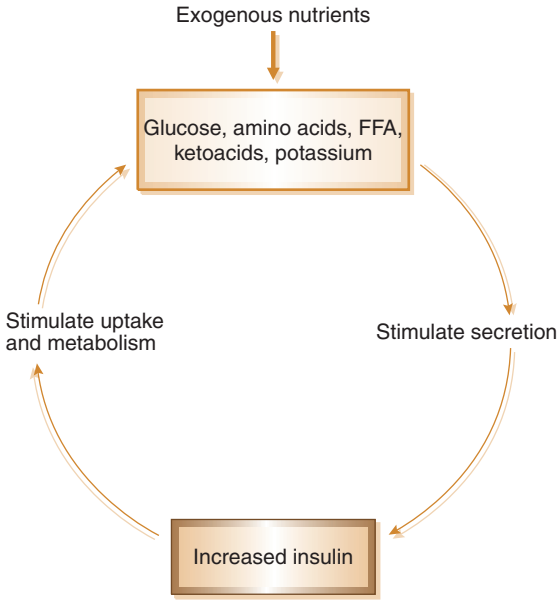


Fig. 8.6-5 Feedback control of insulin release by exogenous nutrients.

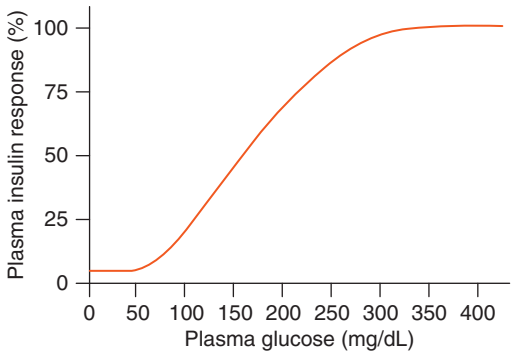


Fig. 8.6-6 A sigmoidal relationship between levels of plasma glucose and insulin secretion response.

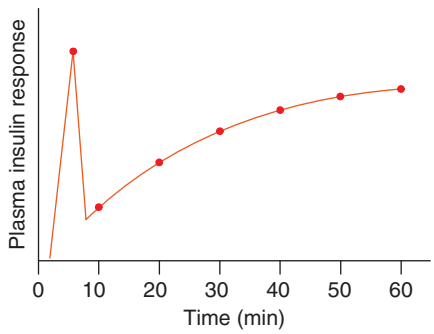


Fig. 8.6-7 Biphasic insulin response to glucose infusion consists of first phase of rapid release and fall followed by second phase of slow rise.

Factors responsible for biphasic response are:

- Initial peak occurs due to release of preformed insulin.
- Second phase occurs due to glucose stimulation of insulin synthesis that sustain the secretory phase.

When glucose is given orally, a greater insulin response is elicited than when plasma glucose is elevated comparably by an intravenous administration. This augmented response to oral glucose is attributed to the gastrointestinal (GIT) hormones like gastrin, secretin, cholecystokinin and gastric inhibitory polypeptide, i.e. GIP (most potent), that cause moderate increase in insulin secretion. Since these hormones are released immediately after meals, so they cause anticipatory rise of insulin before actual absorption of glucose and amino acids.

III. Role of sympathetic and parasympathetic nervous system

Sympathetic nerves and epinephrine. Epinephrine is the most predominant inhibitor of insulin release. In emergency situations like stress, extreme exercise and trauma, the nervous system stimulates adrenal medulla to release epinephrine. The epinephrine suppresses insulin release and promotes energy yielding compounds—glucose from liver and fatty acids from adipose tissue.

Parasympathetic nerves to pancreas and acetylcholine (ACh) increase insulin secretion to some extent and stimulate insulin release. Some of the important factors stimulating and inhibiting insulin secretion are depicted in Table 8.6-1.

PLASMA INSULIN LEVELS, CIRCULATION AND DEGRADATION

Plasma levels of insulin

- Average basal peripheral plasma insulin level is 10 μU/mL.
- After several days of fasting, basal plasma levels of insulin decline over 50%, i.e. become less than 5 μU/mL.
- After prolonged exercise, the plasma insulin levels fall.
- A 3–10 fold increase in plasma insulin level is noted after a typical meal. The peak occurs after 30–60 min of initiating the meal.
- Total daily peripheral delivery of insulin is about 30 units.

Circulation and degradation of insulin

Insulin circulates unbound to carrier protein. *Half-life of insulin* in plasma is 5–18 min.

Table 8.6-1 Factors stimulating and inhibiting insulin secretion

Factors stimulating insulin secretion	Factors inhibiting insulin secretion
• ↑ Blood glucose • ↑ Amino acids • ↑ Fatty acids • Glucagon • GIP • ACh • Growth hormones, cortisol	• ↓ Blood glucose • Somatostatin • Norepinephrine, epinephrine

- A protease enzyme, namely *insulinase* (mainly found in the kidneys and liver), degrade insulin. It splits the disulphide bonds and separates the A and B chains.
- Very little insulin is excreted unchanged in the urine.

MECHANISM OF ACTION OF INSULIN

Insulin acts on the target tissue through insulin receptors.

Insulin receptor

Insulin receptor is a protein kinase receptor that contains enzyme activity. About 2–3 lac insulin receptors are present on the cell membrane of the target tissues for insulin.

Structure. The insulin receptor is a tetramer having two identical α subunits and two β subunits.

- subunits (chains) are located on the outer surface of the plasma membrane and contain the insulin-binding domain.
- subunits (chains) span across the plasma membrane and reside largely within the cytoplasm. These have tyrosine kinase domain.

The activity of insulin receptors ultimately produce following effects on the target cells

1. Gene expression in the nucleus of the target cell leading to biological action.

2. Translocation of glucose transport proteins to the plasma membrane. The glucose transporters are responsible for the insulin mediated uptake of glucose by the cells. As the insulin levels fall, the glucose transporters move away from the membrane to the intracellular pool for storage and recycle.

3. Activation or deactivation of numerous enzymes in glucose and fatty acid metabolism is brought about by increased mRNAs.

4. Protein synthesis. Increased mRNA synthesis (transcription) is followed by translation (protein synthesis). In this way, insulin promotes synthesis of enzymes, such as glucokinase, phosphofructokinase and pyruvate kinase.

ACTIONS OF INSULIN

A. Metabolic effects,
B. Effects on ion transport and
C. Role in cell growth and development.

A. Metabolic effects of insulin

The insulin plays a key role in the metabolism of carbohydrate, lipids and proteins. The major targets for insulin actions are the muscle mass, the liver and adipose tissue.

- Insulin is required for the uptake of glucose by muscles (skeletal, cardiac and smooth), adipose tissue, leucocytes and mammary glands.
- Tissues in which glucose transport is not insulin dependent include nervous tissue, kidney, RBC, retina, blood vessels and intestinal mucosa.

1. Effects on carbohydrate metabolism

Insulin decreases blood glucose concentration by following mechanisms:

(i) Insulin increases uptake of glucose in target cells by translocating the glucose transporter into the cell membranes.

(ii) Insulin promotes glucose utilization by:

- Glycolysis (oxidation of glucose) is increased in muscle and liver by activating enzymes phosphofructokinase and pyruvate.
- Glycogen formation from glucose in muscle and liver is promoted by activating glycogen synthetase enzyme.

(iii) Insulin decreases glucose production by inhibiting:

- Gluconeogenesis by decreasing uptake of precursor amino acids.
- Glycogenolysis by decreasing glucose-6-phosphatase levels.

2. Effects on lipid metabolism

The metabolism of both endogenous and exogenous fat is profoundly influenced by insulin in the target tissues: liver, adipose tissue and muscle

(i) Insulin increases lipogenesis

(a) Lipogenic effects on liver. As described in effects of insulin on carbohydrate metabolism, insulin promotes storage of glucose as glycogen in the liver cells. However, when glycogen concentration increases to 5–6%, glycogenesis is inhibited and the additional glucose entering is converted to fat in the liver cells.

Synthesis of cholesterol. Insulin also favours hepatic synthesis of cholesterol from acetyl-CoA.

(b) Lipogenic effects on adipose tissue. Insulin activity promotes deposition of circulating fat into adipose tissue by activating the key enzyme lipoprotein lipase in the capillary wall of adipose tissue.

(c) Lipogenic effects of insulin in muscle. Within muscle, insulin suppresses the enzyme lipoprotein lipase in inverse proportion to its stimulation of glucose uptake.

(ii) Insulin decreases lipolysis

(a) In adipose tissue, insulin profoundly inhibits *hormone*-sensitive lipase activity. By suppressing lipolysis and the release of stored fatty acids and glycerol, insulin diminishes their delivery to liver and peripheral tissues. It is noteworthy that abdominal visceral fat is less sensitive to insulin than is subcutaneous fat.

(b) In liver also the mobilization of free fatty acid (FFA) to peripheral circulation.

(c) In muscle also the insulin inhibits lipolysis of triglyceride stores.

(iii) Insulin reduces ketogenesis. Insulin is the major and perhaps the sole antiketogenic hormone. Its antiketogenic effects are:

- Insulin leads to decreased FFA flow to the liver from the adipose tissue.
- Insulin also stimulates the use of ketoacids by the peripheral tissue.

(iv) Effect of insulin on lipoprotein metabolism. It appears that insulin is required for the utilization of very-low-density lipoprotein (VLDL) and low-density lipoprotein (LDL). The levels of VLDL and LDL and consequently the concentration of cholesterol are elevated in diabetics, which has been implicated in the pathogenesis of atherosclerosis.

From the effects of insulin on carbohydrate and fat metabolism described above, it is obvious that insulin regulates use of glucose and FFA for energy production.

3. Effects of insulin on protein metabolism

Insulin is an anabolic hormone, it stimulates protein synthesis and inhibits protein degradation.

Insulin stimulates protein synthesis by following effects:

- It increases the transport of many amino acids (especially valine, leucine, isoleucin, tyrosine and phenylalanine) into the cells by increasing membrane permeability. Thus plasma amino acids are lowered.
- Insulin increases the translation of messenger RNA on the ribosomes forming new proteins. In the absence of insulin, ribosomes stop working.
- In the liver, insulin decreases rate of gluconeogenesis and thus conserves amino acids for protein synthesis.

Insulin inhibits protein metabolism. Insulin inhibits proteolysis.

B. Effects of insulin on ion transport

Insulin increases K^+, PO_4^{3-} and Mg^{2+} uptake into the skeletal muscle cell and of K^+ and PO_4^{3-} into hepatic cells from the extracellular fluid by increasing membrane permeability. Another effect of insulin on the electrolytic balance is to increase reabsorption of K^+, PO_4^{3-} and Na^+ by the tubules of the kidney.

C. Role of insulin in cell growth and development

Insulin is an important factor for growth and development with the following roles:

- *Anabolic action* of insulin is as important as growth hormone for promotion of normal growth.
- *Direct stimulatory effect on macromolecules.* Insulin also stimulates the synthesis of macromolecules in tissues such as cartilage and bone and thereby directly contributes to body growth.
- *Stimulation of other growth factors.* The genes for insulin and its receptors are related to genes that encode a variety of tissue growth factors. These growth factors include somatomedins (insulin-like growth factors 1 and 2, i.e. IGF-1 and IGF-2), epidermal growth factor, nerve growth factor and relaxin.

Note. The insulin-deprived young animal or human has a reduced lean body and bone mass, and it may be profoundly retarded in height and maturation.

GLUCAGON

STRUCTURE AND SYNTHESIS

Structure. Glucagon secreted by α-cells of islets of Langerhans is a polypeptide composed of 29 amino acids in a single chain and has a molecular weight of 3500.

Synthesis. Glucagon is synthesized from a preproglucagon precursor by islet α cells. It is actually synthesized as a proglucagon (molecular weight 9000), which on sequential degradation releases active glucagon.

PLASMA LEVELS, CIRCULATION AND DEGRADATION

- *Circulation of glucagon* in plasma is in unbound form.
- *Basal levels* of glucagon in a normal fasting individual are 100–150 pg/mL.
- *Half-life* of this hormone is 6 min (range 5–9 min).
- *Secretion rate* of glucagon is estimated to be 100–150 µg/day.
- *Degradation*—mainly occurs in the liver. It is also degraded in tissues and plasma by an aminopeptide. The kidney is the other major site of glucagon degradation. Less than 1% of glucagon filtered by the glomerulus is excreted in the urine.

MECHANISM OF ACTION OF GLUCAGON

Glucagon binds to the specific receptor on the plasma membrane of target cells and acts through the mediation of cyclic AMP as a second messenger.

ACTIONS OF GLUCAGON

As mentioned above, the actions of glucagon, in almost all respects, are exactly opposite to those of insulin. It promotes mobilization of stored nutrients, such as glucose, fatty acids and ketoacids and thus is a hormone of energy release. Its metabolic effects are as follows:

1. Effects on carbohydrate metabolism. Glucagon predominately acts on the liver and increases the blood sugar level by following actions:

Increased glycogenolysis. In the liver, glucagon exerts glycogenolytic effect through activation of enzyme *glycogen phosphorylase.*

Increased gluconeogenesis. After the glycogen in liver is exhausted, glucagon increases the rate of gluconeogenesis, i.e. formation of glucose from lactate, pyruvate, glycerol and amino acids by activating multiple enzymes involved in the gluconeogenesis especially the enzyme system converting pyruvate to phosphopyruvate (rate limiting step in gluconeogenesis).

2. Effects on lipid metabolism. Glucagon is a powerful *lipolytic agent.* It acts via stimulating cAMP system to activate lipase in adipose tissue, which releases FFA and glycerol into the circulation. In the liver, excess of FFAs are oxidised resulting in energy production and ketone body synthesis (ketogenesis). Thus glucagon is a ketogenic as well as a hyperglycaemic hormone.

3. Effects on protein metabolism. Glucagon increases the amino acid uptake of liver, which in turn, promotes gluconeogenesis. Thus, glucagon *lowers plasma amino acids.*

4. Calorigenic effect. Glucagon also has a calorigenic effect. This is not due to hyperglycaemia but this action requires the presence of glucocorticoids and T_4. It is probably related to increased hepatic deamination of amino acids.

5. Other actions of glucagon. Other miscellaneous actions of glucagon include:

- Inhibition of renal tubular sodium reabsorption resulting in natriuresis.
- Modest increase in force of contraction of the heart by activation of myocardial adenylyl cyclase.
- Stimulation of secretion of growth hormone, insulin and pancreatic somatostatin.
- Glucagon may also be synthesized in central nervous system and it may act locally in the regulation of appetite.

INSULIN–GLUCAGON RATIO

Under basal conditions, the usual molar ratio of insulin to glucagon in plasma is about 2.0.

Under circumstances that require mobilization and increased use of endogenous substrate, such as fasting and prolonged exercise, the insulin–glucagon ratio drops to 0.5 or less. Low insulin: glucagon ratio is helpful in maintaining

glucose supply to central nervous system (CNS) and energy requirements needs of the body by:

- Increasing glycogenolysis,
- Increasing amino acid mobilization and promoting gluconeogenesis and
- Increasing lipolysis, which enhances FFA to muscle and liver for oxidation.

Under circumstances in which substrate storage is advantageous, such as after a pure carbohydrate load or a mixed meal, the insulin–glucagon ratio rise to 10 or more. This high insulin: glucagon ratio helps in storage of substrate.

- Increasing glucose uptake, oxidation and conversion to liver and muscle glycogen, thus
- Suppressing unneeded proteolysis and lipolysis and
- Facilitating clearance of chylomicrons by activation of adipose tissue lipoprotein lipase.

REGULATION OF GLUCAGON SECRETION

I. Role of blood levels of nutrients

There exists a feedback relationship between blood levels of glucagon and nutrients (Fig. 8.6-6).

1. Blood glucose level

Low blood glucose concentration is the most potent stimulus for secretion of glucagon. Hypoglycaemia causes 2–4 fold increase in plasma levels of glucagon (Fig. 8.6-8).

High blood glucose concentration inhibits glucagon secretion (Fig. 8.6-8). Hyperglycaemia lowers glucagon secretion by 50%. The presence of insulin greatly potentiates the suppressive effect of high glucose levels on α cells.

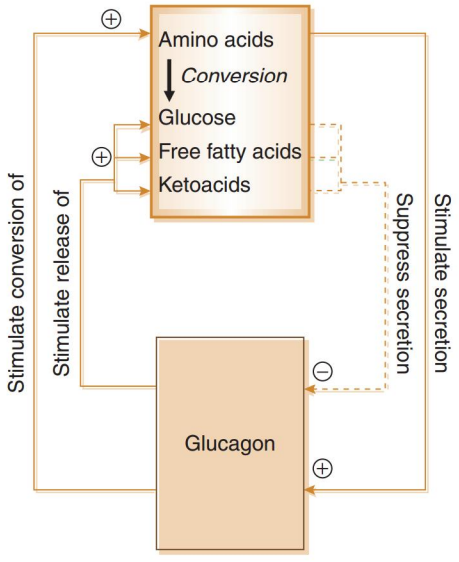


Fig. 8.6-8 Feedback relationship between glucagon and nutrients. Glucagon stimulates production and release of glucose, free fatty acids and ketoacids, which in turn suppress glucagon secretion, and glucagon in turn stimulates the conversion of amino acids to glucose.

2. Plasma amino acids.

Secretion of glucagon is increased by a protein-rich meal. There exists a feedback relationship, i.e. amino acids stimulate glucagon secretion, and glucagon in turn stimulates the conversion of amino acids to glucose (Fig. 8.6-8).

3. Free fatty acids and ketoacids.

Glucagon stimulates production and release of FFA and ketoacids, which in turn suppress glucagon secretion, i.e. there is a feedback relationship (Fig. 8.6-8).

II. Role of gastrointestinal hormones

Gastrointestinal hormones, such as CCK, gastrin and GIP increase glucagon secretion. This accounts for the enhanced glucagon response to orally ingested nutrients as opposed to the responses to intravenously delivered nutrients.

III. Role of nervous system

Sympathetic nerve stimulation to pancreas increases glucagon secretion. Various stresses, fasting, exercise and infection increase the glucagon secretion in part by their stimulatory effect on sympathetic nervous system and partly by release of glucocorticoids.

- Vagal stimulation and acetylcholine also acutely increase glucagon secretion.
- Neurohormone, somatostatin inhibits the secretion of glucagon, probably by paracrine or neurocrine effects made possible by the islet microarchitecture.

SOMATOSTATIN AND PANCREATIC POLYPEPTIDE

SOMATOSTATIN

Structure and synthesis.

Pancreatic somatostatin is a neuropeptide containing 14 amino acids, synthesized by δ cells. It is also synthesized by the intestinal cells and was originally discovered as a hypothalamic neuropeptide that inhibits growth hormone secretion.

Regulation of secretion.

Somatostatin secretion is increased after ingestion of food, because increased blood glucose, amino acids, fatty acids and gastrointestinal tract hormones stimulate its secretion. Glucagon, β-adrenergic and cholinergic neurotransmitters also stimulate somatostatin secretion. Insulin and α-adrenergic neurotransmitters inhibit somatostatin secretion.

Actions.

Somatostatin has following effects:

- Acts on the islets of Langerhans and inhibits secretion of insulin and glucagon.
- Increases the motility of stomach, duodenum and gall bladder.

- Decreases secretion of hydrochloric acid, pepsin, gastrin, secretin, intestinal juices and pancreatic juice.
- Inhibits the absorption of glucose, xylose, and triglycerides across the mucosal membrane.

In a nutshell, actions of somatostatin along with those of insulin and glucagon, probably co-ordinate nutrient input with substrate disposal.

PANCREATIC POLYPEPTIDE

Structure and synthesis. Pancreatic polypeptide has 36 amino acids and belongs to a family of similar molecules including neuropeptide Y in the hypothalamus. It is synthesized by the PP cells of islets of Langerhans.

Regulation of secretion. Pancreatic polypeptide is secreted in response to food ingestion via gastrointestinal secretagogues and cholinergic stimulation. Its secretion is also stimulated by hypoglycaemia and inhibited by glucose administration.

Actions and physiological importance. Its best known action is to inhibit exocrine pancreatic secretion. Its true physiological importance is not known.

HORMONAL REGULATION OF BLOOD GLUCOSE LEVEL

Normal blood glucose levels and body glucose reserves

Normal blood glucose levels

A healthy individual is capable of maintaining the blood glucose level within a narrow range.

- Fasting blood glucose level in a post-absorptive state varies between 70 and 110 mg/dL.
- Post-prandial blood glucose level, i.e. after a large carbohydrate meal or following oral administration of glucose in the dose of 1 g/kg body weight, the blood glucose level increases to about 140 mg/dL (less than 150 mg/dL) in a period of less than 1 h. However, when this response to oral administration of carbohydrate, when plotted on a time scale is called *glucose tolerance curve* (Fig. 8.6-9) and is used clinically as a test to study the maintenance of blood glucose levels.

Normal body reserves of glucose

Free glucose. About 18 g free glucose is present in an adult human body. This amount is just sufficient to meet the basal energy requirements of the body for 1 h.

Stored glucose is present in the form of glycogen in liver and muscles.

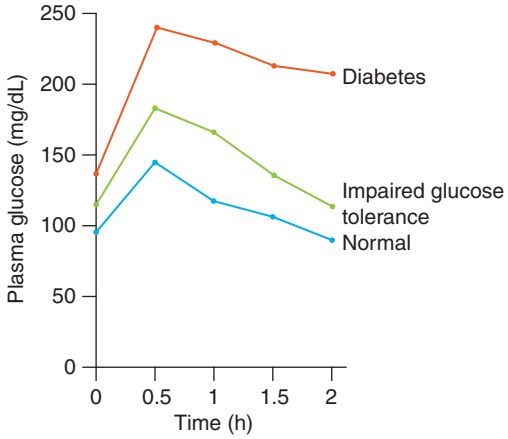


Fig. 8.6-9 Glucose tolerance curve.

- Liver has about 100 g stored glycogen. An adult liver (weighing about 1.5 kg) can provide only 40–50 g of blood glucose from glycogen that can last only for a few hours to meet the body requirement. However, liver is also capable of producing about 125–150 mg glucose/min or 180–220 g/2 h. Therefore, during an overnight fast, the glycogen stores of liver are not totally non-carbohydrate sources (gluconeogenesis).
- Muscle glycogen store is much more than of liver. However, degradation of glycogen in muscle does not directly produce glucose but produces lactate, which is used for gluconeogenesis.

Sources and utilization of blood glucose

Sources of blood glucose (Fig. 8.6-10)

1. Dietary sources. The dietary carbohydrates are digested and absorbed as monosaccharides (glucose, fructose, galactose, etc.). The liver is capable of converting fructose and galactose into glucose, which can readily enter the blood.

2. Gluconeogenesis. The glucose is synthesized in the liver and kidney. Precursors for gluconeogenesis include lactate, glycerol, propionate and some amino acids.

- Lactate is formed by degradation of glycogen stored in the muscle.
- Free glycerol and propionate are formed by breakdown of fat in the adipose tissue.
- Amino acids may be derived from dietary sources or from protein breakdown.

3. Glycogenolysis. Stored glycogen in liver is degraded to glucose, while muscle glycogen after degradation produces lactate, which is used for gluconeogenesis as described above.

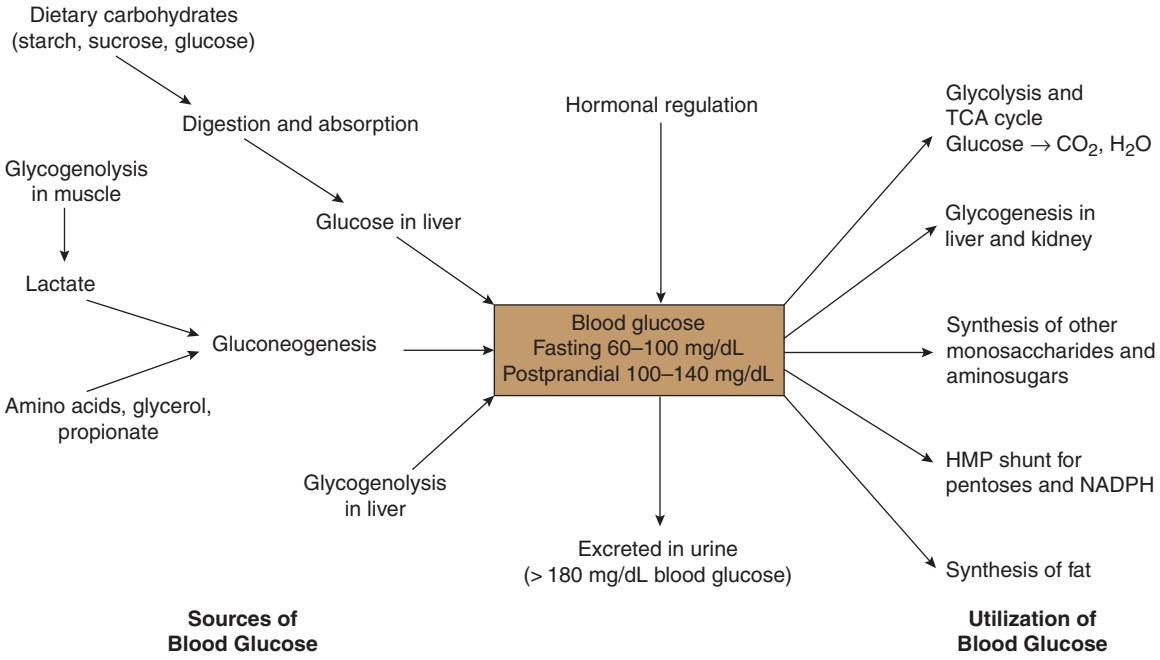


Fig. 8.6-10 Sources and utilization of glucose.

Utilization of blood glucose

The glucose present in the blood is utilized for:

1. *Provision of energy needs to body tissues.* The oxidative pathways in which glucose is used include:
   - Glycolysis and tricarboxylic acid (TCA) cycle.
   - Hexose monophosphate (HMP) shunt for pentoses and NADPH.
   - Uronic acid pathway.
2. *Glycogenesis,* i.e. synthesis of glycogen in liver and kidney,
3. *Synthesis* of other monosaccharides and amino sugar and
4. *Synthesis of fat.*

Role of hormones in regulation of blood glucose

Under normal circumstances, the various hormones play a significant role in maintaining the blood glucose levels within normal physiological range. This is accomplished by preventing the occurrence of hyperglycaemia and hypoglycaemia.

Prevention of occurrence of hyperglycaemia

The occurrence of hyperglycaemia after a pure carbohydrate load or a mixed meal in a healthy individual is prevented by a manifold (4–5 times) increase in insulin secretion (for details see actions of insulin).

Prevention of occurrence of hypoglycaemia

Hypoglycaemia, which may occur due to fasting or prolonged exercise, is prevented in a healthy individual by a number of hormones, which include glucagon, epinephrine, growth hormone and glucocorticoids. It is obvious that there is only one hormone, insulin, which prevents hyperglycaemia, whereas at least four hormones are available for prevention of hypoglycaemia.

APPLIED ASPECTS

Important applied aspects of endocrine pancreas which need mention are:

- Diabetes mellitus and
- Hypoglycaemia.

DIABETES MELLITUS

Diabetes mellitus, commonly called just diabetes, refers to a clinical syndrome of hyperglycaemia occurring due to deficiency of insulin.

Predisposing factors include:

(i) *Heredity* is the most common predisposing factor. The potential candidates to develop diabetes are with strong genetic disposition, e.g. first degree relatives of diabetics.

(ii) *Obesity* refers to an increase in body mass index (BMI) i.e. body weight in kg/(height)^2 in meters. The person with BMI value > 30 is considered as obese.

- In obese persons, the adipose tissues are usually more resistant to actions of insulin as compared to a normal (non-obese). As a consequence of this factor, there is decrease in uptake of glucose by adipose cells and, decrease in release of glucose from the liver.

– In obese person, there is hyperinsulinaemia associated with dyslipidaemia (high levels of circulating FFA and HDL) also called as *Metabolic syndrome or Syndrome X.* In this condition, it is postulated that load on β cells increases to produce more insulin, and ultimately leads to exhaustion of β-cell activity resulting in diabetes in a non-diabetic obese person.

TYPES AND STAGES OF DIABETES MELLITUS

Diabetes mellitus can be classified into following types:

1. Primary diabetes mellitus in which cause is not known. It is of further two types:

- Insulin-dependent diabetes mellitus (IDDM or Type-I), and
- Non-insulin-dependent diabetes mellitus (NIDDM or Type-II).

2. Secondary diabetes mellitus. It is associated with certain pathological conditions, such as pancreatitis, cystic fibrosis, acromegaly, Cushing syndrome, etc.

Insulin-dependent diabetes mellitus

- Insulin-dependent diabetes mellitus, or type I diabetes, is considered an autoimmune disorder in which antibodies destroy the β cells of islets causing an absolute deficiency of insulin.
- Genetic susceptibility is a major determinant while environmental factors act as a trigger.

Characteristic features of type-I (IDDM) are:
- It manifests before 40 years of age (usually between 12 and 15 years) and is also called juvenile onset diabetes. It accounts for 10–20%.
- Patients are usually lean.
- Classical triad of presenting symptoms consisting of polyuria, polydipsia and polyphagia is associated with weight loss.
- Ketosis and acidosis are common complications of this diabetes mellitus.
- Plasma insulin levels are very low or undetectable.

Non-insulin-dependent diabetes mellitus

Non-insulin-dependent diabetes mellitus, or type II diabetes, is also a genetic disorder. It is supposed to occur due to decrease in insulin receptors on the insulin responsive (target) cells.

Characteristic features of type-II (NIDDM) are:
- It manifests after 40 years of age and so is also called as *adult onset diabetes.*
- It is most common and accounts for 80–90% of diabetic population.
- Most of the patients are obese.
- Symptoms begin gradually and may be ignored and many a times diagnosis is made on urine examination which shows glycosuria.
- Plasma insulin levels are often normal or even elevated.
- Ketoacidosis is not very common.

IDDM versus NIDDM

The comparison between IDDM and NIDDM is given in Table 8.6-2.

PATHOPHYSIOLOGY OF DIABETES MELLITUS

Pathophysiology of diabetes mellitus revolves around the metabolic alterations associated with insulin deficiency. Most important among them are hyperglycaemia, ketoacidosis, hypertriglyceridaemia and protein catabolism (Fig. 8.6-11):

1. Hyperglycaemia and its consequences

Hyperglycaemia (elevation of blood glucose concentration) is the characteristic feature of uncontrolled diabetes mellitus. It occurs due to lack of insulin resulting in:

- Decreased peripheral utilization of glucose.
- Increased hepatic output of glucose (owing to glycogenolysis and gluconeogenesis) into the circulation.

(i) Glycosuria and its consequences

- *Glycosuria,* i.e. excretion of glucose into the urine occurs when the blood glucose level rises above the renal threshold point, i.e. above 180 mg/dL (see page 398).
- *Polyuria,* i.e. passage of large amount of urine frequently. It is the result of osmotic diuresis caused by renal excretion of osmotically active glucose molecules (see page 407).
- *Loss of electrolytes* (sodium, potassium and phosphate) in urine also occurs as a side effect of osmotic diuresis.
- *Cellular dehydration.* High glucose concentration increases osmotic pressure of the ECF and osmotic transfer of water from cells to the ECF leading to dehydration of cells. In addition to it, osmotic diuresis causes increased loss of water from the body thereby reducing ECF volume, which also causes compensatory dehydration of cells.
- *Polydipsia,* i.e. excessive drinking of water results from activation of thirst mechanism caused by cellular dehydration.
- *Increased caloric loss* is the result of loss of glucose in urine.
- *Polyphagia,* i.e. excessive eating occurs due to stimulation of satiety centre caused by deficient utilization of glucose in the hypothalamic ventromedial nuclei. Increased caloric loss also results in compensatory polyphagia.
- *Loss of body weight* occurs because of loss of calories in the urine and mobilization of fats and proteins for energy production. Since loss of body weight occurs in

Table 8.6-2 IDDM versus NIDDM

Feature	IDDM (Type I)	NIDDM (Type II)
General features		
Defect	Insulin deficiency due to β-cell destruction	Resistance of target tissues to insulin
Prevalence	10–20% of diabetic population	80–90% of diabetic population
Age of onset	<40 years	>40 years
Body weight	Low (thin and lean)	High (obese or normal)
Gene locus	Chromosome 6	Chromosome 1
Family history	Mild or moderate	Very strong
Clinical features		
Duration of symptoms	Weeks (rapid)	Months to year (slow)
Presenting symptoms	Polyuria, polydipsia, polyphagia	Usually patients present with different complications
Complication at the time of diagnosis	Absent	Present (in 10–20% cases)
Acute complication	Ketoacidosis	Hyperosmolar coma
Biochemical features		
Plasma insulin	Decreased or absent	Normal or increased
Autoantibodies	Frequently found	Rare
Ketonuria	Present	Absent
Treatment		
Treatment of choice	Insulin (oral hypoglycaemics not useful)	Oral hypoglycaemics (insulin usually not required)
Mortality if not treated	High	Low

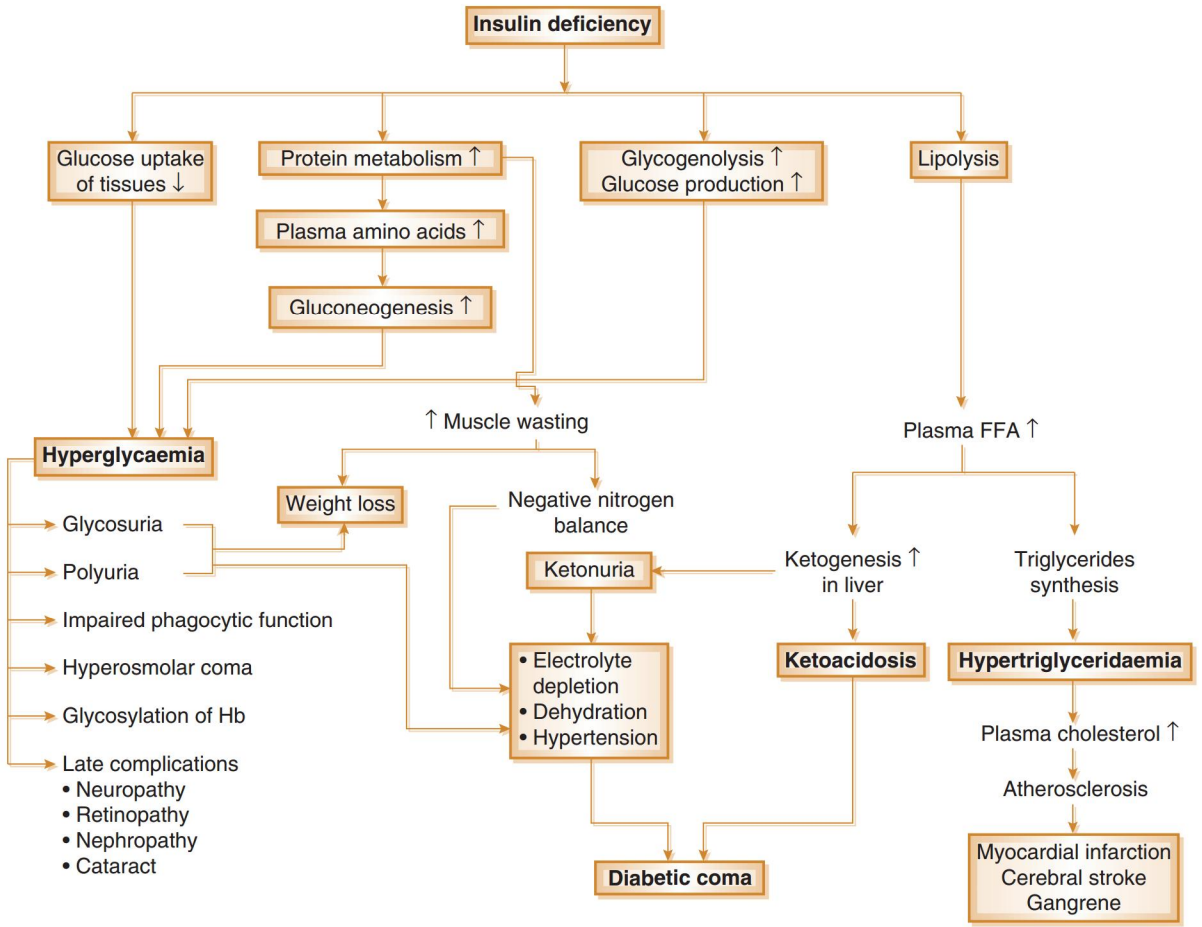


Fig. 8.6-11 Pathophysiology of diabetes mellitus and its complication.

spite of excessive food intake, diabetes is called a *condition of starvation in the midst of plenty.*

(ii) Impaired phagocytic function. Hyperglycaemia impairs all aspects of leucocytic phagocytic function, i.e. adherence, diapedesis, phagocytosis and intracellular killing. Because of impaired phagocytic function, the diabetics are more prone to infections as compared to the non-diabetics.

(iii) Hyperosmolar effects. Osmolarity of the blood goes on increasing with the increasing blood sugar levels. Under such circumstances, the plasma osmolality may be over 375 mOsm/kg. Such a high hyperosmolality may cause dehydration in central nervous system leading to impairment of cerebral functions. Ultimately, coma may result, which may be even fatal.

(iv) Glycosylation of proteins. Glycosylation of proteins refers to the post-translation, non-enzymatic addition of sugar residues to amino acids of proteins.

Glycosylation of haemoglobin. Glycosylated haemoglobin refers to the glucose-derived products of normal haemoglobin (HbA). Among the glycosylated haemoglobins, the most abundant form is HbA_{1C}, which is produced by condensation of glucose with N-terminal valine of each β chain of haemoglobin A (HbA).

- Normally, HbA_{1C} concentration is about 3–5% of the total haemoglobin.
- During sustained hyperglycaemia, as in diabetes mellitus, the concentration of HbA_{1C} may be elevated to 10–20% of the total haemoglobin.
- Determination of HbA_{1C} has become an important tool for monitoring of diabetes control and proper regulation of insulin therapy.

Glycosylation of tissue proteins occurs when the blood glucose levels remain elevated for a prolonged duration (years). Glycosylation leads to irreversible changes in the chemical structure of tissue proteins. These chemical changes have been implicated in producing long-term complications of diabetes mellitus, such as:

- Diabetic nephropathy,
- Diabetic retinopathy,
- Diabetic neuropathy and so on.

2. Ketosis, hypertriglyceridaemia and their consequences

Since due to insulin deficiency the utilization of glucose is poor, the body turns to fats for obtaining energy by lipolysis. As a result of lipolysis, plasma levels of FFAs are increased. Excessive FFAs in plasma leads to:

- Hypertriglyceridaemia and
- Ketosis.

Consequences of ketosis include:

- *Cellular dehydration*. Ketone bodies being hyperosmolar, remove water from the cells producing cellular dehydration.
- *Ketoacidosis*. Ketone bodies being strong acids dissociated readily and release H^+ ions. In the blood, these H^+ ions are buffered by bicarbonate ions (HCO_3^-) to form carbonic acid. Fall in bicarbonate level in the blood leads to acidosis called ketoacidosis.

Features of ketoacidosis are:

- Rapid, deep respiration (dyspnoea, Kussmaul breathing),
- Acetone smell in patient's breath and
- Urine becomes highly acidic.
- *Electrolyte loss.* The electrolyte and water loss further added to cellular dehydration.
- *Hypovolaemia and hypotension* may ultimately result from water and electrolytic loss and cellular dehydration.
- *Coma and death*. Depression of consciousness to the level of coma may eventually ensure owing to marked acidosis and dehydration which may finally lead to death.

3. Protein catabolism

Insulin is an anabolic hormone, i.e. it promotes protein synthesis and it also inhibits proteolysis. Therefore, in diabetes, due to insulin deficiency the protein anabolism is suppressed and catabolism is increased.

Consequences of suppression of protein anabolism and increased catabolism include:

- Protein depletion in the body,
- Muscle wasting and
- Negative nitrogen balance.

CLINICAL FEATURES, COMPLICATIONS AND DIAGNOSIS OF DIABETES MELLITUS

Clinical features and complications of diabetes mellitus

Cardinal symptoms include polyuria, polydipsia, polyphagia, weight loss. Occurrence of these symptoms has been explained in pathophysiology.

Biochemical signs include hyperglycaemia, glycosuria, ketosis, ketonuria and ketoacidosis. These have been fully elucidated in pathophysiology.

Complications include:

- *Pre-disposition to infections* due to impaired phagocytic function and protein depletion.
- *Acute complications* include ketotic coma and non-ketotic hyperosmolar coma.
- *Chronic complications* include:
  - *Atherosclerosis,* i.e. deposition of lipids underneath the tunica intima of blood vessels. The common sites are coronary, cerebral and peripheral arteries. It occurs

due to longstanding hyperlipidaemia and hypercholesterolaemia.

- *Microangiopathy,* a vascular lesion in which the capillary basement membrane becomes thicker, probably due to structural changes caused in tissue proteins by their glycosylation. It is responsible for common complications of longstanding diabetes, which includes:
  - Diabetic retinopathy,
  - Diabetic nephropathy, and
  - Diabetic neuropathy.

Diagnosis of diabetes mellitus

In clinically suspected cases, diagnosis is confirmed by following investigations:

1. Urine examination for glycosuria (see page 435). This is a rapid, simple and easy test for diagnosis of diabetes mellitus. Amount of glucose excreted in urine depends upon the severity of disease.

Disadvantage. Glycosuria depends upon the renal threshold level which itself is variable; hence, both overdiagnosis (false positive) and underdiagnosis (false negative) of diabetes are possible.

2. Urine examination for ketone bodies. Presence of ketone bodies (acetone) in urine along with glycosuria is almost diagnostic of diabetes mellitus. Other causes of ketonuria are starvation, prolonged fasting, following high-fat diet and after repeated vomiting.

3. Fasting and post-prandial blood glucose levels. Samples for estimation of fasting blood glucose are taken after overnight fast and that for post-prandial are taken after 2 h of normal diet. Normal values of plasma glucose are:

- Fasting: 70–110 mg/dL
- Post-prandial (after 2 h of meals): < 140 mg/dL.

4. Glucose tolerance test (GTT)

- *In a normal person,* fasting plasma glucose levels range between 70 and 110 mg/dL, after glucose intake. The peak value of about 140 mg/dL is reached in an hour or so which returns to fasting level within 2–2½ h (Fig. 8.6-9). Urine does not show the presence of glucose.
- *In diabetes mellitus,* glucose tolerance curve is abnormal. Fasting glucose level is high (≥ 126 mg/dL), after glucose intake peak is also high (≥ 200 mg%), and does not return to fasting level for a long time (4–6 h) (Fig. 8.6-9). This slow fall of glucose level indicates failure to control due to lack of insulin secretion following sugar ingestion.
- *Impaired glucose tolerance.* The fasting plasma levels between 110 and 126 mg/dL and peak values (after glucose ingestion) between 140 and 200 mg/dL are classified as impaired glucose tolerance (Fig. 8.6-9). Such patients are potential candidates to develop diabetes later on. Therefore, they need further supervision and repeated blood sugar estimations at frequent intervals to detect development of diabetes mellitus.

HYPOGLYCAEMIA

Hypoglycaemia refers to a clinical condition caused by blood glucose levels below 45 mg/dL (2.5 mmol/L). The human body has developed a well-regulated system for an efficient maintenance of blood glucose concentration (see regulation of blood glucose page 608). However, still hypoglycaemia (though not common) is observed under some circumstances.

TYPES AND CAUSES OF HYPOGLYCAEMIA

Broadly hypoglycaemia may be divided into two types:

- Hypoglycaemia in non-diabetics, and
- Hypoglycaemia in diabetics (more common).

A. Hypoglycaemia in non-diabetics

1. Post-prandial hypoglycaemia, also known as *reactive hypoglycaemia,* occurs typically after meals within 4 h after ingestion of food. It is caused by a transient rise in insulin levels and symptoms are short lasting. It is more common in patients who have undergone gastric resection.

2. Post-absorption or fasting hypoglycaemia usually does not occur in normal fasting patients. It is seen in patients with:

- Insulin secreting tumours (adenomas) of pancreatic islets causing *hyperinsulinism*
- In hepatic failure, degradation of insulin is less, which may result raised levels of insulin and hypoglycaemia.

B. Hypoglycaemia in diabetics

Hypoglycaemia in diabetics is more common in diabetics than in non-diabetics. About 4% deaths of IDDM are said to be due to hypoglycaemia.

Causes of hypoglycaemia in diabetics include:

- Overdose of antidiabetic drugs especially insulin is comparatively common cause of hypoglycaemia.

Other factors responsible for hypoglycaemia in patients on regular antidiabetic treatment are:

- Intake of too little or no food
- Heavy exercise
- Mismatch between insulin administration and food habits
- Alcohol intake etc.

SYMPTOMS AND SIGNS OF HYPOGLYCAEMIA

Symptoms and signs of hypoglycaemia occur due to effects of low levels of glucose per se (mainly on nervous system

Table 8.6-3 Hypoglycaemic versus hyperglycaemic coma in diabetics

Sr. No.	Feature	Hypoglycaemic coma	Hyperglycaemic coma
1.	Cause	Regular dose of insulin and no food leading to fall in blood glucose level	Too little or no insulin with regular food intake leading to high blood glucose level
2.	Precipitating factor	Severe unaccustomed exercise	Untreated/hidden infection
3.	Rate of onset	Rapid, develops within minutes	Invariably slow takes hours or days to develop
4.	Symptoms and signs		
	(i) Vomiting	• No or occasional vomiting	• Frequent vomiting with abdominal pain
	(ii) Breathing	• Laboured breathing • No abnormal smell in breath	• Rapid and shallow breathing (Kussmaul) • Air hunger present
	(iii) Pulse	• Bounding	• Weak/feeble
	(iv) Skin and tongue	• Moist as no dehydration	• Dry due to dehydration
	(v) CNS sign	• Tendon reflexes brisk • Plantar is extensor	• Diminished • Plantar is normal (flexor)
5.	Investigations		
	(i) Urine	• No glucose • No ketone bodies	• Glycosuria marked • Ketonuria marked
	(ii) Blood	• Low blood glucose (usually < 30 mg%) • Bicarbonate level normal • pH normal	• High blood glucose (usually > 400 mg%) • Low bicarbonate • Low pH

especially brain) and because of sympathetic stimulation (mainly on CVS, GIT and skin).

1. CNS symptoms are called *neuroglycopenic symptoms.* Since metabolism of brain mainly depends on the blood glucose level, it is depressed when glucose level falls below 50–70 mg/dL. Central nervous system becomes quite excitable (due to facilitation of neuronal activity by hypoglycaemia) which results into hallucinations, extreme nervousness, tremors, confusion, difficulty in concentration, incoordination, convulsions and drowsiness. When blood glucose levels fall further (< 30 mg/dL) hypoglycaemic coma may develop, which needs to be differentiated from hyperglycaemic coma in diabetics (Table 8.6-3), and needs an emergency treatment by immediate administration of large quantity of glucose intravenously.

2. CVS symptoms in hypoglycaemia are palpitation, tachycardia and cardiac arrhythmias.

3. GIT symptoms include nausea and vomiting.

4. Skin symptoms are sweating and hypothermia.

GASTROINTESTINAL HORMONES

The glandular cells secreting GT hormones are individually scattered in the epithelium of stomach and small intestine and not in the form of clusters of cells as in the endocrine glands. Hence GIT may be considered as the largest mass of cells that secrete hormones.

The GT hormones based on their physio-anatomical similarities can be broadly classified into three groups:

1. Gastrin family of hormones.
2. Secretin family of hormones.
3. Other GT hormones.

For details see page 454.

Chapter 8.7

Endocrinal Functions of Other Organs and Local Hormones

HORMONES OF THE HEART
- Structure
- Secretion
- Actions
- Natriuretic peptide receptors

HORMONES OF THE KIDNEY
- Renin
- Erythropoietin

PINEAL GLAND
- Functional anatomy
- Melatonin

THYMUS
- Functional anatomy
- Functions

LOCAL HORMONES
- Prostaglandins and related substances
- Other local hormones synthesized in tissues
- Local hormones produced in blood

In addition to the main endocrinal glands described in the previous chapters, the other organs which have endocrinal functions are heart, kidney, pineal gland thymus and others.

HORMONES OF THE HEART

The heart also acts as an endocrine organ. The hormones secreted by heart include:

- Atrial natriuretic peptide (ANP),
- Brain natriuretic peptide (BNP), and
- C-type natriuretic peptide (CNP).

STRUCTURE

Atrial natriuretic peptide (ANP). It was the first natriuretic hormone isolated from the heart. It is a polypeptide and has 28 amino acid residues. It is formed from a large precursor molecule containing 151 amino acid residues.

Brain natriuretic peptide (BNP). It was the second natriuretic hormone, first isolated from the porcine brain and hence named as BNP. In humans, it is present in the heart and to a lesser extent in brain also. It is a polypeptide having 32 amino acid residues.

C-type natriuretic peptide (CNP). It was the third natriuretic hormone to be isolated in sequence and so named C-type natriuretic peptide.

In the heart, it is present in very small amount. It is mainly present in the brain, the pituitary, the kidneys and vascular endothelial cells. It appears to be primarily a paracrine hormone, as very little amount is present in circulation.

SECRETION

Atrial natriuretic peptide secretion is proportionate to the degree to which atria are stretched by an increase in central venous pressure. Therefore, ANP secretion is affected by following conditions:

- *Increase in the extracellular fluid volume* following infusion of isotonic saline.
- *Immersion of body in water up to neck* increases central venous pressure by counteracting the effect of gravity on the circulation.
- *Rising from the supine to the standing position* lowers the central venous pressure and thus decreases the ANP secretion.

ACTIONS

1. Increase in sodium excretion by kidneys. ANP and BNP increase excretion of sodium ion in urine by their following effects:

- Increasing glomerular filtration by dilating afferent arterioles and relaxing mesangial cells and
- Inhibiting Na^+ reabsorption at the level of renal tubules.

2. Lowering of blood pressure. ANP lowers the blood pressure by their peripheral and central effects.

(i) Peripheral blood pressure lowering effects include:

- Increasing the capillary permeability leading to extravasation of fluid and decline in blood pressure.
- Relaxing vascular smooth muscle in arterioles and venules.
- Inhibit renin secretion and thus counteract the pressor effects of catecholamines and angiotensin II.

(ii) Central blood pressure lowering effect is exerted through the ANP containing neural circuits in the brain which project from the anteromedial part of the hypothalamus to the areas in the lower brainstem that are concerned with neural regulation of cardiovascular system.

NATRIURETIC PEPTIDE RECEPTORS

Three types of natriuretic peptide receptors (NPR) are known:

1. NPR-A. It has an intracellular guanylyl cyclase domain. Atrial natriuretic peptide has greatest affinity for this receptor.

2. NPR-B. It also has an intracellular guanylyl cyclase domain. C-type natriuretic peptide has the greatest affinity for this receptor.

3. NPR-C. It has only a small cytoplasmic domain. It probably does not trigger any intracellular change. It removes natriuretic peptides from the blood stream and then releases them later, helping to maintain a steady blood level of the hormones; and is thus also called clearance receptor.

HORMONES OF THE KIDNEY

The kidneys secrete three hormones:

- Renin,
- 1,25-dihydroxycholecalciferol (see page 376) and
- Erythropoietin.

RENIN

Structure. Renin is a glycoprotein with a molecular weight of 37,326 in humans secreted by the granular cells of juxtaglomerular apparatus of the kidneys into the blood stream.

Actions. The only action of active renin is to convert angiotensinogen (renin substrate) into angiotensin-I. For further details about renin–angiotensin system (see page 417).

ERYTHROPOIETIN

Structure. Erythropoietin is glycoprotein with 165 amino acid residues and four oligosaccharide chains that are necessary for its activity in vivo.

Source. In adults, erythropoietin is mainly (85%) secreted by the juxtaglomerular apparatus of the kidneys with some contribution (15%) from the perivenous hepatocytes in the liver.

Actions. The main role of erythropoietin is to stimulate the bone marrow and cause erythropoiesis (for details see page 104).

PINEAL GLAND

FUNCTIONAL ANATOMY

Pineal gland, also known as epiphysis, is a small structure (5 mm × 7 mm) shaped like a pine cone. It is situated in the groove between the two superior colliculi in diencephalic area of brain above the hypothalamus (Fig. 8.7-1).

Structure. The pineal stroma has two types of cells: neuroglial and parenchymal.

Parenchymal cells are large epithelial cells with features suggesting that they have a secretory function.

- *In infants,* the pineal gland is large and the cells tend to be arranged in alveoli.

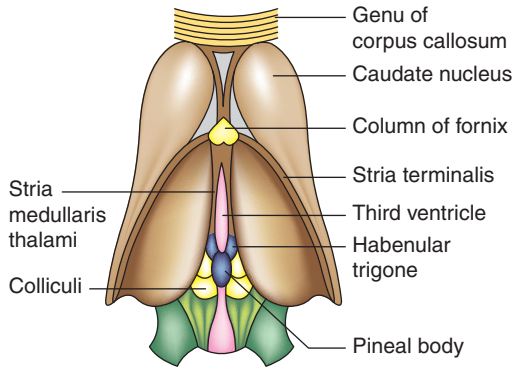


Fig. 8.7-1 Location of pineal body in the groove between the two superior colliculi.

- *In adults,* the pineal gland gets calcified, i.e. small concretions of calcium phosphate and carbonate *(pineal sand)* appear in the tissue.

MELATONIN

Structure, synthesis, plasma levels and metabolism

Structure and synthesis. The hormone melatonin is an indole (N-acetyl-5 methoxy-tryptamine). It is synthesized by the parenchymal cells of the pineal gland.

Plasma levels of melatonin show fluctuations with night time rise. The nocturnal plasma levels of melatonin are much higher in children than adults and they decline with age. The average plasma levels of melatonin at various age groups are:

- 1–3 years of age: 250 pg/mL
- 8–15 years of age: 120 pg/mL
- 20–27 years of age: 70 pg/mL
- 67–84 years of age: 30 pg/mL

Metabolism. In the liver, circulating melatonin is rapidly metabolized by 6-hydroxylation followed by conjugation. More than 90% of melatonin that appears in the urine is in this form. The exact pathway for melatonin metabolism in brain is not known.

Functions of melatonin

1. Role in circadian rhythm of the body. The dark-light cycle through suprachiasmatic nuclei of hypothalamus initiate neural and humoral signals that entrain a wide variety of well-known circadian rhythms including diurnal variation in melatonin secretion. The nocturnal peaks of secretion of melatonin also play an important role in circadian rhythm.

2. Effects on the gonads. Both inhibitory and facilitatory effects of melatonin on the gonads. This variability in the effect has led to the hypothesis that the diurnal change in the melatonin secretion that functions as some sort of timing signal which co-ordinates internal events with the light-dark cycle in the environment.

3. Effect on melanocyte-stimulating hormone (MSH) and adrenocorticotropic hormone (ACTH) secretion. An inhibitory effect of melatonin on MSH and ACTH secretion has been reported.

4. Other actions of melatonin include induction of sleep and inhibition of puberty.

Regulation of melatonin secretion

Melatonin secretion shows diurnal variation. It is secreted more during dark period of the day than during the day light hours. This correlates with various internal activities in different periods of the day, i.e. circadian rhythm.

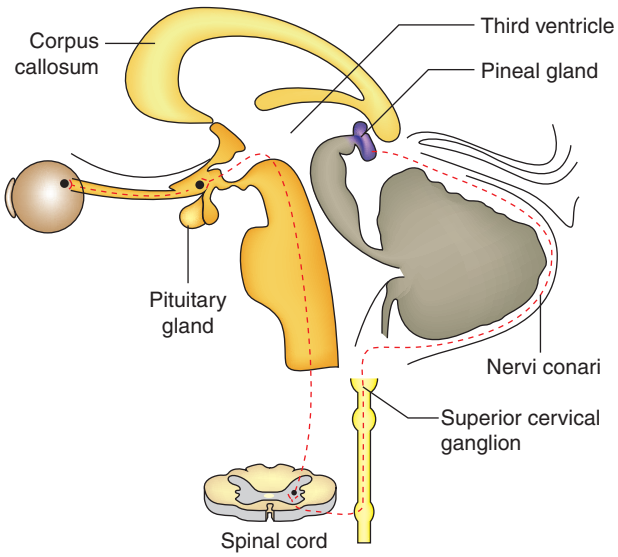


Fig. 8.7-2 Sagittal section of human brainstem showing pineal gland and its innervation (dotted line). ***Note.*** The pineal body forms the posterior boundary of third ventricle and lies under the posterior end of corpus callosum.

Hypothalamus is responsible for circadian fluctuations of melatonin secretion. Hypothalamus exerts its effect through the norepinephrine secreted by post-ganglionic sympathetic nerves (nervi conari) that innervate the pineal gland.

The neural pathway involved is (Fig. 8.7-2):

- *Retino-hypothalamic fibres* involved in light-dark cycle synapse in the suprachiasmatic nucleus of the hypothalamus.
- *Descending pathways from the suprachiasmatic nucleus of the hypothalamus* converge on the intermediolateral grey column of the thoracic spinal cord and end on the pre-ganglionic sympathetic neurons.
- *Pre-ganglionic fibres* pass from the spinal cord to superior cervical ganglion.
- *Post-ganglionic neurons* from the superior cervical ganglion project to the pineal in the nervi conari (Fig. 8.7-2).

THYMUS

FUNCTIONAL ANATOMY

Thymus is a small lymphoid structure located in the lower part of neck in front of the trachea, below the thyroid gland. At birth, it is small (weighing 10–12 g), gradually enlarges till puberty when it weighs 20–30 g, and then it starts decreasing in size and in old age weighs about 3–6 g. The sex glands exert a depressant effect on the thymus, therefore,

castration (removal of gonads) prolongs the period of persistence of the thymus.

Histologically, thymus consists of the inner medulla and outer cortex.

- *Medulla.* It comprises reticular epithelial cells, a few lymphocytes and concentric *corpuscles of Hassall.*
- *Cortex.* It includes actively multiplying, closely packed lymphocytes and contains no Hassall's corpuscles.

FUNCTIONS

Thymus has two functions:

- Immunological functions and
- Endocrinal functions.

1. Immunological functions of thymus

(i) *Development of immunologically competent* T-lymphocytes is an essential function of the thymus (see page 133).

(ii) *Maintenance of adequate pool of T-lymphocyte.* The hormone thymosin produced by the thymus also stimulates lymphopoiesis in the peripheral lymphoid tissue and thus plays a role in maintenance of an adequate pool of T-lymphocytes in adult life.

2. Endocrine function of thymus. Thymus tissue secretes two hormones, thymosin and thymin.

(i) *Thymosin.* It is a peptide, which, as described above, promotes proliferation of T-lymphocytes in the thymus and peripheral lymphoid tissue.

(ii) *Thymin*, also called thymopoietin, inhibits acetylcholine release at motor nerve endings and thus suppresses neuromuscular activity. Therefore, in hyperactivity of thymus, there occurs myasthenia gravis (see page 65).

LOCAL HORMONES

As described earlier, the endocrine glands secrete hormones into the blood stream, which show their actions at some distant places. In contrast, the local hormones are the substances which are produced in many tissues, and when activated in certain circumstances, execute their actions in the same area or in immediate neighbourhood. Commonly produced local hormones are:

- Prostaglandins (PGs) and related substances, such as thromboxanes, prostacyclin, leukotrienes and lipoxins.
- Other local hormones include acetylcholine, serotonin (5HT), histamine, adenosine derivatives, e.g. AMP, ADP and ATP; and plasma polypeptides, e.g. angiotensin, plasma kinins, etc.
- Local hormones produced in the blood, such as bradykinin, serotonin and angiotensinogen.

PROSTAGLANDINS AND RELATED SUBSTANCES

Prostaglandins and related substances include thromboxanes, prostacyclin, leukotriene and lipoxin. These substances are called *eicosanoids,* reflecting their origin from *arachidonic acid,* linoleic and linolenic acid.

Prostaglandins were so named by Von Euler in 1937, because first of all they were isolated from prostatic secretion in semen. However, now they are known to be synthesized in almost all tissues of the body. Presently, a variety of PGs are identified but the active forms are PGD_2, PGE_2 and PGF_2.

Synthesis of prostaglandins and related substances

Steps involved in the synthesis of PGs and the related substances are (Fig. 8.7-3):

- *Phospholipids* of the cell membrane are released by the action of phospholipase A_2 and converted into arachidonic acid.
- *Arachidonic acid* is converted into prostaglandin H_2 (PGH_2) by the action of enzymes cyclo-oxygenase 1 and 2 (Cox-1, and 2).
- *PGH_2* is converted into various other PGs, thromboxanes and prostacyclin by various tissue isomerases (Fig. 8.7-3).

Note. In addition to the above mentioned hormones, arachidonic acid is converted into two more hormones:

- By the action of 5-lipoxygenase into 5-hydroperoxy-eicosatetraenoic acid (5-HPETE), which is converted into leukotrienes.
- By the action of 15-lipoxygenase into 15-HPETE, which is then converted into lipoxins.

Actions of prostaglandins and related substances

Actions of prostaglandins. Prostaglandins have multitudinous and varied actions on almost all tissues of the body. Many of them are discussed in the chapters on the systems in which they play an important role. Some important actions of PGs are:

1. Actions on cardiovascular system. Peripheral arteriolar dilatation, especially in splanchnic and muscular bed, is caused by PGA_1 and PGA_2.

2. Actions on kidneys. PGA_2 increases renal cortical blood flow and increases urinary excretion of sodium, potassium and water.

3. Actions on female reproductive system

- $PGF_{2\alpha}$ is reported to initiate labour by stimulating contraction of gravid uterus.

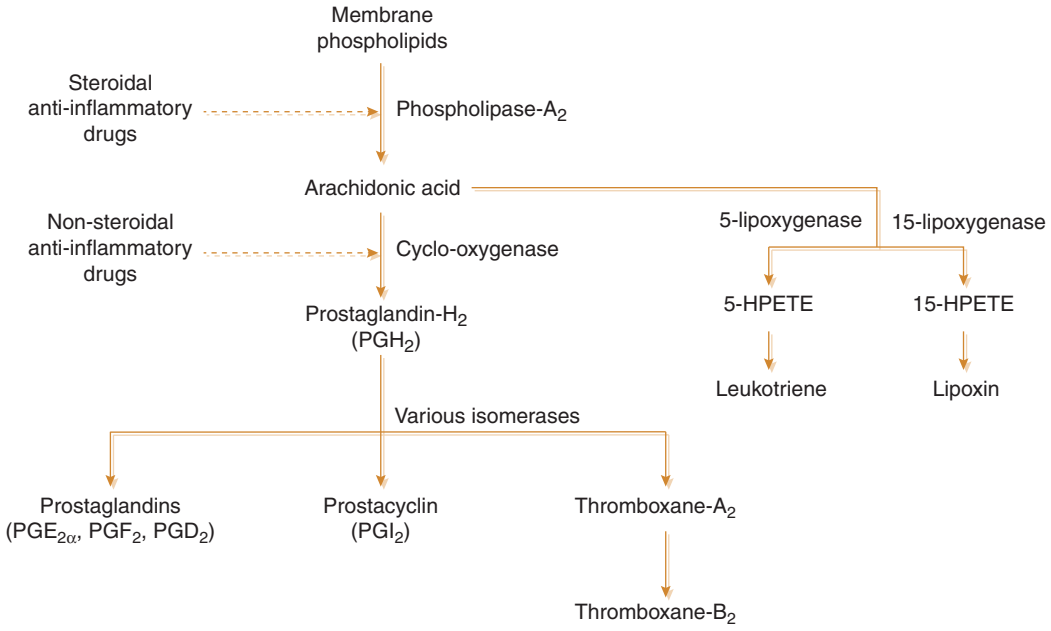


Fig. 8.7-3 Synthesis of prostaglandins and related substances.

- $PGF_{2\alpha}$ is also reported to be responsible for painful uterine contractions during menstruation (dysmenorrhoea).
- PGE_2 and PGF_2 promote secretion of hypothalamic gonadotropin releasing hormone (GnRH).

4. Role of PGs in inflammation. Prostaglandins are reported to mediate following effects of inflammation:

- Histamine-induced vascular permeability.
- Pain producing effect of bradykinin by sensitizing cutaneous nerves.
- Increase in vascular permeability and cellular infiltration.

5. Actions on blood platelets

- PGE_1 inhibits platelet aggregation through activation of adenylyl cyclase.

6. Action on bronchial musculature

- $PGE_{2\alpha}$ causes contraction of bronchial smooth muscles and may precipitate bronchial asthma.
- PGF_2, on the other hand, relax bronchial smooth muscles.

7. Actions on gastrointestinal tract

- PGE_1, PGE_2 and PGA_1 inhibit the secretion of gastric HCl.
- PGE_2 and $PGF_{2\alpha}$ cause inhibition of sodium and water absorption producing profuse watery, cholera-like diarrhoea.
- PGE and PGF increase intestinal motility.

8. Metabolic actions of PGs in vivo are variable. In vitro, PGE_1 inhibits the ACTH, growth hormone, glucagon and epinephrine-induced lipolysis.

9. Actions on nervous system

- On central nervous system, the PGs function as transmitters or modulators of neuron activity.
- In the ANS, the PGEs stimulate cholinergic neuroeffector junctions and inhibit the release and response to norepinephrine.

10. Actions on the eye. PGE_2 and $PGF_{2\alpha}$ occur in the iris and produce miosis.

Actions of thromboxane A_2. Thromboxane A_2 is synthesized by platelets. It promotes:

- Vasoconstriction and
- Platelet aggregation.

Actions of prostacyclin. Prostacyclin is produced in vascular endothelium. It produces vasodilatation.

Actions of leukotrienes. Leukotrienes are mediators of allergic responses and inflammation. Their release is provoked when specific allergens combine with IgE antibodies on the surfaces of mast cells. They produce:

- Bronchoconstriction,
- Arteriolar constriction,
- Increased vascular permeability and
- Attract neutrophils and eosinophils.

Actions of lipoxins. Physiological role of lipoxins is uncertain. Their actions include:

- Dilatation of microvasculature (by lipoxin A)
- Inhibition of cytotoxic effects of natural killer cells (by lipoxin A and lipoxin B).

OTHER LOCAL HORMONES SYNTHESIZED IN TISSUES

In addition to PGs and related substances, the other local hormones synthesized in the tissues are:

- Acetylcholine (see page 63),
- Serotonin (see page 790),
- Histamine (see page 791),
- Substance P (see page 793),
- Heparin (see page 158), and
- Gastrointestinal hormones (see page 454).

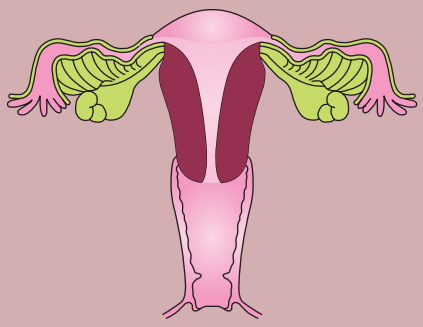
LOCAL HORMONES PRODUCED IN BLOOD

Local hormones produced in blood are:

- Serotonin (see page 152),
- Angiotensinogen (see page 592) and
- Bradykinin (see page 260).

SECTION 9

Reproductive System



Reproduction is a multidimensional subject with physiological, biochemical, genetic, psychological, emotional, social, economic, moral and many other aspects. The physiology of reproductive system begins with *sex determination,* i.e. genetic differentiation which occurs during fertilization. An ovum always contains 22 + X chromosomes, while a sperm may contain either 22 + X or 22 + Y chromosomes. Therefore, after fertilization the zygote's chromosomal patterns can be either 44+XX (i.e. female genotype) or 44+XY (i.e. male genotype). The next step in reproductive physiology is *sex differentiation* which begins at about 7–8 weeks of intrauterine life when the primitive, bipotential sex gland or gonad (gone=seed) differentiates into either testis or ovary depending upon the genotype (gonadal differentiation). From this point onwards, there will occur development of male or female accessory sex organs *(genital differentiation or phenotype)*. After remaining quiescent during childhood, the gonads suddenly awaken into vigorous activity for a period of 3–4 years during which gonadal development and maturation reaches to the point where reproduction is possible for the first time. During this phase, there also occurs a sudden spurt of physical growth and the child grows into an adult. This transitional period between the childhood and adulthood is called period of *puberty,* during which the secondary sex characters develop.

The adult *male reproductive physiology* and *female reproductive physiology* involves gametogenic and endocrinal functions of testes and ovaries, respectively. In a sexually active female, the *physiology of pregnancy* involves: fertilization at a

proper time and place (fallopian tube), development of the fetus in mother's womb and finally birth of a new human being, either male or female.

After child birth mother continues to provide nourishment by breastfeeding for a further period of 6–9 months *(physiology of lactation)*.

The understanding of reproductive physiology is not only important for the promotion of conception and normal fetal growth, but also for the prevention of conception *(physiology of contraception)* for population control which is of global concern.

The point when reproductive physiology ends is called *climacteric* (literally meaning, a major turning point). In females it is *menopause;* but in males there is no sharp point.

Chapter 9.1

Sexual Growth and Development

PRE-PUBERTAL SEXUAL GROWTH AND DEVELOPMENT

In human embryo, sexual growth involves two processes:

- Sex determination
- Sex differentiation.

SEX DETERMINATION

Sex determination, also known as *genetic differentiation,* refers to the genotype of the fetus, whether male or female. The genotype is determined by the presence of sex chromosomes, hence also known as *chromosomal sex differentiation.*

Human chromosomes. Each cell (except ovum and sperm) in a normal adult male and female possesses 46 chromosomes (44 autosomes + 2 sex chromosomes) usually arranged in an arbitrary pattern (karyotype).

- *Sex chromosomes* are called X and Y chromosomes.
- *The females* possess 44 autosomes plus 2X chromosomes (44+XX).
- *The males* possess 44 autosomes plus a X chromosome and a Y chromosome (44+XY).

Human gametes. The mature male gametes are called sperms and mature female gametes are called ova. During gametogenesis, there occurs meiosis *(reduction division);* therefore the mature sperm and ovum contain half the number of chromosomes, i.e. 23 (22 autosomes+one sex chromosome). This is called *haploid* number.

- Since, the primitive female germ cells *(oogonia)* from which mature ova are formed contain 44+XX chromosomes, so each ovum will contain 22+X chromosomes (Fig. 9.1-1).
- The primitive male germ cells (spermatogonia) from which mature sperms are formed contain 44+XY chromosomes, so half of the normal sperms will contain 22+X and other half will have 22+Y chromosomes (Fig. 9.1-1).

Genetic sex determination of the embryo occurs during fertilization, i.e. penetration of the ovum by the sperm as:

- When an ovum (22+X) is fertilized by a sperm containing 22+X chromosomes, the resultant zygote's chromosomal pattern will be 44+XX (*female genotype*).
- When an ovum (22+X) is fertilized by a sperm containing 22+Y chromosomes, the resultant zygote's chromosomal pattern will be 44+XY (*male genotype*).

Note. The human Y chromosome is smaller than the X chromosomes. The sperms containing the Y chromosomes are lighter and able to swim faster up in the female genital

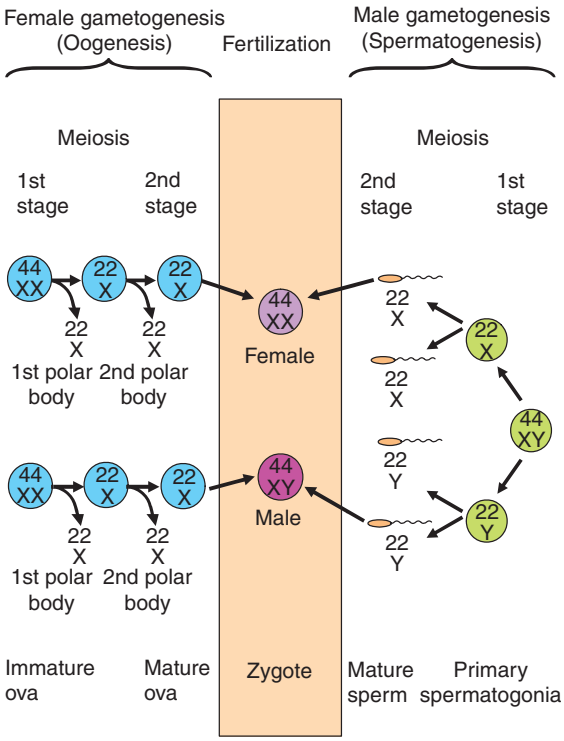


Fig. 9.1-1 Basis of genetic sex determination.

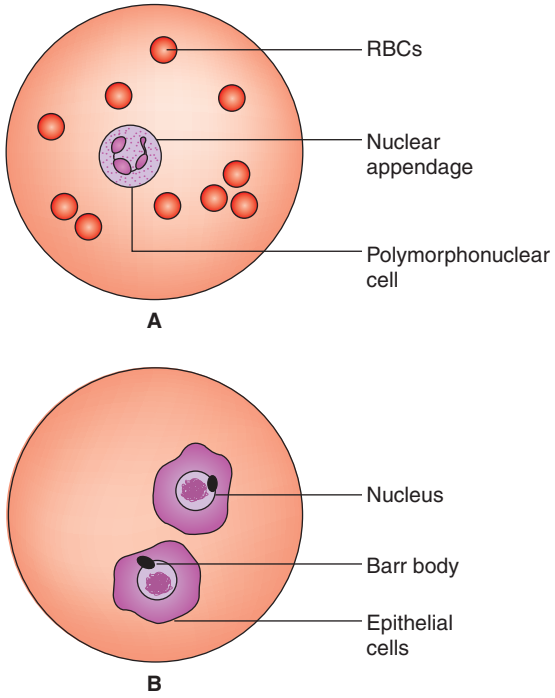


Fig. 9.1-2 Sex chromatins (Barr body) seen in: A, polymorphonuclear cell and B, epithelial cells of epidermal spinous layer.

tract, thus reaching the ovum rapidly. This probably accounts for the fact that the number of males born is slightly greater than the number of females.

Formation of Barr body

During embryonic development, the somatic cells start multiplying immediately after fertilization. It has been seen that one of X chromosomes of somatic cells in the female embryo becomes inactive while the other remains active. The exact details about the inactivation process of X chromosomes are not known.

The inactive X chromosome of each somatic cell forms a condensed mass called *sex chromatin or Barr body*. The Barr body can be seen near the nuclear membrane of the cells (Fig. 9.1-2).

Significance of Barr body

1. Identification of sex genotype. Since Barr bodies are present in the somatic cells of females only, so the sex genotype can be identified by a cytological test. The most suitable cells for this test are polymorphonuclear cells (in about 15% of the polymorphonuclear cells, the Barr bodies are seen as drumsticks projecting from the nuclei).

2. Identification of abnormal genotypes. In the abnormal cells with three or more X chromosomes, there are two or more Barr bodies.

SEX DIFFERENTIATION

After fertilization, the normal sex differentiation in the embryo proceeds sequentially. The stages of sex differentiation are:

- Gonadal differentiation,
- Genital differentiation and
- Psychological differentiation.

GONADAL DIFFERENTIATION

Gonadal differentiation or *gonadogenesis* refers to the formation of gonads, i.e. testes in males and ovaries in females. Gonadal sex differentiation is dependent on the genotype of the embryo.

Genital ridge or the urogenital ridge (the condensation of mesenchymal tissue present on each side near the adrenal glands) is the site where gonads develop. The primordial germ cells migrate into the genital ridge, where proliferation of both germinal and non-germinal cells leads to formation of bipotential gonads.

Bipotential gonads. Bipotential gonads are also known as primordial or primitive or indifferent or ambisexual gonads. Up to 6 weeks of gestation the bipotential gonads are identical in both sexes and have the rudiments of both male and female gonads (Fig. 9.1-3).

Structure. The bipotential gonad consists of a medulla, a cortex and primordial germ cells. The germ cells are embedded in the layer of cortical epithelium surrounding a core of medullary mesenchymal tissue (Fig. 9.1-3A, B).

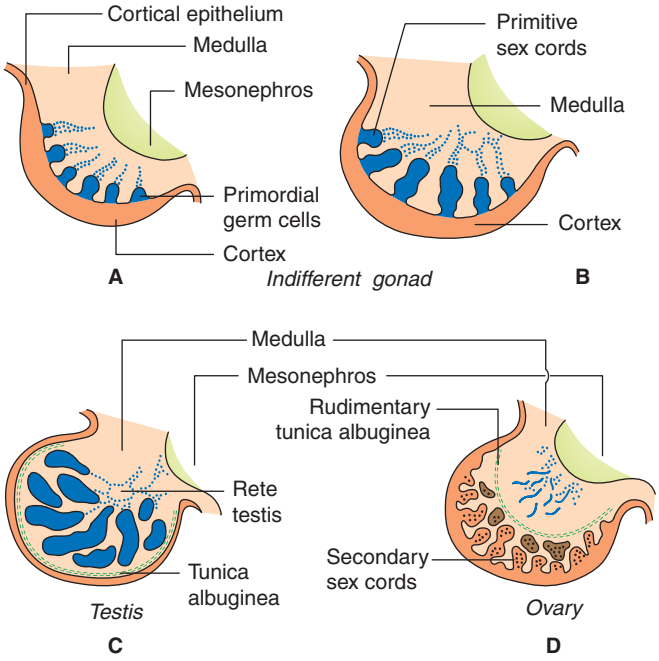


Fig. 9.1-3 Diagrammatic structure of human bipotential gonad (A and B), and development of testis from the medulla (C) and ovary from the cortex (D).

Testicular differentiation

In genetic male (44+XY) embryo, the bipotential gonads begin to differentiate into testes at approximately sixth week (Fig. 9.1-3C). The Y chromosome plays a key role in the process of testicular differentiation.

At about 6th week of gestation, the testicular differentiation begins with the appearance of primitive seminiferous cords (sex cords) from the germinal epithelium covering the medulla of bipotential gonad (Fig. 9.1-3).

At about eighth week of gestation, Leydig (interstitial) cells appear in the interstitial spaces of seminiferous tubules and continue to proliferate. The Leydig cells are derived from the sex cords that are not canalized. The membrane of Leydig cells has receptors for human chorionic gonadotropins (HCG) and for luteinizing hormone (LH).

At about ninth week of gestation, the Leydig cells synthesize and secrete testosterone in response to HCG secreted by placenta.

At about the 35th week of gestation, there occurs descent of testes through inguinal canal into scrotum. This marks the final stage of testicular differentiation.

Role of Y chromosome in testicular differentiation

- Two transcription genes, one for testicular differentiation and another for the formation of Mullerian duct inhibitory substance (MIS), are present on the Y chromosome.
- The gene responsible for testicular differentiation is called SRY gene (SRY=Sex-determining region of the Y chromosome).
- The SRY gene encodes the *testis-determining factor* (TDF), which triggers the testicular differentiation.
- The TDF gene product causes Sertoli cell differentiation, which is critically important for all events in male sexual differentiation.

Ovarian differentiation

- In genetic female (44+XX) embryo, by about 10th week of gestation, ovarian differentiation occurs in the absence of TDF.
- The ovaries develop on each side from the cortical region of the bipotential gonad (Fig. 9.1-3D).
- The coelomic epithelial cells (cortical cells) proliferate and form granulosa cells, which surround the germ cells and commit them to oocyte formation.
- Meanwhile, the medulla from which testes develop, regresses.
- During 11–12th week of gestation, oogonia undergo meiotic division to form oocyte, which is the end point of ovarian differentiation.
- Embryonic ovary, like testis, does not secrete any hormone.

Note. Presence of XX chromosome is must for ovarian development. Therefore, in chromosomal abnormality having XO constitution, there is no ovarian development.

GENITAL DIFFERENTIATION

Genital sex differentiation, also known as *phenotypic sex* differentiation, refers to the differentiation of internal and external genitalia.

Differentiation of internal genitalia

- The internal genitalia differentiate from the *neutral sex anlagen,* which develops during sixth week of gestation along with the development of bipotential gonad.
- The primordia of internal genitalia are a paired set of Wolffian (male) ducts and a paired set of Mullerian (female) ducts. By seventh week of gestation, the embryo has both male and female primordial ducts (Fig. 9.1-4).

Differentiation of male internal genitalia. In the genetic male fetus (44+XY) with functioning testes (Fig. 9.1-4B):

- The testosterone secreted by the Leydig cells stimulate the Wolffian ducts to form the epididymis, vas deferens and seminal vesicles.
- The MIS causes regression of the Mullerian ducts by apoptosis.

Differentiation of female internal genitalia (Fig. 9.1-4C)

- In the genetic female fetus (44+XX), in the absence of MIS, the female ducts (Mullerian ducts) proliferate and form oviduct (uterine tubes), uterus and upper two-thirds of vagina.
- In the absence of testosterone, Wolffian ducts degenerate.

Differentiation of external genitalia

The external genitalia in both sexes develop from *common anlagen,* which are the urogenital sinus, the genital sinus, the genital tubercle, the genital swelling and the genital (urethral) folds (Fig. 9.1-5).

- The external genitalia are bipotential till eighth week of gestation, i.e. it can develop along either male or female lines.

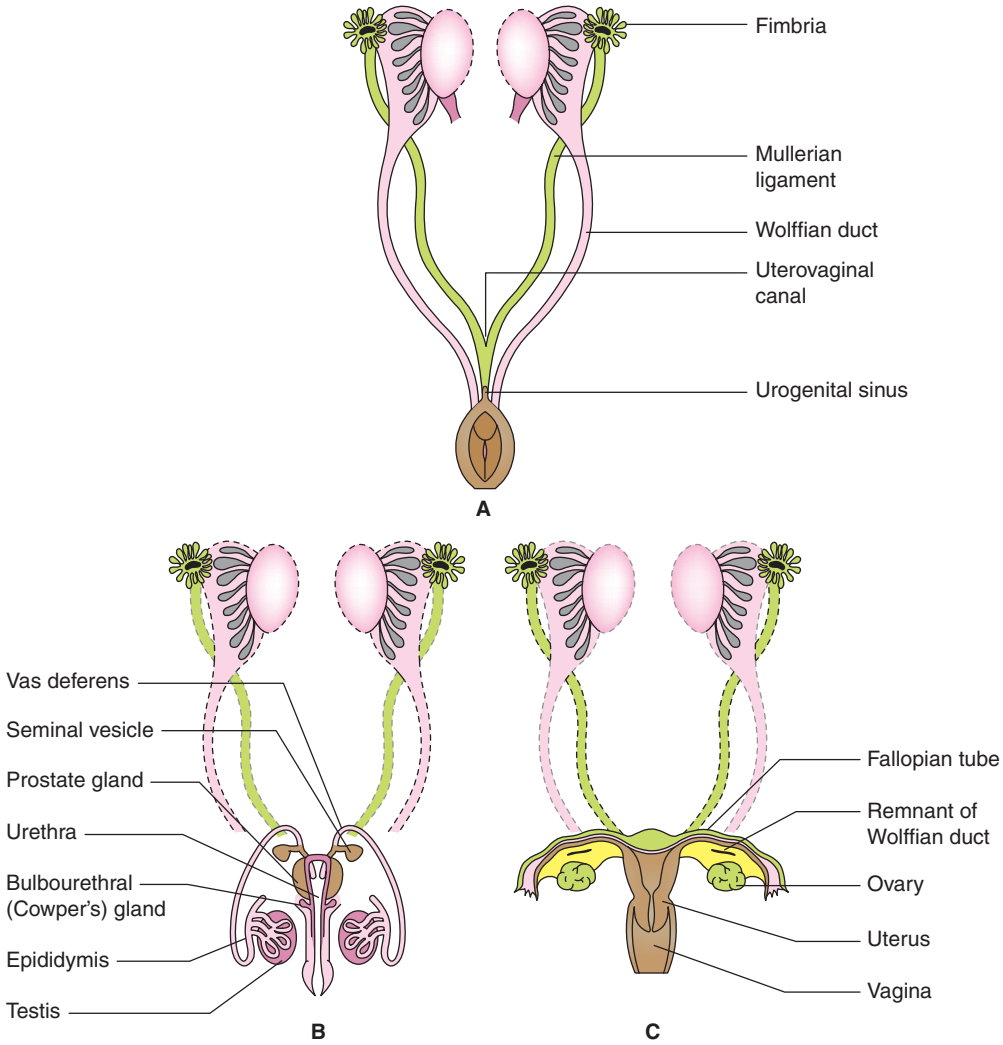


Fig. 9.1-4 Development of male (B) and female (C) internal genitalia from primordial genital ducts (A).

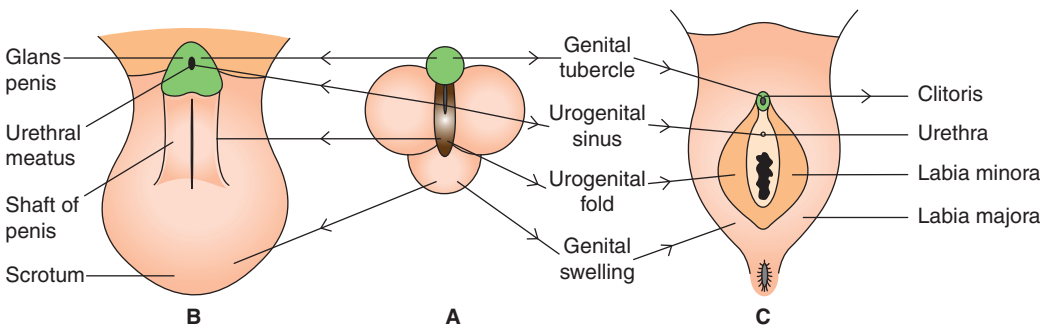


Fig. 9.1-5 Differentiation of external genitalia in male (B) and female (C) from common anlagen (A).

- *In male fetus* having functional testis secreting testosterone and dihydrotestosterone, the external genitalia acquire male characteristics by the fifth month of gestation.
- *In female fetus*, in the absence of any hormone, external genitalia differentiation occurs along the female line.

The external genitalia derived from the common anlagen in male and female are shown in Table 9.1-1.

PSYCHOLOGICAL DIFFERENTIATION

Psychological sex differentiation refers to a normal sexual behaviour in adult male and female. It is determined by the effect of androgens on the development of brain in the embryonic stage.

DISORDERS OF SEXUAL DEVELOPMENT

Abnormalities of sexual development occur due to:

- Defect in sex chromosomes leading to genetic abnormalities.
- Hormonal abnormalities leading to defect in gonadal and genital differentiation

Table 9.1-1 The male and female external genitalia derived from the common anlagen

Anlagen part	Male derivative	Female derivative
Urogenital sinus	Prostate and prostatic urethra	Urethra
Urethral fold	Penile urethra and shaft of penis	Labia minora
Genital swelling	Scrotum	Labia majora
Genital tubercle	Glans penis	Clitoris

CHROMOSOMAL ABNORMALITIES

Chromosomal abnormalities include:

1. Trisomy

Chromosomal abnormalities usually arise during gametogenesis due to *non-disjunction* of sex chromosomes (Fig. 9.1-6). The presence of extra X or Y chromosome (Trisomy) (Fig. 9.1-7) gives rise to many syndromes; associated with abnormal development, mental retardation and abnormal growth. Trisomy is presented as:

(a) Individual with XXY pattern of chromosomes (Klinefelter syndrome) is an abnormal male due to presence of Y chromosome. It is the most common sex chromosome disorder, has an incidence of 1 in 500 males.

Characteristic features in a person with Klinefelter syndrome are (Fig. 9.1-8):

- Poor development of testis with hyalization of seminiferous tubules, leading to sterility. Therefore, this disorder is also known as *seminiferous dysgenesis.*
- Patient has normal male internal and external genitalia.
- Patients are usually tall (due to growth of lower body segment) and obese.
- Gynaecomastia (development of breast in male).
- The secondary sex characters are poorly developed.

(b) Individual with XXX (genotype) pattern of chromosomes is referred to as 'superfemale'. However, there is nothing super about them, because there is poor sexual development (infantile), scanty menstruation and mental retardation.

- Other associated features include:
  - Low or normal plasma testosterone level,
  - High plasma level of gonadotropins (LH and FSH),
  - High plasma level of oestradiol and

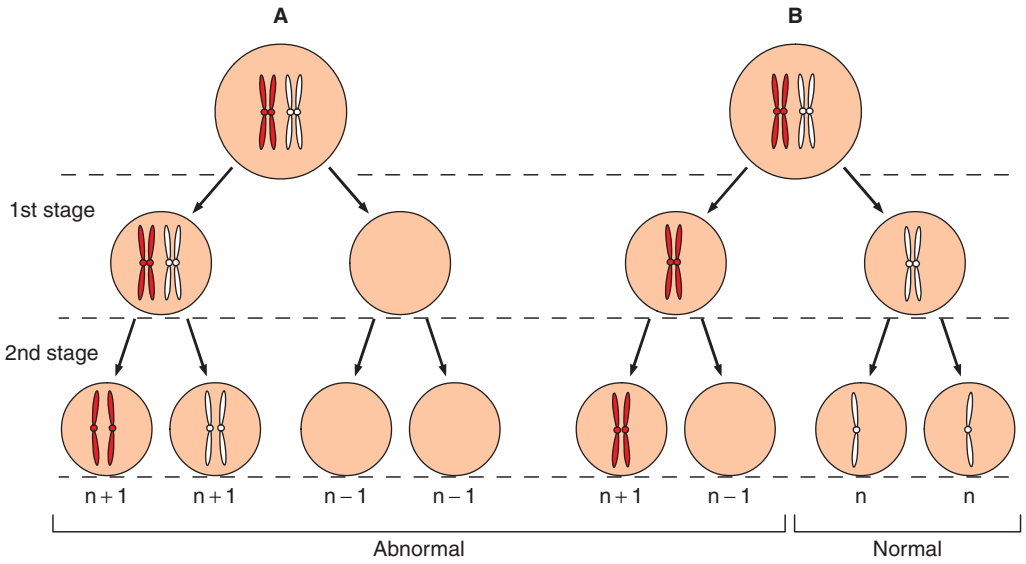


Fig. 9.1-6 Non-disjunction of sex chromosomes during meiotic division; A, in Ist stage and B, in 2nd stage.

– Positive sex chromatin test (as genetic chromosomal pattern is of female).

Note. *Down syndrome* (also known as mongolism) is an example of autosomal chromosomal trisomy of (21 chromosome).

2. Monosomy

As shown in Fig. 9.1-6, when both chromosomes of a pair go to one gamete the other gamete resulting from such a division has only 22 chromosomes; and at fertilization the zygote formed will have only 45 chromosomes. Hence one pair is represented by single chromosome, so it is called monosomy. Depending upon the presence of X or Y chromosome, there will be either female phenotype (44+XO) or male phenotype (44+YO). The best known example of monosomy disorder is *Turner's syndrome* and other example is individual with 44+YO karyotype.

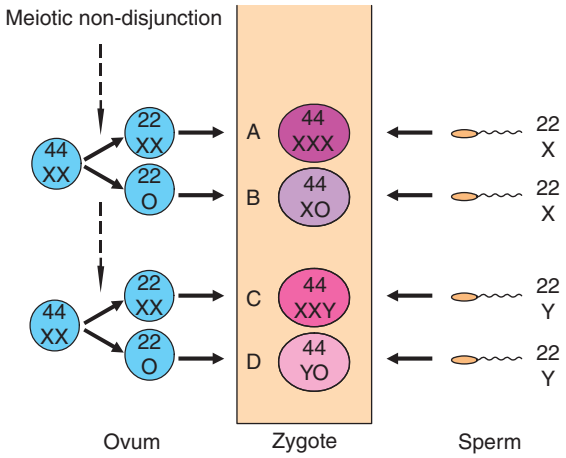


Fig. 9.1-7 Defects due to maternal non-disjunction of sex chromosomes at the time of meiosis: A, superfemale; B, gonadal dysgenesis (Turner's syndrome); C, seminiferous dysgenesis (Klinefelter syndrome) and D, lethal monosomy.

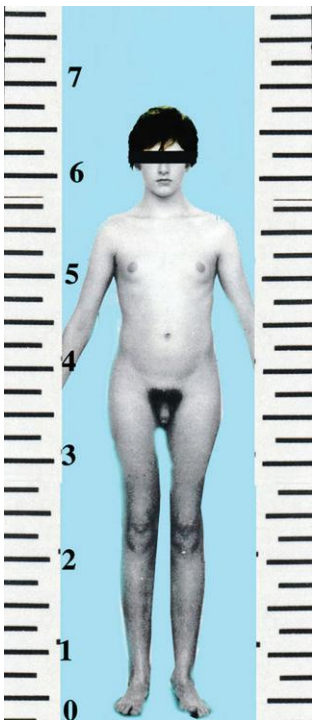


Fig. 9.1-8 Photograph of a patient with Klinefelter syndrome showing tall stature, famine stigmata, bilateral gynaecomastia and small size external genitalia.

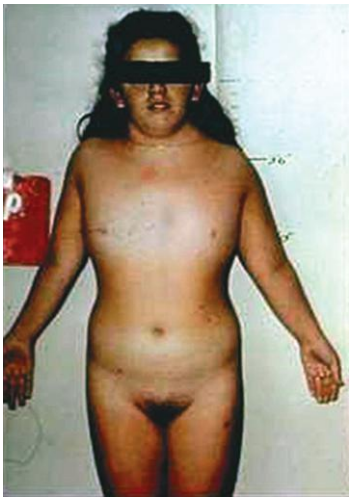


Fig. 9.1-9 Photograph of a patient with Turner's syndrome (ovarian dysgenesis) showing short stature, webbed neck and underdeveloped secondary sexual characters (sexual infantilism).

Turner's syndrome. Characteristic features of Turner's syndrome are (Fig. 9.1-9):

- Patient's chromosomal pattern (karyotype) is 44+XO; Y chromosome is absent hence patient is phenotypically female.
- There is ovarian dysgenesis because of XO karyotype.
- Puberty is delayed. Though there is female type of sexual development but it is characterized by scanty menstruation, amenorrhoea (no menstruation), primary infertility and amastia.
- Other important associated feature is mental retardation.
- Among skeletal abnormalities, dwarfism is very common. The characteristic features are webbed neck (folds of skin on the side of the neck present), face is peculiar with low hair line, ptosis (drooping of eyelids), epicanthus (low set ears), micrognathia (small jaw) and *co-arctation* of aorta.

Diagnosis of chromosomal abnormalities

Early diagnosis (in utero) of these types of disorders is very important. It is made possible by following techniques:

1. Amniocentesis. In this procedure, amniotic fluid is collected by inserting a needle into the amniotic cavity through anterior abdominal wall. The fetal cells present in the amniotic fluid are examined.

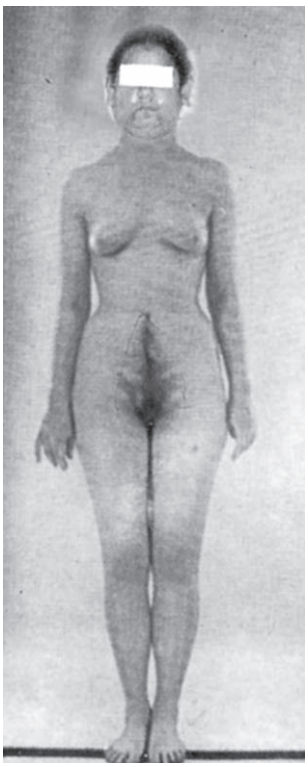


Fig. 9.1-10 Photograph of a patient with female pseudohermaphroditism (congenital virilizing adrenal hyperplasia) showing partial masculinization (diamond-shaped pubic hair).

2. Chorionic villus sampling. In early pregnancy, the fetal cells are obtained by a needle biopsy of chorionic villi.

HORMONAL ABNORMALITIES

The most common developmental disorder due to hormonal abnormalities is pseudohermaphroditism.

Pseudohermaphroditism

Pseudohermaphroditism means individual having genotype (gonads) of one sex (either testes or ovaries) and genitalia of other sex. It occurs in two forms:

- Female pseudohermaphroditism and
- Male pseudohermaphroditism.

A. Female pseudohermaphroditism

There is exposure to increased levels of androgens to the genetic female fetus.

Characteristic features of female pseudohermaphroditism are (Fig. 9.1-10):

- Genotypically, the individual is female (XX).
- Gonads and internal genitalia are feminine like (ovaries, oviduct and uterus are present), but at pre-pubertal age masculinization occurs in the form of diamond-shaped pubic hair growth and development of penis.
- Increased plasma levels of testosterone and androgens.

B. Male pseudohermaphroditism

In this condition, person is genetically male (XY) but have feminisation (female internal and external genitalia). Male pseudohermaphroditism results in following conditions:

1. Androgen resistance means androgen levels are normal but cannot exert their full effect on the target tissue. The effect varies from mild defect to complete loss of responsiveness of receptors to androgens.

- *In mild defect*, patient is infertile and may or may not be associated with gynaecomastia.
- *In case of complete loss of responsiveness of androgen receptors* to androgens, patient presents with *testicular feminising syndrome*. In this condition, MIS is present and testosterone is secreted at normal or at high rate. The patient presents with following features:
  - The external genitalia are of female type but vagina ends blindly.
  - There is no female internal genitalia because testicular hormone suppresses Mullerian duct derivatives (no uterus and oviducts), thus at puberty there occurs primary amenorrhoea due to lack of uterus.

This condition cannot be diagnosed until patient seeks consultation for primary amenorrhoea.

2. Defective testicular development. It leads to deficiency of MIS or MRF (Mullerian regression factor), which is responsible for feminization in a genetic male individual.

3. Congenital 17α hydroxylase deficiency. This enzyme converts adrenal androgens into testosterone. Thus its deficiency causes feminization due to deficient testosterone.

4. Congenital blockade of pregnenolone formation (see page 593). This congenital blockade of pregnenolone formation is associated with male pseudohermaphroditism.

True hermaphroditism

It is a very rare condition, in which gonads of both sexes are present (an ovary on one side and testis on other side), thus resulting in numerous variations in phenotypic (internal and external genitalia) differentiation (Fig. 9.1-11).

Sex chromatin test may or may not be positive.

The true hermaphroditism results due to combined abnormalities.

PUBERTY AND ADOLESCENCE

INTRODUCTION

Puberty and adolescence are the phases of growth between childhood and adulthood.

- Puberty refers to the stage of gonadal development and maturation to the point where reproduction is possible for the first time.
- Adolescence refers to the period of sudden spurt of physical growth between childhood and adulthood (for details page 963).

Since these two phases (adolescence and puberty) of growth are overlapping, hence the terms are interchangeable. The total period of growth spurt ranges between 3 and 5 years. It starts from the age of 8 years. The average age of onset of puberty is 12 years in girls and 14 years in boys.

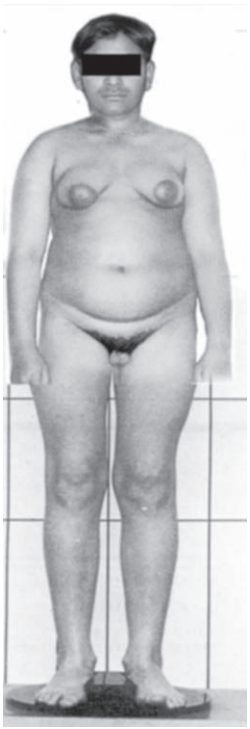


Fig. 9.1-11 Photograph of a patient with true hermaphroditism showing breast development and underdeveloped male genitalia.

COMPONENTS OF PUBERTY

The two principal components of puberty are: sudden spurt of physical growth and appearance of secondary sex characters.

1. Sudden spurt of physical growth

During sudden spurt of physical growth, there is increase in height, muscle mass and muscle strength of an individual. The height increases by 7–12 cm in boys and about 6–11 cm in girls. The increase in height is mainly of the trunk part rather than of limbs.

- The muscle mass and muscle strength also increases in both the sexes but the increase is far greater in boys as compared to in girls.

2. Appearance of secondary sex characters

Stages of development of secondary sex characters. The sequence of events of puberty which occurs in 3–5 years period have been discussed in five stages (Table 9.1-2).

Types of secondary sex characters. The secondary sex characters are almost fully developed by the stage 5 of the puberty both in male and females. These can be grouped as (Table 9.1-3):

- Structural,
- Functional and
- Psychological.

HORMONAL CHANGES DURING PUBERTY

Besides ovaries and testes, other endocrinal glands (adrenal, thyroid and anterior pituitary) also grow in size and

Table 9.1-2 Sequence of events during puberty in male and female

Stage of puberty	Females		Males	
	Bone age in years	Characteristics	Bone age in years	Characteristics
Stage 1	Up to 7½	Pre-adolescent age	7½ years	Pre-adolescent age
Stage 2	10½	Appearance of breast buds (thelarche)	12	Genital development begins (enlargement of testis)
Stage 3	11½	(i) Axillary and pubic hair appear (pubarche) (ii) Enlargement of breast (elevation) (iii) Sudden increase in height (height spurt)	14	(i) Axillary and pubic hairs start appearing (ii) Enlargement of penis
Stage 4	13	(i) Menstruation starts (menarche) (ii) Breast areola begins to elevate and project	15	(i) Further growth of testis, penis and genitalia (ii) Sudden increase in height spurt
Stage 5	14	(i) Adult genitalia (ii) Secondary sex characters	16½	Adult genitalia and secondary sex characters

Table 9.1-3 Secondary sex characters in female and male

Group	Female	Male
A. Structural		
(i) Body configuration	Narrow shoulders, broad hips (broad pelvis) Thighs converge Arms diverge (wide carrying angle)	Shoulders are broader than pelvis
(ii) Skin	Skin is smooth and light	Skin is thick, dark and oily (sebaceous glands secretion thickens and predisposing to acne)
(iii) Hair growth on:		
• Body	Body hair fine and scanty	Body hair rough and dark
• Face	–	– Moustaches and beard appeared
• Scalp	Thick growth, frontal hairline rounded	– Frontal hairline indented at the side
• Pubic region	Concave	Convex and extends towards umbilicus (triangle with apex up)
(iv) Muscularity	Muscles are soft (+)	Muscle bulk and strength is far greater (+++)
(v) Subcutaneous fat	Female distribution of fat due to deposition of fat in breast and hips, which gives characteristic curves and contours to the body	
(vi) Genitalia and accessory sex organs	Adult type: • Clitoris increases in size, labia majora and minora get enlarged • Breasts are developed • Uterus and vaginal growth increases and their activity starts	Adult type: • Penis and scrotum increase in size and become pigmented, scrotal skin thickens and rugal folds appear • Prostate, seminal vesicles, bulbourethral glands enlarged and their secretion begins
B. Functional		
(i) Voice	No change (remains soft and shrill)	Larynx enlarges and vocal cords get thickened, therefore, voice becomes loud, bass (low pitched) deep and breaks
(ii) Basal metabolic rate (BMR)	Lower	5–10% higher than female
(iii) RBC count and Hb, concentration	Lower	Higher
(iv) Menstrual cycle	Begins	Absent
C. Psychological	Girls are more emotional, shy, introvert and sexually attracted towards males	Behaviour is more aggressive, extrovert, competitive and interested in opposite sex

their activity increases at the onset of puberty. The hormonal changes noticed at the time of puberty are:

1. Gonadotropins. In both sexes, the levels of *gonadotropins:* follicle stimulating hormone (FSH) and luteinizing hormone (LH) secreted from the *anterior pituitary gland* rise slowly from birth of the child up to pre-adolescent age, but at the time of puberty (early teen age) their levels suddenly increase. In pre-pubertal stage, the gonadrotropin secretion is not under the check of gonadal hormones (oestrogen and testosterone).

2. Adrenal androgens. There is an increase in the secretion of adrenal androgens at puberty. The onset of this stage of increase or activation is called adrenarche. It occurs at 8–10 years of age in girls and at 10–12 years of age in boys.

Functions subserved by adrenal androgens at puberty are: Growth of pubic and axillary hair in both sexes, and growth of muscle mass and its strength.

3. Growth hormone. Normally from birth up to pre-pubertal stage, the growth hormone secretion is intermittent (a few peaks every 24 h) but at the time of puberty, though basal level of growth hormone does not rise but there is an increase in the frequency and amplitude of the peaks. It is responsible for generalized growth spurt at adolescence. For details page 964.

4. Thyroid gland secretions (thyroxine) also increase during puberty. Thyroxine is necessary for normal growth and development (see page 555).

5. Gonadal hormones (sex hormones). There is slow increase in secretion of sex hormones in children between the age of 7 and 10 years. But, there is a rapid rise in oestrogen secretion (in girls) and testosterone in boys in early teenage.

CONTROL OF ONSET OF PUBERTY

The exact mechanism of onset of puberty is still not fully understood, but experimental and clinical observations

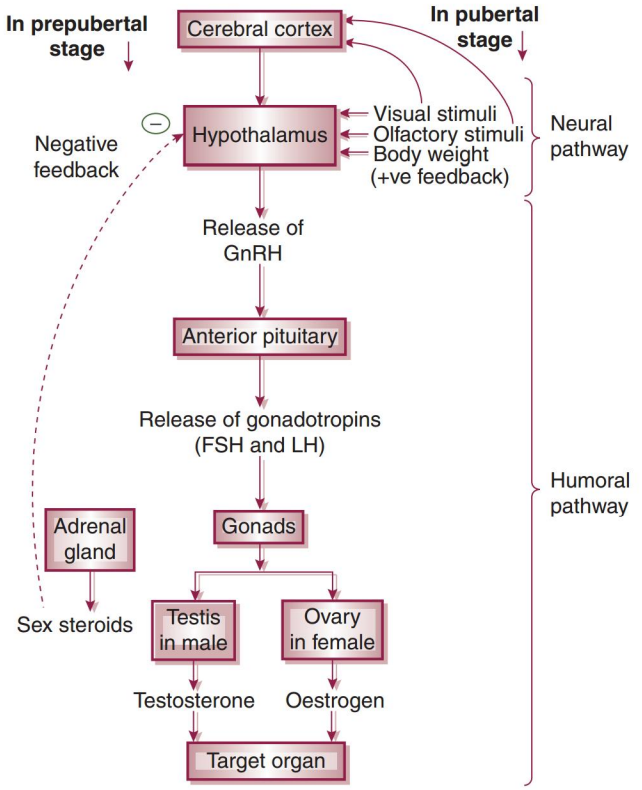


Fig. 9.1-12 Neurohormonal mechanism of onset of puberty.

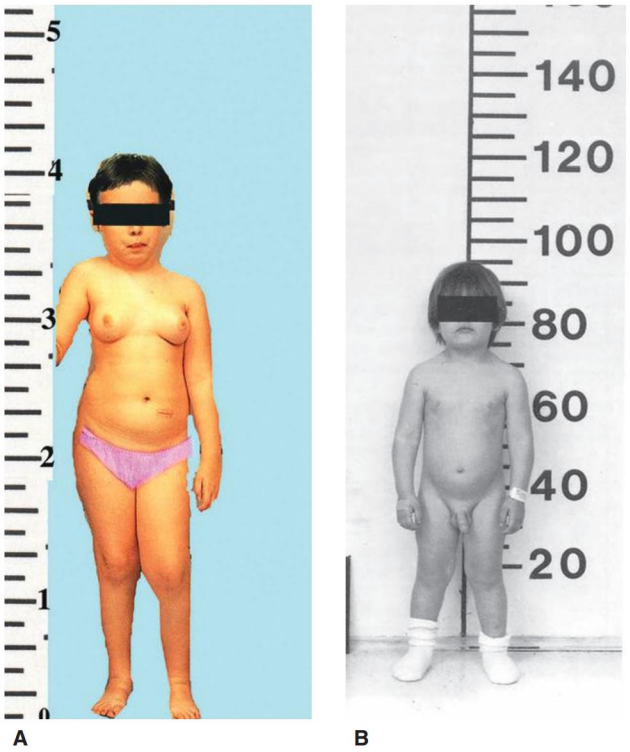


Fig. 9.1-13 A, Photograph of a 7-year-old girl with true precocious puberty. B, Photograph showing precocious puberty in a male child (well-developed secondary sex characters).

support that the *hypothalamus* is intimately involved in this process, being a nodal point between nervous and hormonal circuits. The plausible mechanism involved is as (Fig. 9.1-12):

Role of hypothalamus

- *From birth to pre-adolescent age.* Gonadotropin releasing hormone (GnRH) secretion from the hypothalamus is highly sensitive to feedback inhibition by circulating sex hormones (source being adrenal cortex and prepubertal testes). Therefore, there is very less or slow release of gonadotropins (FSH and LH) from the anterior pituitary.
- *At the time of puberty* (12–14 years), hypothalamic cells become more mature and their sensitivity for circulating sex hormones (for negative feedback) decreases.
- As a result, there occurs an increase in secretion of pituitary gonadotropins (FSH and LH). They act on gonads to release gonadal sex hormones.
- *Awakening of hypothalamus.* Hypothalamus at puberty is also positively stimulated (awakening of hypothalamus) by the critical body mass, visual, external, olfactory and other sensory stimuli.
- *Role of leptin.* Leptin (a Greek word meaning thin) is a circulating protein formed in the fat cells. It acts on the hypothalamus by feedback control mechanism. Therefore, leptin acts as a link between critical body weight and onset of puberty.

DISORDERS OF PUBERTY

Disorders of puberty are related to the time of its onset, e.g.

- Early onset of puberty (precocious puberty) and
- Late onset of puberty (delayed or absent puberty).

1. Precocious puberty

Precocious puberty refers to the onset of puberty in a child before 8 years of age. It is more commonly seen in girls. There is early development of secondary sex characters and gametogenesis also starts earlier (Fig. 9.1-13A & B). Precocious puberty is of two types: true precocious puberty and pseudoprecocious puberty.

True precocious puberty

There is early increased secretion of gonadotropins either due to *decreased inhibition of release of GnRH from the pulse generator* (hypothalamus) or due to chronic stimulation of hypothalamic cells by some irritative focus. Hence the condition is also called *gonadotropin-dependent precocious puberty.*

Pseudoprecocious puberty

In pseudoprecocious puberty, there occurs early development of secondary sex characters without gametogenesis. It occurs due to abnormal exposure of sex hormones to immature child. In this type of precocious puberty, child

may not remain isosexual and normal sequence of events of puberty is also altered.

Causes. Following conditions involving adrenal or gonads result in pseudoprecocious puberty are:

- Congenital virilizing hyperplasia (see page 595)
- Androgen-secreting tumours in males and
- Oestrogen-secreting tumours in females.

2. Delayed or absent puberty

Puberty is considered to be pathologically delayed in case of female, if menarche does not occur by 17 years of age, or in case of male, testicular development and maturation fails to occur by the age of 20 years.

Delayed puberty is more commonly observed in boys than in the girls. Delayed or absent puberty is a matter of great concern when it occurs in following conditions (Fig. 9.1-14):

- *Failure of hypothalamus/pituitary* to secrete gonadotropins, as in panhypopituitarism.
- *Primary gonadal failure.* It refers to the developmental failure or gonadal dysgenesis, which occurs in Klinefelter syndrome in males and Turner's syndrome in females.

Features of delayed or absent puberty are:

- Lack of pubertal development,
- Short stature (dwarf),
- Presence of associated features of other endocrinal abnormalities and
- Low levels of gonadotropins.

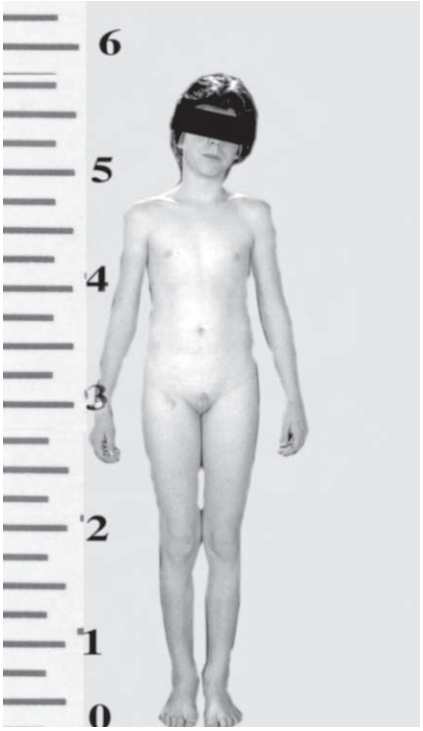


Fig. 9.1-14 Photograph showing characteristic features of delayed puberty due to gonadotropin failure in a 16-year-old boy.

Note. In some cases puberty is absent even when gonads are present and other endocrines are functioning normally. In male, this condition is known as *eunuchoidism* and in female, it is called *primary amenorrhoea*.

Chapter 9.2

Male Reproductive Physiology

AN OVERVIEW OF MALE REPRODUCTIVE SYSTEM

The male reproductive system comprises the internal and external genital organs, which can be functionally organized as (Fig. 9.2-1):

I. Gonads or primary male sex glands

Gonads or primary male sex glands are a pair of testes. The main functions of the testes are to produce sperms and secrete testosterone (male sex hormones).

II. Accessory sex glands

1. Seminal vesicles are two lobulated glands situated on either side of the prostate between the urinary bladder and rectum.

- They secrete a thick alkaline fluid that mixes with the sperms as they pass into the ejaculatory ducts and urethra.
- The duct of each seminal vesicle joins the ductus deferens to form the ejaculatory duct.

2. *Bulbourethral (Cowper's) glands* are two pea-sized glands.

3. *Prostate gland* is the largest accessory gland of the male reproductive system.

- It secretes a thin milky fluid which forms 30% the volume of semen.

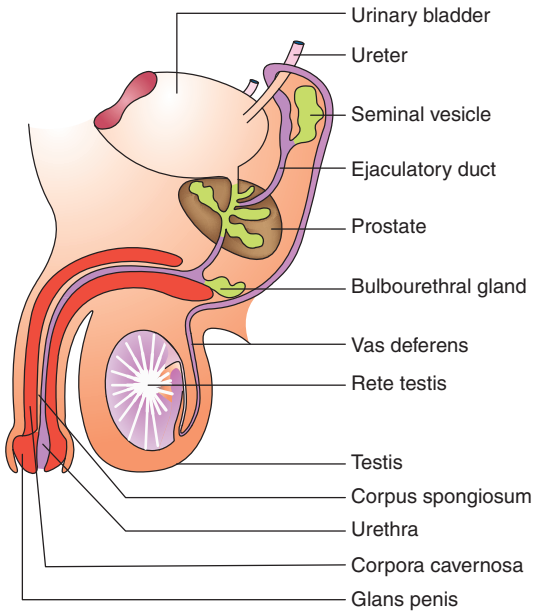


Fig. 9.2-1 Male reproductive system.

III. Ducts of male reproductive system

These include:

1. Epididymis. It is formed by minute convolutions of the duct of the epididymis, so tightly compacted that they appear solid.

- The efferent ductus transports the sperms from the rete testis to the epididymis where they are stored. The sperms can remain viable for a month in the epididymis.

- Secretions of epididymis provide nourishment to the spermatozoa and help them to mature.
- Non-motile spermatozoa become motile after passing through epididymis.

2. Ductus deferens or vas deferens. It is the continuation of the tail of epididymis. It ends by joining to the duct of seminal vesicle. It serves as a secondary storehouse for spermatozoa which will be released at the time of ejaculation.

3. Ejaculatory ducts. Each ejaculatory duct is a slender tube that arises by the union of the ductus deferens with the duct of seminal vesicle. The ejaculatory ducts open as minute slit-like opening into the prostatic urethra.

4. Urethra. The male urethra is a muscular tube (18–20 cm long) that conveys urine from the internal urethral orifice of the urinary bladder to the external urethral orifice at the tip of the glans penis. The urethra also provides an exit for semen (sperms and glandular secretions).

IV. Supporting structures of male reproductive system

1. Spermatic cord. It suspends the testes in the scrotum and contains structures that pass through the inguinal canal to and from the testis viz. ductus deferens, vessels and nerves of the testis.

2. Scrotum. It is a cutaneous fibromuscular sac (can be considered an outpouching of the lower part of the anterior abdominal wall) which houses testes, epididymis and the lower ends of the spermatic cords.

The scrotum maintains the temperature lower than the normal body temperature (about 32°C), which is necessary for normal spermatogenesis.

3. Penis. It is the male copulatory organ and the common outlet for urine and semen. Penis can be divided into three parts: root, body and glans penis. It is composed of three cylindrical bodies of erectile cavernous tissue—the corpora cavernosa and corpus spongiosum.

FUNCTIONAL ANATOMY OF TESTES

GROSS ANATOMY

Location. The testes are ovoid bodies suspended by spermatic cords into the scrotum.

Weight. In an adult male, the average weight of each testis is 25 g (range 10–40 g). Weight of testis decreases in old age.

Coverings. Each testis, from interior to exterior is covered by following three layers (Fig. 9.2-2):

(i) Tunica vasculosa. It is the innermost covering made up of loose connective tissue rich in blood vessels.

(ii) Tunica albuginea. It is also called capsule of the testis consists of closely packed collagen fibres intermingling with many elastic fibres.

(iii) Tunica vaginalis is the outermost covering composed of mesothelial cells.

Coverings of the testis provide protection from trauma and allow free movement of the testis in the scrotum.

Blood supply

The arterial blood supply to the testes is by testicular arteries (arise from abdominal aorta). The testicular artery or one of its branches anastomose with artery of ductus deferens. The venous blood is drained by testicular veins emerging from testes and epididymis and join to form a venous network (pampiniform plexus) consisting of 8–12 veins. The pampiniform plexus is a part of thermoregulatory mechanism which maintains constant temperature (lower than normal body temperature).

Lymphatic drainage of the testes to lumbar (lateral aortic) and pre-aortic lymph nodes.

Nerves (innervation) of the testes

The autonomic nerves of the testes arise as a testicular plexus of nerves on the testicular artery, which contain vagal *parasympathetic fibres* and *sympathetic fibres* from T_7 segment of the spinal cord.

STRUCTURE OF TESTES

Each testis is divided into many lobules by the fibrous septa which project from the mediastinal testis into the tunica albuginea. Each lobule is roughly conical in shape, (Fig. 9.2-2A). Each lobule of the testes consists of:

- Seminiferous tubular compartment
- Interstitial compartment

SEMINIFEROUS TUBULAR COMPARTMENT

The seminiferous tubular compartment of each lobule of the testis contains about 2–3 seminiferous tubules. The seminiferous tubules constitute about 80–90% of the testicular volume.

Each seminiferous tubule is about 80 cm long and 150 μm in diameter. It consists of two parts: the convoluted part and the straight part. The convoluted part forms the loops and continues as two straight ends. Near the apex of the

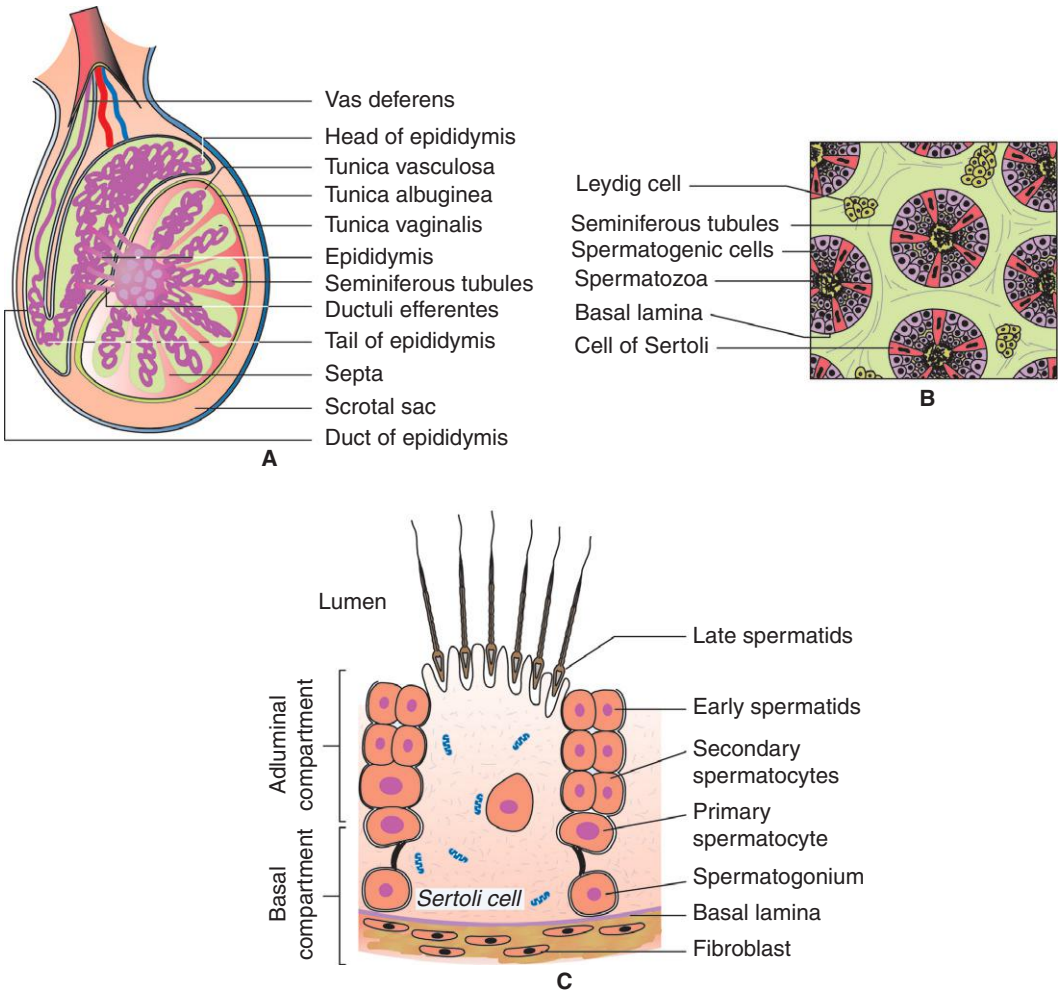


Fig. 9.2-2 Structure of testis: A, lateral view showing the cut section of testis, epididymis and distal part of spermatic cord; B, histology of testis and C, electron microscopic structure of seminiferous epithelium.

lobules, the straight ends join one another to form 20–30 larger straight tubules (tubule recti). The straight tubules pass through the fibrous tissue of mediastinal testis and unite to form a network called *rete testis*. At the upper end of each testis, the rete testis gives of 10–20 efferent ductules which continue into the head of epididymis (Fig. 9.2-2A).

Histological structure of seminiferous tubule

Histologically, the wall of seminiferous tubules is comprised by three layers (Fig. 9.2-2B):

1. Outer capsule or tunica propria. It consists of fibroelastic connective tissue containing few muscle-like cells (myoid cell). The contraction of myoid cells helps in movement of spermatozoa along the wall of the seminiferous tubules.

2. Basement membrane (basal lamina). It is a thin homogeneous lamina lying next to the tunica propria.

3. Epithelial layer of the seminiferous tubules. This is complex stratified epithelium.

The epithelium contains mainly two types of cells:

- The germ cells (spermatogenic cells) and
- Supporting cells or sustentacular cells (Sertoli cells).

Spermatogenic cells

The spermatogenic cells lie in between the Sertoli cells. These are arranged in an orderly manner in 4–8 layers, which extend from the basal lamina to the lumen of the seminiferous tubule.

In a sexually mature individual, the spermatogenic cells of all stages of differentiation are seen, arranged in an orderly manner (Fig. 9.2-2C).

- Basal compartment or deep part of the epithelium is occupied by cells of early stages of spermatogenesis (spermatogonia and primary spermatocytes).
- Adluminal compartment or superificial compartment contains cells of later stages of spermatogenesis, from periphery to lumen, these cells are secondary spermatocytes, early spermatid, late spermatid and spermatozoa.

Sertoli cells

Under electron microscope, each Sertoli cell appears as a slender cell having irregular pyramidal shape. The base of Sertoli cells rest on the basal lamina and each cell stretches from the basal lamina to the lumen of tubule.

Nucleus lies near the base and has prominent nucleolus. The sides and the apices of the cells are marked by recesses, which are occupied by developing cells of different stages of spermatogenesis (spermatogonia, spermatocytes, spermatids and spermatozoa) from basal to luminal region.

Tight junctions. There is no cytoplasmic continuity between the two adjacent Sertoli cells, but plasma membranes of two adjoining cells are connected by tight junctions in the basal region.

The functions attributed to Sertoli cells are:

1. Physical support and nutrition. The Sertoli cells provide physical support to maturing germ cells (the germ cells are present in the recesses on the side walls of the cells), nourish them, (being rich in glycogen contents) and also remove waste products from the germ cells.

2. Phagocytic function. The residual cytoplasmic by products, which are cast off from the spermatozoa during conversion of spermatids into sperms are phagocytized by the Sertoli cells.

3. Maintenance of blood–testis barrier. The tight junction forms effective permeability barrier within the seminiferous epithelium, which is defined in man as blood–testis barrier.

Significance of blood–testis barrier. It limits the transport of many substances from blood to the seminiferous lumen:

- This barrier maintains germ cells in a privileged location, because mature sperm cells are very immunogenic when introduced into the systemic circulation. Thus blood–testis barrier protects the cells of different stages of spermatogenesis from blood-borne toxic substances and from circulating antibodies.
- It prevents the entry of byproducts of gametogenesis into the blood (that is why autoimmune reactions do not occur).

4. Secretory functions. Sertoli cells secrete following hormones and substances:

- *Mullerian duct inhibitory substance* or Mullerian duct inhibition factor,
- *Inhibin,*
- *Androgen-binding protein (ABP)* or androgen-binding globulin,
- *Oestrogen,*
- *Transport proteins,* such as transferrins (the iron transporting protein), and ceruloplasmin (copper binding protein),
- *Plasminogen activator.* It is required for proteolytic activity for the disruption of tight junctions during migration of maturing germ cells from basal to luminal compartment, is formed by Sertoli cells.
- *Seminiferous tubular luminar fluid.* Sertoli cells secrete watery, solute-rich (K^+ and HCO_3^-) fluid into the lumen of seminiferous tubules. The fluid movement provides a driving force for non-motile spermatozoa.

INTERSTITIAL COMPARTMENT

The interstitial spaces between the seminiferous tubules constitute about 10–20% volume of the testis. The interstitial compartment of each lobule is filled by loose connective tissue and Leydig cells. The *Leydig cells* or the so-called interstitial cells are present in groups. They have an endocrine function of secretion of male sex hormone (testosterone).

FUNCTIONS OF TESTES

The two principal functions of testes are:

- Gametogenic function (spermatogenesis) and
- Endocrine function.

SPERMATOGENESIS

Spermatogenesis refers to the process of formation of spermatozoa from the primitive germ cells (spermatogonia).

PHASES OF SPERMATOGENESIS

The phases of spermatogenesis are as follows (Fig. 9.2-3):

1. Phase of mitotic division of spermatogonia. Each spermatogonium divides mitotically five times to form 32 spermatogonia. The division occurs in the basal compartment of the seminiferous tubule.

2. Phase of formation of primary spermatocytes by mitotic division. The 32 spermatogonia (44 + X + Y) undergo mitosis to form 64 primary spermatocytes (44 + X + Y). Primary spermatocytes are large cells with large nucleus having diploid number of chromosomes (2n).

3. Phase of formation of secondary spermatocyte by meiotic division. Each primary spermatocyte undergoes meiotic division:

- After first reduction division (meiosis), the 64 tetraploid primary spermatocytes (4n) are converted into

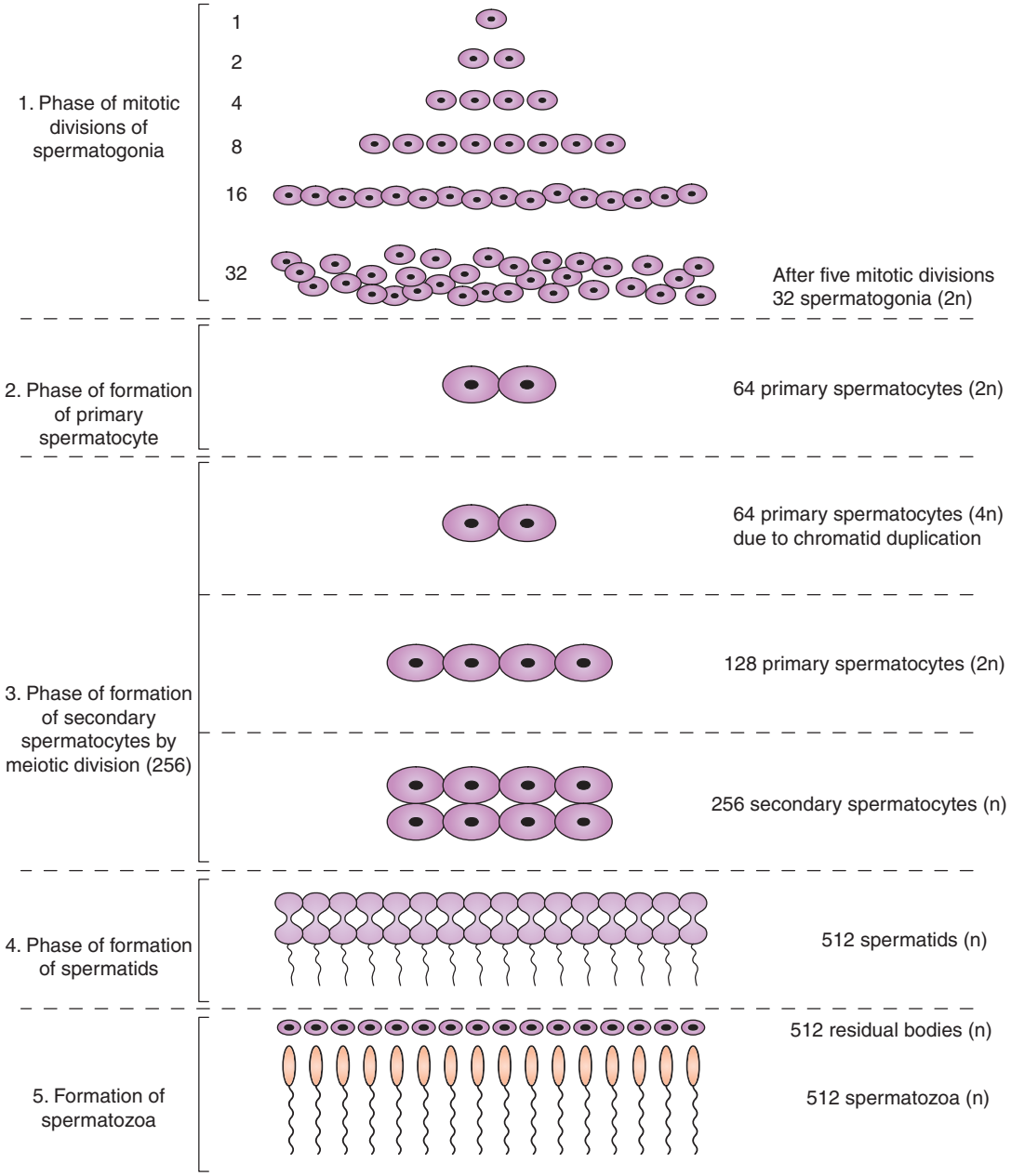


Fig. 9.2-3 Phases of spermatogenesis.

128 primary spermatocytes with diploid number of chromosomes (2n).

- The 128 primary spermatocytes (meiosis) to form 256 secondary spermatocytes having haploid number of chromosomes (n), i.e. either 22 + X or 22 + Y. Therefore, 50% of sperms will have X chromosome and other 50% will have Y chromosome.

4. Phase of formation of spermatid. Each secondary spermatocyte divides mitotically to give rise to two spermatids. Thus, a total of 512 spermatids are formed from a single spermatogonium.

5. Phase of formation of spermatozoon (spermiogenesis). The spermatids do not divide further but undergo morphological changes to form sperms or spermatozoa. The spermatid undergoes changes in the shape and orientation of its organelles. The spermatids mature into spermatozoa in the deep folds of the cytoplasm of the Sertoli cells.

Structure of spermatozoon

A fully formed spermatozoon is about 55–65 µm in length. It comprises following parts (Fig. 9.2-4):

1. Head. The head is about 4–5 µm long, flattened from anterior to posterior. It is oval when seen from the front. It is surrounded by acrosome.

Acrosome is a thick cap-like structure which covers the anterior two-thirds part of the head. It contains a number of

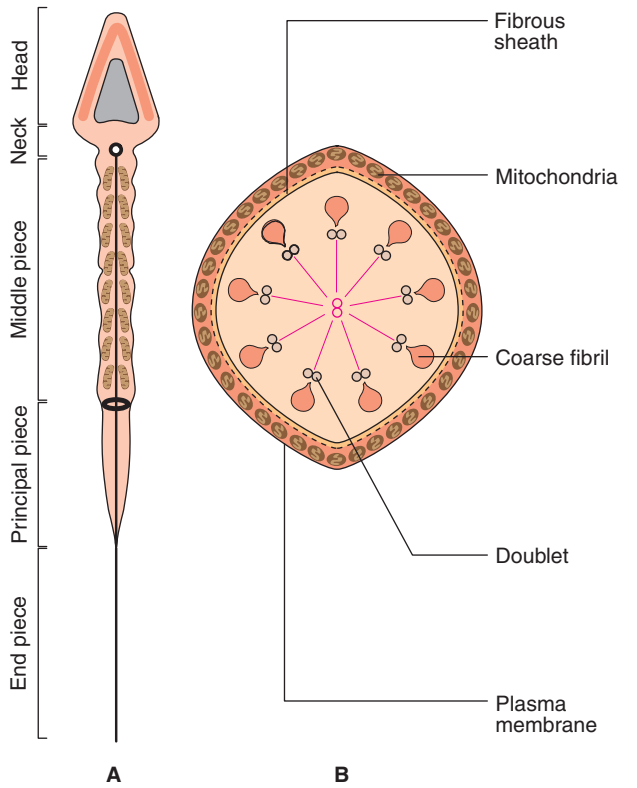


Fig. 9.2-4 Structure of a mature human spermatozoon (A) and transverse section of the middle piece of tail part showing its detail structure (B).

enzymes (hyaluronidase, proteolytic enzymes and acid phosphatase), which help the sperm in penetrating ovum during fertilization.

2. Neck. It is a narrow constricted part. It contains a funnel-shaped basal body and a spherical centriole.

3. Tail of the sperm is the motile portion and is also called the flagellum. It can be divided into three parts:

- Middle piece,
- Principal piece and
- End piece.

Structure. The tail of the sperm consists of following components:

- *Axoneme or axial filament.* It forms the central skeleton of the tail. The axoneme begins just behind in the neck and extends through the entire length of tail. It is made up of nine pairs (doublets) of microtubules arranged in a circle, surrounding central pair.
- *Coarse fibrils.* The nine petal-shaped coarse fibrils are present in the *middle piece* and principal piece, but do not extend into the end piece of tail. Fibrils are contractile components of the sperm.
- *Fibrous sheath* is present outside the coarse fibrils.
- *Mitochondria.* In the region of middle piece, i.e. proximal part of tail, the fibrous sheath is surrounded by spirally arranged mitochondria in abundance. The mitochondria synthesize ATP, which supplies energy for the motility of tail.
- *Plasma membrane* encloses the entire sperm.
- The principal piece contains Ca^{2+} channel known as *Catsper protein,* its activation causes sperm motility.

Storage of spermatozoa

About 120 million sperms are formed each day. A small quantity of them is stored in the epididymis but most of them are stored in vas deferens and ampulla of vas deferens. They can remain stored maintaining their fertility for about a month.

Maturation and capacitation of spermatozoa

Role of epididymis. The fully formed spermatozoa are released into the lumen of seminiferous tubules, from where they reach the epididymis. Epididymis is the site of extra testicular maturation of spermatozoa. When the sperms arrive in the epididymis they are non-motile and acquire some motility only after passing through the epididymis.

Role of seminal vesicles and prostate gland. The secretions of seminal vesicles and the prostate have a stimulating effect on the sperm motility, but the spermatozoa become fully motile only after ejaculation.

Role of female genital tract. Spermatozoa acquire ability to fertilize the ovum only after they have been in the female genital tract for sometime (1–10 h). This final step in their maturation is called capacitation (see page 663).

SEMEN

Semen or the seminal fluid refers to the fluid ejaculated during the orgasm at the time of male sexual act.

Characteristic features

- *Volume.* The average volume of semen per ejaculation is 2.5–3.5 mL after an abstinence of 2 days. Volume of semen decreases with repeated ejaculations.
- *Appearance* of semen is milky due to prostatic secretions.
- *Specific gravity* is about 1.028.
- *Reaction* is alkaline with a pH of 7.5. The alkalinity is due to the prostatic secretions. The alkaline semen brings the vaginal pH from 3.5–4 to 6–6.5, the pH at which sperms show optimum motility.
- *Nature* of the semen when ejaculated is liquid but soon it coagulates in vitro or in the vagina, and finally undergoes secondary liquefaction after about 15–30 min. The clotting of semen soon after ejaculation helps to retain it in the vagina for sometimes. Lysis later on would release the sperms for their free movement into the uterine cavity for fertilization.

Components of semen and their characteristics

The semen comprises following components:

1. Spermatozoa. The normal sperm count varies from 35 to 200 million/mL of semen with an average of 100 million/mL.

2. Secretions of seminal vesicles. Secretions of seminal vesicles contribute 60% of the semen volume:

- The secretion from seminal vesicles is mucoid and viscous fluid.
- It is neutral or slightly alkaline in nature.
- It contains fructose, phosphorylcholine, ergothioneine, ascorbic acid, flavins and prostaglandins.

Functions subserved by the seminal vesicle secretions are:

- *Nutrition to sperms* after being ejaculated into the female genital tract is provided by the fructose and other nutritive substances from the seminal vesicle secretions.
- *Clotting of semen* soon after ejaculation into the female genital tract occurs due to fibrinogen present in the seminal vesicle secretions.
- *Fertilization of ovum* may be enhanced by the prostaglandins present in the seminal vesicle secretion.

3. Secretion of prostate gland. Secretion of prostate gland forms about 10% of the total semen bulk:

- It contributes milky and alkaline fluid part of the semen.

Functions subserved by prostatic fluid component of semen are:

- *Maintenance of optimum pH for fertilization* (6–6.5) is the function of alkaline prostate fluid, which neutralizes the acidity of vaginal secretion. At this pH, the sperms become motile and the chances of fertilization are enhanced.
- *Clotting of semen* by converting fibrinogen (from seminal vesicles) into a coagulum is caused by the clotting enzymes present in the prostatic fluid.

4. Secretion of bulbourethral gland. Secretion of bulbourethral gland and other mucous glands provide mucoid consistency to the semen after puberty.

ENDOCRINE FUNCTIONS OF TESTES

- The Leydig cells of testes produce male sex hormones known as androgens.
- The Sertoli cells of testes secrete oestrogen, *inhibin* and *activin*.

SECRETION AND TRANSPORT OF ANDROGENS

Testes secrete the following androgens (male sex hormones):

1. Testosterone. The most important testicular hormone is testosterone.

- Leydig cells are numerous in newborn male infants and in adult males. So, the androgens are secreted in infancy and after puberty.
- The androgen secretion starts decreasing after 40 years and becomes almost zero by the age of 80 years.
- A normal man secretes 4–9 mg of testosterone daily. Plasma testosterone level in adult males is about 0.65 µg%. More than 98% of secreted testosterone is bound to plasma proteins; 68% is bound to albumin and 30% is bound to testosterone-binding globulin also called sex steroid binding globulin, (SSBG) or gonadal steroid binding globulin, i.e. GBG (because it binds oestradiol as well). A very small percentage of the plasma testosterone is unbound. The free fraction alone is physiologically active in the target tissues.

2. Androstenedione is an important steroid precursor for blood oestrogens in men. It is secreted by the testes at a rate of about 2.5 mg/day.

3. Dihydrotestosterone (DHT) is another important androgen present in the blood.

- Only 20% of the plasma DHT is formed in the testes by the action of 5α-reductase (from the Sertoli cells) on the testosterone (secreted by Leydig cells).
- About 80% of the plasma DHT is derived from the peripheral conversion of testosterone.
- DHT has more than twice the biologic activity of testosterone.

Adrenal cortex also secretes androgens normally testosterone, androstenedione and dehydroepiandrosterone (of these last one is more important). The actions of adrenal androgens are unimportant under normal physiological conditions, because their quantity is insignificant.

SYNTHESIS OF ANDROGENS

Salient points about synthesis of androgens:

- Androgens (C-19 structure) are synthesized in the Leydig cells from the cholesterol (C-27 structure).
- The key step in the synthesis of androgens is the conversion of cholesterol to pregnenolone.
- The CYP 11 Al enzyme is the rate-limiting enzyme for steroid synthesis in all steroid producing tissues.
- In the developing male fetus, the stimulus for testosterone synthesis is human chorionic gonadotropin (HCG), which is the placental hormone secreted in highest amounts during the first trimester of pregnancy.

Biochemical pathways of synthesis of androgens

The enzymatic processes involved in the conversion of cholesterol to androgens are shown in Fig. 9.2-5.

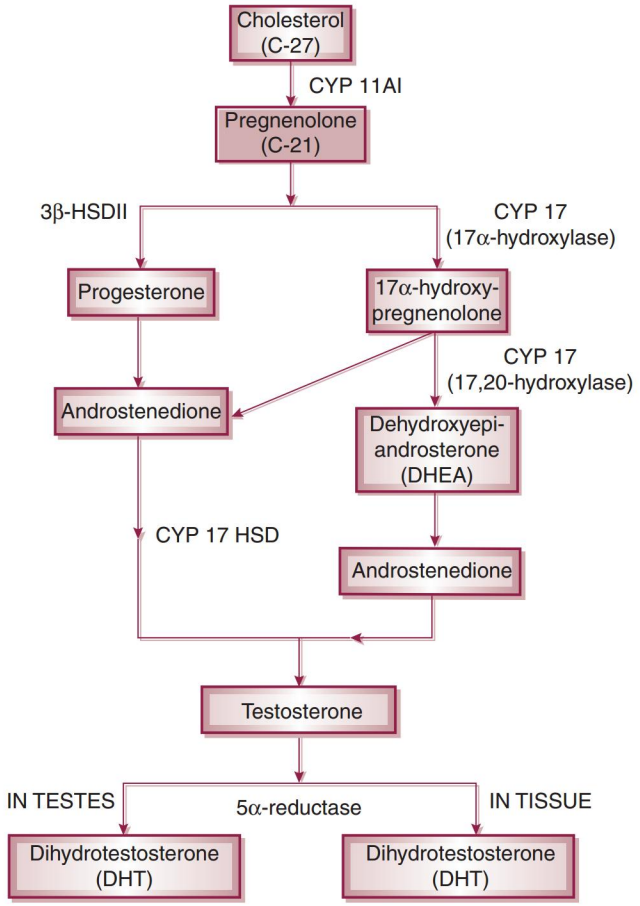


Fig. 9.2-5 Pathway of formation of androgens from cholesterol.

METABOLIC DEGRADATION AND EXCRETION OF ANDROGENS

Within 25–30 min of entry of testosterone into circulation, it is either fixed to the target tissues or degraded in the liver. Testosterone can have one of the following fates:

1. Conversion into dihydrotestosterone (DHT) occurs in skin and male reproductive tract (prostate gland, seminal vesicles and epididymis) by the enzyme 5α-reductase. Dihydrotestosterone crosses the cell membrane of the target cells and produces its effects.

2. Conversion into oestradiol and oestrone. Circulating testosterone and androstenedione can be converted to oestradiol and oestrone, respectively, by the action of aromatase (Fig. 9.2-6).

3. Conversion to 17-ketosteroid. The degradation products of testosterone include androsterone, epiandrosterone and etiocholanolone. These three products are grouped together as 17-ketosteroids.

These are conjugated with glucuronic acid or sulphuric acid in the liver and excreted in the bile and urine.

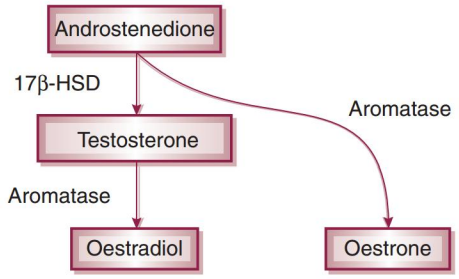


Fig. 9.2-6 Conversion of androgens to oestradiol and oestrone.

FUNCTIONS OF ANDROGENS

A. Functions of androgens in fetal period (in utero)

The testosterone is secreted by the fetal testes at about 2nd–4th month of embryonic life. The functions of androgens in fetal period are:

1. Effect on sex differentiation in fetus

Gonadal sex differentiation (as discussed on page 624) is dependent on the genotype of the embryo.

Genital differentiation (see page 625).

2. Effect on descent of testes

The testes developed in the abdominal cavity are pushed into the scrotum through inguinal canal just before birth. Testosterone is necessary for this descent of testes.

B. Functions of androgens at puberty

1. Effects on external genitalia

Testosterone causes pubertal enlargement of penis. Scrotum increases in size and becomes pigmented. Rugal folds appear in scrotal skin.

2. Effects on accessory sex organs

- Testosterone (with or without DHT) causes enlargement of seminal vesicles.
- DHT promotes growth of prostate and stimulates prostatic secretions.

3. Development of male secondary sexual characters (see Table 9.1-3)

4. Effects on psyche

- *Psychological differentiation.* A brief exposure of fetal hypothalamus to androgens (from its own testes) during early embryonic period causes male pattern of sexual behaviour during puberty. It occurs due to constant secretion of pituitary gonadotropins.
- *Libido.* During puberty, testosterone initiates sexual drive (libido) and erectile function (potency).

- *Aggressive behaviour.* Testosterone produces aggressive behaviour and interest in the opposite sex.

5. Anabolic and general growth promoting effects

Testosterone causes nitrogen retention in the body (positive nitrogen balance) and causes accelerated growth of the body and skeletal muscles in particular. Androgens increase the rate of linear growth of the bones causing a rapid increase in stature at puberty (pubertal growth spurt).

C. Functions of androgens in adults

1. Hair growth. Androgenic patterns of hair growth are maintained. With increasing age, male baldness may be initiated (see page 593).

2. Psyche. Behavioural attitudes and sexual potency are maintained in post-pubertal adults.

3. Bone. Bone loss and osteoporosis are prevented by the androgens in adult males.

4. Spermatogenesis is maintained in adulthood by testosterone along with follicle-stimulating hormone (FSH). The testosterone, acts on both Sertoli cells and germ cells and thus maintains spermatogenesis.

5. Haematopoiesis. Testosterone stimulates erythropoiesis. Therefore, accounts for the greater haemoglobin concentration and RBC count in males.

6. Effects on circulating and stored body fats. Testosterone increases circulating levels of low-density lipoproteins cholesterol and decreases plasma high-density lipoproteins cholesterol.

7. Regulation of gonadotropin secretion. Androgen suppression of luteinizing hormone releasing hormone (LHRH) and luteinizing hormone (LH) by negative feedback effect.

CONTROL OF TESTICULAR FUNCTIONS

The two main functions of testes, viz. spermatogenesis and secretion of testosterone, are controlled by the hypothalamic–hypophyseal–testicular axis.

CONTROL OF SPERMATOGENESIS

The hypothalamic–hypophyseal–testicular (seminiferous tubular) axis controlling the spermatogenesis is as follows (Fig. 9.2-7):

I. Stimulatory control

1. Role of hypothalamus. At puberty, hypothalamic cells become more mature and their sensitivity for circulating sex hormones (negative feedback) decreases so much that there is a pulsatile release (8–14 pulses/day) of gonadotropin-releasing hormone (GnRH) from the hypothalamus. The GnRH stimulates anterior pituitary to secrete LH and FSH.

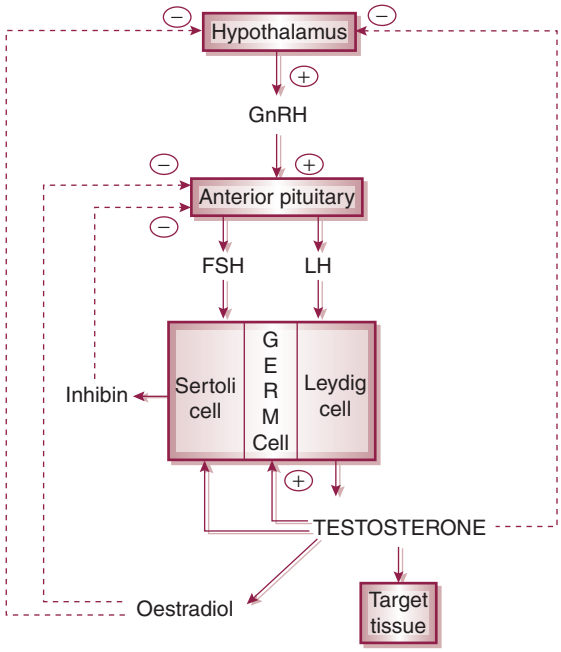


Fig. 9.2-7 Stimulatory and feedback inhibitory control of secretion of spermatogenesis.

2. Role of anterior pituitary. The anterior pituitary controls spermatogenesis through the gonadotropic hormones (FSH and LH) and growth hormone. Neither FSH nor LH acts on the spermatogonia. Yet, normal spermatogenesis requires both FSH and LH as described:

- Follicle-stimulating hormone *stimulates cells of Sertoli,* which play following roles during spermatogenesis:
  - Sertoli cells help in conversion of spermatids to sperms.
  - They secrete ABP, which stabilizes the high supply of testosterone to the developing germ cells in the seminiferous tubular lumen.
  - FSH also promotes the synthesis of inhibin by Sertoli cells.
- Follicle-stimulating hormone indirectly affects testosterone synthesis by increasing the number of LH receptors on the Leydig cells.
- *Role of LH (LH also called interstitial cell stimulating hormone, i.e. ICSH).* The LH stimulates Leydig cells to cause testosterone secretion. The testosterone is required for normal spermatogenesis.
- *Role of growth hormone.* Growth hormone specifically promotes early division of the spermatogonia themselves. In its absence, as in pituitary dwarfs, spermatogenesis is severely deficient or absent.

3. Role of testicular hormones. Testosterone secreted by the Leydig cells by a paracrine effect acts on both Sertoli cells and germ cells and thus maintains spermatogenesis. It is important to note that:

Oestrogen formed from the testosterone by Sertoli cells (when stimulated by FSH) is probably also essential for spermatogenesis.

IMPORTANT NOTE

- Exogenous testosterone cannot promote spermatogenesis in men lacking Leydig cells, because it is not possible to achieve the required high concentration by parenteral therapy.
- Dihydrotestosterone is not required for normal spermatogenesis; therefore, normal sperm development occurs in men with five α reductase deficiency.

II. Feedback inhibitory control

The spermatogenesis is controlled by following negative feedback mechanisms (Fig. 9.2-8):

- Inhibin secreted by the Sertoli cells acts directly on the anterior pituitary and inhibits the secretion of FSH.
- Testosterone and oestradiol inhibit LH secretion by negative feedback mechanism:

Oestradiol exerts the negative feedback effect at both the hypothalamic (GnRH) and pituitary (LH) levels.

Testosterone has its feedback effect mainly at the hypothalamic (GnRH) level.

Note. Plasma physiologic level of testosterone does not produce significant feedback inhibition of FSH secretion.

CONTROL OF TESTOSTERONE SECRETION

In fetus

During fetal life, HCG secreted by placenta stimulates the development of Leydig cells in the testes of fetus and causes testosterone secretion.

In adults

I. Stimulatory control

The hypothalamic–hypophyseal–testicular (Leydig cell) axis controls the secretion of testosterone in adults as (Fig. 9.2-8).

Hypothalamus produces GnRH, which stimulates anterior pituitary to secrete FSH and LH.

Anterior pituitary controls secretion of testosterone (steroidogenesis) primarily through LH.

Leydig cells of testes have LH receptors located on their plasma membrane. Luteinizing hormone binds to these receptors to activate cyclic AMP synthesis, which triggers testosterone synthesis and its secretion.

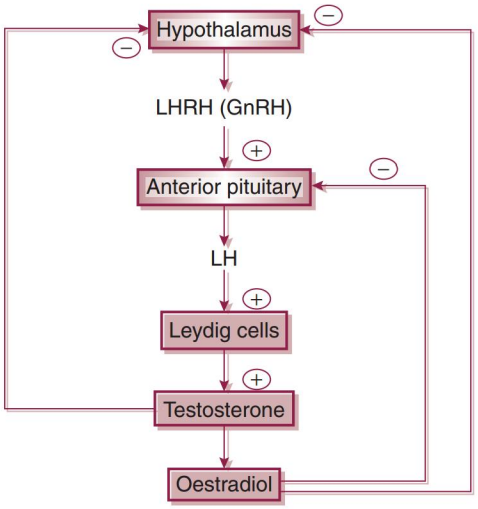


Fig. 9.2-8 Stimulatory and feedback inhibitory control of secretion of testosterone.

II. Feedback inhibitory control

Plasma testosterone level is maintained at a constant level by a feedback control exerted by testosterone and oestradiol independently to control LH (Fig. 9.2-8):

- Testosterone negative feedback is exerted mainly on the opioidergic neurons that project to GnRH (LHRH) neurons.
- Oestradiol exerts negative feedback effects at both hypothalamic and pituitary levels.
- Oestradiol formed by local hypothalamic conversion of testosterone also inhibits LHRH (GnRH).

APPLIED ASPECTS

Some of the important applied aspects, in relation to male reproductive physiology are:

- Cryptorchidism,
- Extirpation,
- Hypogonadism in males and
- Hypergonadism in males.

CRYPTORCHIDISM

The testes develop in relation to the lumbar region of the posterior abdominal wall. During fetal life, they gradually descend to the scrotum by the end of the eighth month of gestation. *Cryptorchidism* refers to a condition in which the descent of the testes may fail to occur or may be incomplete.

Characteristic features of cryptorchidism are:

- The undescended testes may lie in the lumbar region, in the iliac fossa, in the inguinal canal, or in the upper part of scrotum.
- Spermatogenesis often fails to occur in cryptorchidism (due to high temperature of the abdominal cavity) resulting in sterility.

Treatment. Cryptorchidism should be treated as early as possible to prevent male sterility. Surgical correction is advised for correction of undescended testes. However, in some children administration of testosterone or gonadotropic hormone (which stimulate the Leydig cells) can cause the testis to descend provided the inguinal canal is large enough to allow passage of testis.

EXTIRPATION

Extirpation or castration refers to the removal of testes. It will produce following effects:

Effects of extirpation of testes before puberty

The removal of testes before puberty results in a clinical condition, which is known as *eunuchoidism*. It is characterized by:

Permanent sterility, as there is no testis so there are no sperms.

Underdevelopment of external genitalia (i.e. penis and scrotum) and accessory sex organs (i.e. seminal vesicles and prostate gland).

Underdevelopment of secondary sexual characters, that is:

- *Hair growth* on face, trunk and in axilla is scanty.
- *Voice is high pitched* like that of child due to underdevelopment of larynx.
- Muscle mass and shoulder girdle development is poor.
- *Female like body configuration* occurs due to abnormal deposition of fat on buttocks, hips, pubis and breasts.

Abnormal bone growth due to delay in the union of epiphysis may lead to increase in height of the individual, but the bones are weak and thin.

Effects of extirpation of testes after puberty

Under such circumstances, some of the male secondary sexual characters and accessory organs (which depend on testosterone) not only for development but also for maintenance are depressed, while some of the masculine features are retained as:

- *Accessory sex organs* are depressed, i.e. seminal vesicles and prostate undergo atrophy.
- *Penis* remains normal in size.
- *Sexual desire* and sexual activity is slightly impaired. Erection occurs but ejaculation is rare because of atrophy of accessory sex organs and lack of sperm.
- *Voice*, usually remains masculine as the growth of larynx is completed during adolescence.
- *Other secondary sexual characters* are not lost.
- *Other body functions* including life span, senility, intelligence, etc. are not affected.

HYPOGONADISM IN MALES

Causes. Hypogonadism in males results from absent or deficient testicular functions, which may occur in following conditions:

- Congenital non-functioning of testes,
- Underdeveloped testes,
- Cryptorchidism (undescended testes),
- Extirpation of testes and
- Absence of androgen receptors in testes.

Effects of male hypogonadism depend upon whether the testicular deficiency occurs before or after puberty.

If it occurs before puberty it leads to:

- Permanent sterility
- Underdevelopment of external genitalia
- Underdevelopment of secondary sexual characters
- Abnormal bone growth

If it occurs after the onset of puberty, it leads to atrophy of accessory sex organs.

Frohlich's syndrome also known as adipose genital syndrome or hypothalamic eunuchoidism refers to hypogonadism, which occurs due to:

- Hypothalamic disorders,
- Pituitary disorders, or
- Genetic inability of hypothalamus to secrete LHRH, i.e. GnRH.

Features. In this condition, hypogonadism is often associated with abnormal stimulation of feeding centre. Therefore, the affected person overeats and consequently, obesity occurs along with eunuchoidism.

HYPERGONADISM IN MALES

Hypergonadism in males results from excessive secretion of male sex hormones (androgens) as occurs in tumours of Leydig cell. It is characterized by:

- Rapid growth of musculature and bones.
- But, the height is less due to early closure of epiphysis.
- There is excessive development of sex organs and secondary sexual characters at an early age.
- The tumours can also secrete oestrogenic hormones, which can cause overgrowth of breasts (gynaecomastia).

Chapter 9.3

Female Reproductive Physiology

AN OVERVIEW OF FEMALE REPRODUCTIVE SYSTEM

The female reproductive system comprises internal and external genitalia which can be organized as (Fig. 9.3-1):

I. Primary sex organs or ovaries

The primary sex organs are a pair of ovaries which correspond with testes in males. The main functions of ovaries are:

- To produce ova
- To secrete female sex hormones

II. Accessory sex organs

The accessory sex organs of females include internal genital organs and external genitalia.

FEMALE INTERNAL GENITALIA

The internal genital organs include uterus, fallopian tubes and vagina.

I. Uterus

Uterus is a hollow, thick-walled muscular organ, situated between the urinary bladder and the rectum. It can be divided into two parts (Fig. 9.3-2):

1. Body of the uterus. It forms the upper 2/3rd part of the uterus. Its lower limit is marked by a constriction which corresponds to narrowing of uterine cavity at *internal os.* Body of the uterus can be divided into two parts:

Fundus is the rounded part of the body that lies superior to the opening of the fallopian tubes.

Isthmus is the relatively constricted region of the body (approximately 1 cm long) just above the cervix.

2. Cervix of the uterus. It is the cylindrical lower part which protrudes into the upper most vagina. It is approximately 2.5 cm long in an adult non-pregnant woman. Its cavity extends from the *internal os* to *external os* which opens into the vagina.

Structure of uterus

The wall of body of uterus is consisting of three layers:

1. Perimetrium is the external serosal layer.

2. Myometrium is the middle muscular layer comprising bundles of smooth muscles amongst which there is connective tissue.

- The muscle fibres run in various directions and distinct layers are difficult to define.
- The muscle cells of the uterus are capable of undergoing great elongation in association with the great enlargement of organ in pregnancy.
- Contractions of myometrium are responsible for the expulsion of the fetus at the time of child birth.

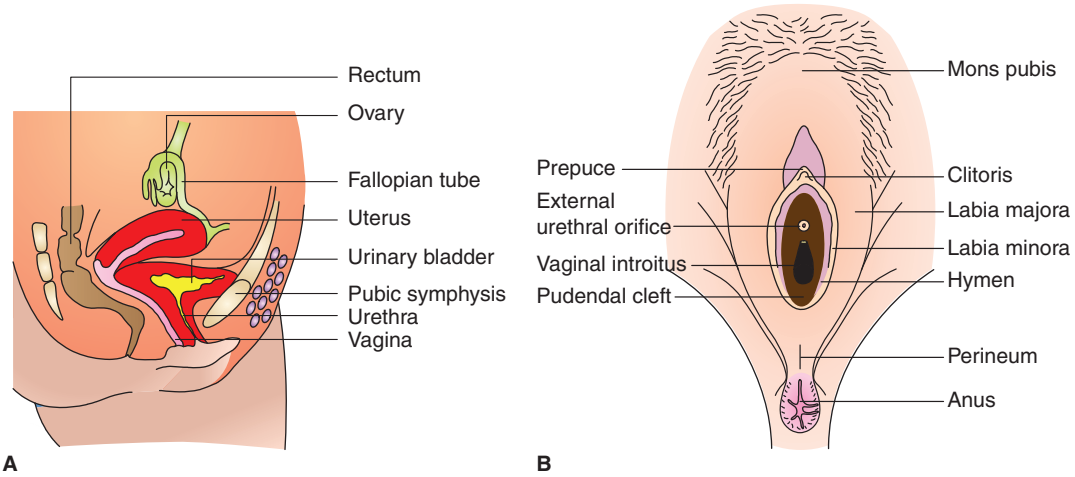


Fig. 9.3-1 Female reproductive organs: A, lateral view showing position of internal reproductive organs in relation to pelvic viscera and B, female external genitalia.

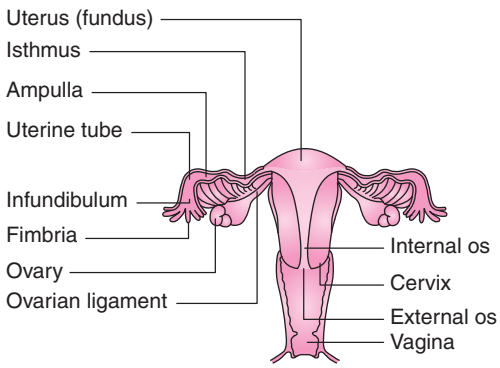


Fig. 9.3-2 Parts of the uterus and fallopian tube.

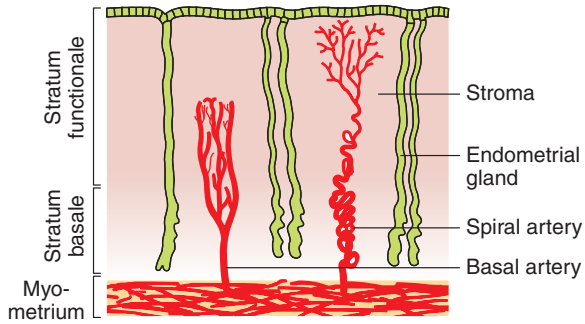


Fig. 9.3-3 Histological structure of the endometrium.

3. Endometrium is inner layer of uterus which consists of epithelial lining and the stroma (Fig. 9.3-3):

- *Epithelial lining* is made up of columnar cells. Before menarche (i.e. the age of onset of menstruation) the cells are ciliated, but thereafter most of the cells may not have cilia.
- *Stroma* of the endometrium is highly cellular and contains numerous blood vessels and numerous simple tubular uterine glands, which are lined by the columnar epithelium.

Functional divisions of endometrium. Functionally, the endometrium of body of uterus can be divided into two strata:

1. Stratum functionale includes the superficial two-thirds thickness of endometrium, which undergoes monthly cyclic changes in preparation for the implantation of fertilized ovum and is shed during menstruation. This portion of endometrium is supplied by long and spiral (coiled) arteries.

2. Stratum basale is the deeper one-third layer of endometrium. It does not participate in the cyclic changes but functions as a regenerative layer. This part of endometrium is supplied by short and straight *basal arteries.*

Structure of cervix

Structure of cervix of the uterus is somewhat different from that of the body.

1. Perimetrium is the outermost serous layer.

2. Myometrium layer of the cervix is much less muscular as compared to the body of uterus and contains more connective tissue. During child birth, when the myometrium of body of uterus contracts, the myometrium of cervix dilates, consequently the cervical canal becomes large enough for the fetal head to pass through.

3. Endocervix refers to the innermost mucosal layer of cervix in contrast to the endometrium of the body of uterus. Endocervix is not shed at the time of menstruation. Endocervix consists of:

- *Epithelium.* The mucous membrane of the upper two-thirds of cervical canal is lined by *ciliated columnar epithelium,* but its lower one-third epithelium is non-ciliated columnar. Near the external os, the canal is lined by stratified squamous epithelium.
- *Stroma.* The stroma of the cervix is less cellular than that of the body of uterus.

II. Fallopian tubes

Each fallopian tube (also known as uterine tube) is approximately 10 cm in length and 8 mm in diameter. It has a medial or *uterine end* which is attached to and opens into the uterus and a *lateral end* opens into the peritoneal cavity near the ovary.

Parts. Each fallopian tube can be divided into four parts (Fig. 9.3-2):

1. Uterine or interstitial part is the most medial part which passes through the thick uterine wall.

2. Isthmus is the relatively narrow and thick-walled part which is just next to the uterine part. It is about 2.5 cm in length.

3. Ampulla is the next thin walled and dilated part of the uterine tube. It is the largest part (7 cm) of uterine tube.

4. Infundibulum refers to the funnel-shaped lateral end of the tube. It is prolonged into a number of finger-like processes known as fimbria. One fimbria is longer than rest of the fimbriae and is attached to the outer pole of ovary.

Structure. Fallopian tubes consist of same three coats as of the uterus viz. endometrium, myometrium and perimetrium.

Functions. The uterine tubes convey ova, shed by the ovaries, to the uterus. Ova enter the tube at its fimbriated end. The sperms enter the uterine tube at its medial end after traversing the vagina and uterine cavity. Secretions present in the tubes provide nutrition, oxygen and other requirements for ova and spermatozoa passing through the tube. Fertilization takes place in the ampulla and the fertilized ovum travels towards the uterus through the tube. The ciliated epithelial cells lining the tube help to move ova towards the uterus.

III. Vagina

- The vagina is a musculomembranous tube (about 8–10 cm long) located anterior to the rectum and posterior to the urethra and urinary bladder.
- Its upper end surrounds the lower part of the cervix and its lower end i.e. vaginal orifice opens into the vestibule of vagina (the cleft between the labia minora).

Structure. The wall of vagina consists of a mucous membrane, a muscle coat and an outer fibrous coat or adventitia.

1. Mucous membrane shows numerous longitudinal folds. In adult female, the vaginal mucosa lined by stratified squamous epithelium. The epithelial cells are rich in glycogen and this property is oestrogen dependent.

2. Muscle coat is made up of an outer layer of longitudinal fibres and a much thinner layer of circular fibres. Many elastic fibres are present among the muscle fibres. The lower end of vagina is surrounded by striated fibres of the bulbospongiosus muscle that form a sphincter for it.

3. Adventitial coat surrounds the muscle coat and is made up of fibrous tissue containing many elastic fibres.

Functions. The vagina serves following functions:
- It serves as the excretory duct for menstrual fluid,
- It forms the inferior part of pelvic (birth) canal,
- It receives the penis and ejaculate during sexual intercourse.

Note. No glands open into the vagina. The small amount of secretion present in the vagina is derived partly from the mucous discharge from the cervix and partly from the transudation of fluid from the vaginal epithelium, which contains glycogen. Action of bacteria on the glycogen present in the vaginal secretion produces lactic acid, which maintains the vaginal pH around 4.5. Acidic environment of vagina prevents the growth of pathogenic organisms.

FEMALE EXTERNAL GENITALIA

The external genital organs include mons pubis, labia majora, labia minora, clitoris, vestibule of vagina, bulbs of vestibule and greater vestibular glands (Fig. 9.3-1B).

Clitoris is an erectile organ located where the labia minora meet anteriorly. The clitoris is analogous to male penis. It functions solely as an organ of sexual arousal.

Vestibule is the space between the labia minora that contains the opening of urethra, vagina, and ducts of greater and lesser vestibular glands. The vaginal orifice is surrounded by a thin fold of mucous membrane called *hymen* which is usually ruptured after first intercourse or otherwise. After child birth, only a few remnants of the *hymen*—*hymenal* caruncles (tags)—are visible.

The synonymous terms vulva and pudendum include all these parts. The vulva serves:

- As sensory and erectile tissue for sexual arousal and intercourse
- To direct the flow of urine
- To prevent entry of foreign material into the urogenital tract

OVARIES

FUNCTIONAL ANATOMY

A pair of ovaries is located (one on each side) behind and below the fallopian tubes. The ovaries are ovoid glands with a combined weight of 10–20 g during reproductive years, which decreases with an increasing age. Each ovary is about

3–5 cm in length and is attached to the uterus by the broad ligament and round ligament of ovary.

Structure

Histologically, each ovary consists of following parts (Fig. 9.3-4):

1. Germinal epithelium. Germinal epithelium refers to the epithelium lining the outer surface of ovary and consists of a single layer of cuboidal cells. The term germinal epithelium is a misnomer, as it does not produce germ cells.

2. Cortex. The cortex is the outer thick main part of the substance of the ovary. It consists of following tissues:

- *Tunica albuginea* is the outer condensation of the connective tissue present immediately below the germinal epithelium.
- *Stroma* of the cortex, present deep to the tunica albuginea, is made up of reticular fibres and numerous fusiform cells that resemble mesenchymal cells.
- *Ovarian follicles* at various stages of development are scattered in the stroma. Each follicle contains a developing ovum.

3. Medulla. The medulla is the inner small part of the substance of ovary. It consists of connective tissue in which numerous blood vessels (mostly veins), smooth muscles and elastic fibres are present.

4. Hilum. The hilum refers to the area where ovary attaches to mesentery. It is the site for entry of blood vessels and lymphatics.

FUNCTIONS OF OVARIES

The two principal functions of ovaries are:

- Gametogenic function, i.e. oogenesis
- Endocrine function, i.e. secretion of female hormones called ovarian hormones

OOGENESIS

Oogenesis refers to the process of formation of ova from the primitive germ cells.

Primitive germ cells. When the bipotential gonads differentiate into ovaries in genetic female (44 + XX) embryo by 10th week of gestation, the primitive germ cells increase in number by mitosis to form oogonia.

Oogonia are the stem cells from which ova are derived. The oogonia proliferate by mitosis to form primary oocytes.

Primordial follicles. The *diploid primary oocytes* become enveloped by a single layer of flat granulosa cells and in this form are called *primordial follicles.*

- After puberty, the oogenesis or formation of ovum occurs in a highly cyclic fashion, once every 28 days till menopause.
- Every month, in each ovary, more than one primordial follicles start undergoing maturation process but only one reaches maturity and the rest undergo atresia at different stages of development. Thus throughout the whole normal reproductive life of about 30 years (from 13–42 years) about 450 ova are expelled and the remainder degenerate.
- The different stages of maturation of primordial follicle into graafian follicle *(folliculogenesis)* are described below.

Phases of folliculogenesis

The follicles at different stages of maturation are (Fig. 9.3-5):

1. Primordial follicles are the fundamental reproductive units of ovary. At the time of puberty, both ovaries contain about 300,000 primordial follicles.

- Primordial follicles are formed in fetal life. Each primordial follicle consists of the primary oocyte in prophase of the first meiotic division surrounded by a single layer of spindle-shaped (flat) cells called the granulosa cells.

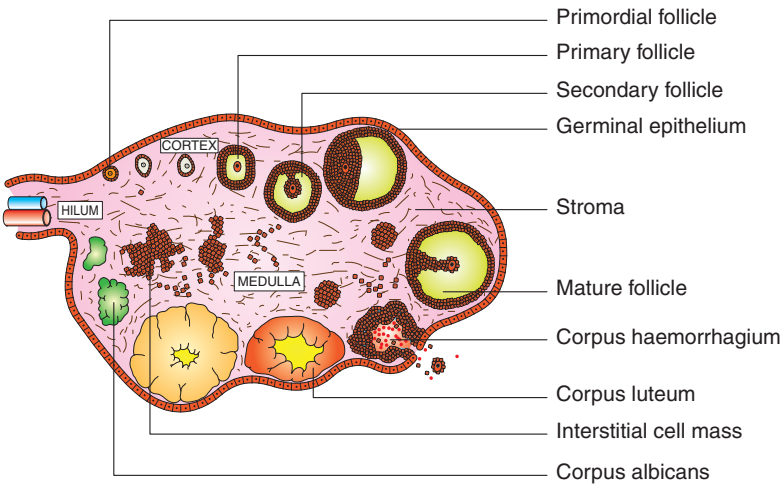


Fig. 9.3-4 Schematic diagram of the histology of ovary depicting various stages of development of follicles and corpus luteum.

- Both the granulosa cells and the primary oocyte are enveloped in a thin membrane called basal lamina (Fig. 9.3-5A).
- The granulosa cells believed to provide nutrition to the ovum and also secrete *oocyte maturation inhibiting factor,* which keeps the ovum in immature stage till puberty.

2. Primary follicle. The primary follicle is formed when the primordial follicle undergoes following developmental changes (Fig. 9.3-5B):

- *Granulosa cells,* which are flat (spindle-shaped) in primordial follicle become columnar and undergo mitotic division to form a multilayered stratum granulosum.

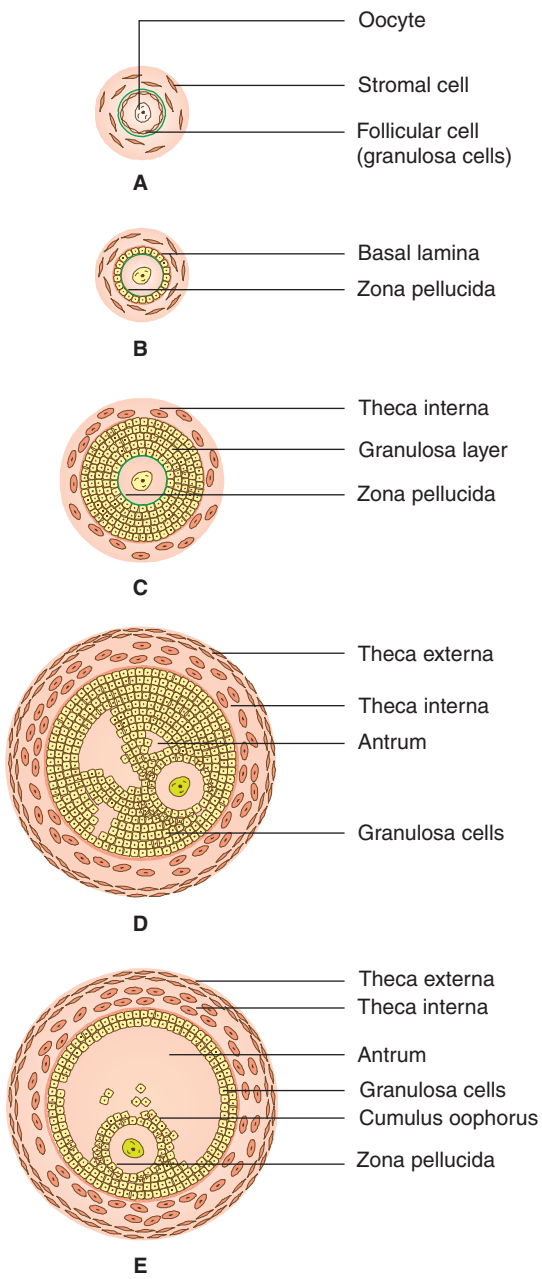


Fig. 9.3-5 Phases of folliculogenesis.

- *Oocyte enlarges* and becomes about 20 µm in size.
- *Zona pellucida,* a homogeneous membrane appears consisting of glycoprotein between the granulosa (follicular) cells and the oocyte. With the appearance of zona pellucida, the follicle is now referred to as a multilaminar primary follicle.

3. Secondary follicle is formed from the primary follicle when (Fig. 9.3-5C):

- *Granulosa cells* undergo further proliferation.
- *Oocyte* further increases in size up to 100 µm. Its nucleus becomes larger and vesicular forming germinal spots.
- *Theca folliculi* or follicular sheath is formed outside the basal lamina from the spindle-shaped cells from the stroma of cortex in ovary. The theca folliculi consist of an inner rim of secretory cells called *theca interna* and an outer rim of thickly packed fibres and spindle-shaped cells called *theca externa* (that merges with the surrounding stroma).

4. Tertiary follicle. After proliferation, the granulosa cells start secreting follicular fluid; this causes cavity to be formed in the stratum granulosum (cavitation), which is called antrum or follicular cavity. The fluid filled in the antrum is called liquor folliculi which also contains oestrogen. The granulosa cells continue to proliferate and the size of follicle is increased (Fig. 9.3-5D).

5. Graafian (antral) follicle. After about seventh day of sexual cycle, one of the tertiary follicle increases in size in response to gonadotropins [both follicle stimulating hormone (FSH) and luteinizing hormone (LH)] and forms the mature follicle called graafian or antral or vesicular follicle (Fig. 9.3-5E). A fully matured graafian follicle is characterized by following features:

- *Size* of the follicle increases markedly to about 2–5 mm. The growth of the graafian follicle is accomplished by granulosa and theca proliferation.
- *Antrum becomes larger.*
- *Theca interna becomes more prominent.* The oestrogen-secreting cells of theca interna increase.
- *Formation of secondary oocyte.* Just prior to ovulation, the primary oocyte of the fully matured graafian follicle completes the first meiotic division (which began in fetal life at about 28th week of gestation, i.e. before birth) and forms the secondary oocyte with a haploid nucleus and the first polar body.

ENDOCRINE FUNCTION OF OVARIES

The endocrine function of the ovaries is to produce female sex hormones which include:

- Oestrogens
- Progesterone

OESTROGENS

Oestrogens are (C-18) steroids. The naturally occurring oestrogens include:

- *Oestradiol.* It is the principal and physiologically most potent oestrogen. Ovarian oestradiol accounts for more than 90% of the circulating oestrogens.
- *Oestrone.* It is a weak ovarian oestrogen.
- *Oestriol.* It is the degradation product of oestradiol and oestrone. It is the weakest of all naturally occurring oestrogens.

Synthesis, plasma levels and transport of oestrogens

Sites

In the normal female, oestrogens are mainly secreted by theca interna and granulosa cells of the ovarian follicles. A small quantity is also produced by the adrenal cortex, breast, some areas of brain, placenta (during pregnancy) and by Sertoli cells (in males).

Biosynthesis

The salient points of oestrogen synthesis are:

- Oestrogens are mainly synthesized from cholesterol derived from blood and to a slight extent from acetyl *co-enzyme A.*
- During synthesis, progesterone and testosterone are synthesized first and then these are converted into oestrogens.
- The biochemical steps for synthesis of oestrogen are shown in Fig. 9.3-6. There are two pathways (Δ5 pathway and Δ4 pathway) for oestrogen synthesis. The first step, i.e. the conversion of cholesterol into pregnenolone is common in both the pathways. This reaction is stimulated by gonadotropins (FSH and LH).

Δ5 pathway. This pathway involves:

- Synthesis of 17α-hydroxypregnenolone and dihydroxyepiandrosterone. The reactions are catalyzed by enzymes 17α-hydroxylase and 17,20-hydroxylase, respectively.

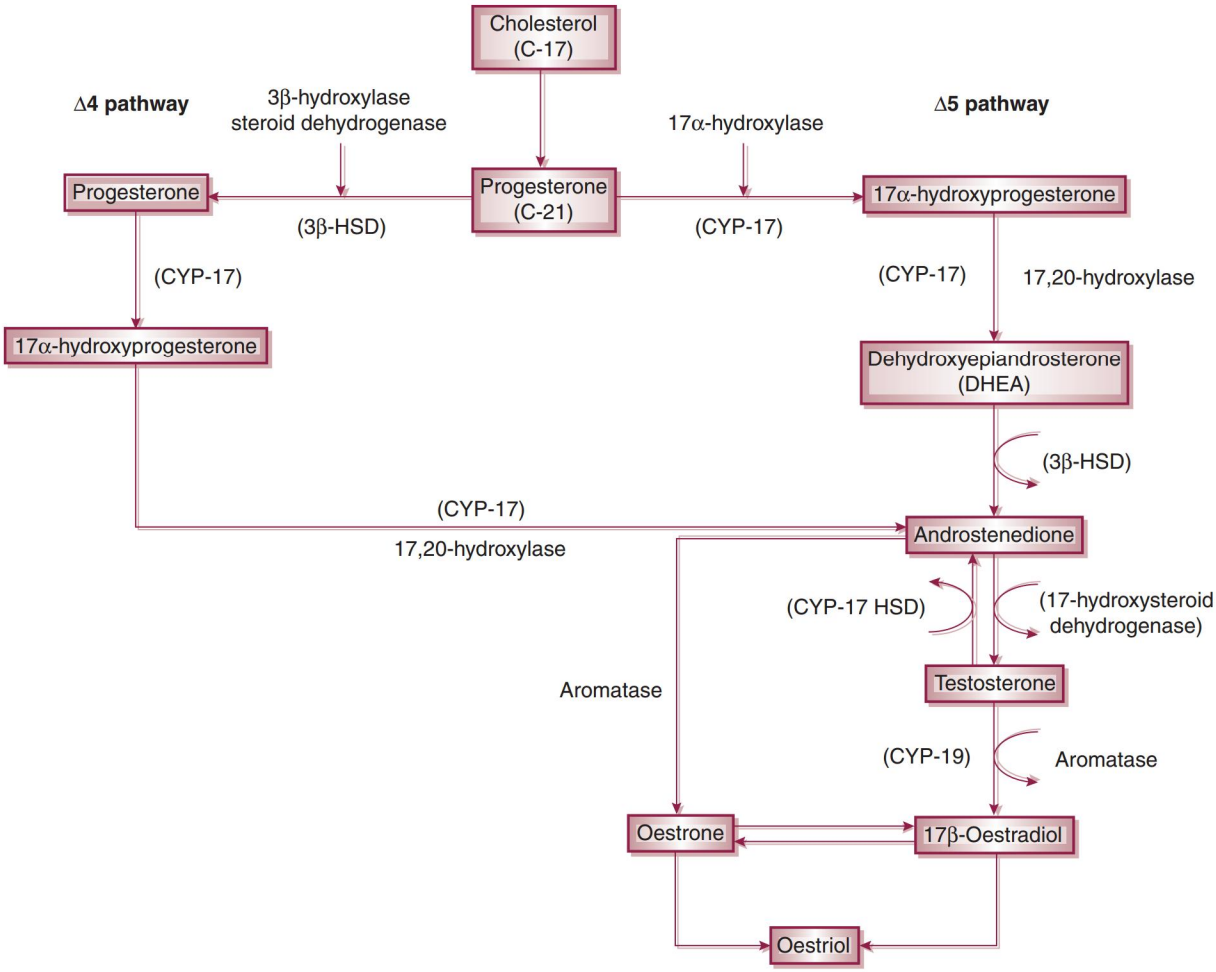


Fig. 9.3-6 Biosynthesis and metabolism of oestrogen.

Δ4 pathway. In this pathway, progesterone is the initial compound which is formed from pregnenolone by 3β-hydroxysteroid dehydrogenase and Δ5 isomerase.

- Then progesterone is converted to 17α-hydroxyprogesterone by 17α-hydroxylase (CYP-17). 17α-hydroxyprogesterone is another precursor for androstenedione via 17–20 hydroxylase.
- Then both androstenedione and testosterone are converted to oestradiol and oestrone by aromatases (CYP-19).

Mechanism of biosynthesis of oestrogen

- Gonadotropins (FSH and LH) from the anterior pituitary stimulate the synthesis of female sex hormones by acting on the receptors.
- Theca interna cells have many LH receptors. The LH therefore increases the conversion of cholesterol to androstenedione via cyclic AMP.
- The granulosa cells also possess many FSH receptors. FSH, therefore, facilitates the secretion of oestradiol by acting on these receptors through activation of cyclic AMP, which increases the activity of aromatase enzyme. The mature granulosa cells also acquire LH receptors; therefore, LH stimulates the oestradiol production from the granulosa cells also.

Plasma levels

In a normal adult woman, the plasma levels of oestrogen vary in different phases of ovarian cycle (Fig. 9.3-7). As shown in Fig. 9.3-7D, there are two peaks of oestrogen secretion. The first occurs just before the ovulation (12–13th day of sexual cycle) and is called *oestrogen surge* and the second peak occurs in the mid-luteal phase. The secretion rate of oestrogen in different phases is:

- In early follicular phase: 36 µg/day
- Just before ovulation: 380 µg/day
- During mid-luteal phase: 250 µg/day

After menopause, oestrogen level falls to minimum of 50 µg/day.

Transport

In the circulation, oestrogen is present in two forms: bound (98%) and free (2%). The oestradiol is mainly bound to plasma proteins: 60% to albumin and 38% to β-globulin. The β-globulin is also known as *gonadal steroid binding globulin protein*. It is the same protein to which testosterone binds.

Metabolism and excretion of oestrogens

The liver is the main site for metabolism of ovarian hormones. In the liver, the degradation metabolites are conjugated with glucuronic acid and sulphuric acid to form water-soluble compounds (glucuronides and sulphates of oestriol and catecholoestriol).

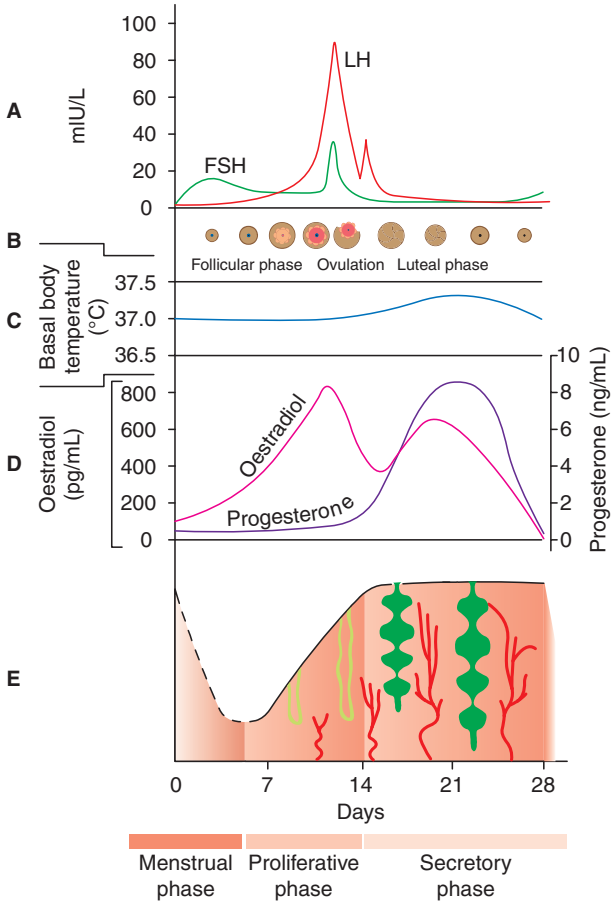


Fig. 9.3-7 Correlation of plasma concentration of gonadotropins (FSH and LH) (A); ovarian cycle changes (B); basal body temperature (C); ovarian hormones (D) and endometrial changes (E) during female sexual cycle.

Excretion. Most (4/5th) of the water-soluble compounds of oestrogens are excreted by kidney into the urine and a small amount (1/5th part) of them is secreted into the bile and gets reabsorbed into the blood by the enterohepatic circulation.

Functions of oestrogens

The functions of oestrogen for descriptive purposes can be grouped as: reproductive actions and other actions.

I. Reproductive actions

A. At puberty. At puberty, oestradiol is secreted in larger amounts which cause following changes:

1. Growth and development of genital organs

(i) Ovaries increase in size and complete ovarian cycles start, which are characterized by folliculosis, ovulation and corpus luteum formation.

(ii) Fallopian tubes become functional and show certain changes, such as epithelium becomes more ciliated, motility of fallopian tubes also increases at ovulation to transport shedded gametes.

(iii) Uterus. It enlarges in size, endometrium gets thickened due to increase in stroma and blood flow. The rhythmic cyclic changes (proliferative and secretory) occur with onset of menstrual cycle.

(iv) Cervix also enlarges and with onset of menstrual cycle, endocervix undergoes cyclic changes (see page 656).

(v) Vagina increases in size. Its epithelial lining increases in height (from 2–3 layers cuboidal epithelium to 10–12 layers cornified squamous epithelium).

(vi) External genitalia. The following changes occur in external genitalia:

- Increase in size of clitoris,
- Labia majora and labia minora increase in size and get widened.

2. Appearance of secondary sex characters. Oestrogen is responsible for the appearance of secondary sex characters (see page 631):

B. In an adult woman. Oestrogens along with progesterone regulate the ovarian cycle, menstrual cycle and cyclic changes in the cervix, vagina and fallopian tubes (see page 658) in non-pregnant state.

- It plays an important role in the maintenance of pregnancy and then during parturition (see page 667).
- It is important for breast development.

II. Other actions

The other functions of oestrogens include the effects on following:

1. Effects on bones

(i) Oestradiol accelerates the linear growth of bones at puberty by its osteoblastic activity.

(ii) Oestradiol enlarges the hip and widens the inlet of the pelvic bone to facilitate child birth.

(iii) Oestrogens maintain balance between bone formation and bone resorption by the following ways:
- It promotes bone formation by deposition of bone matrix by causing Ca^{2+} and HPO_4^{2-} retention and
- Inhibits bone resorption by inhibiting the production of osteoclasts and their activity. These effects are achieved by inhibiting the production of lymphokines, such as interleukin-I (IL-1), TNF α and granulocyte–macrophage colony stimulating factor which promote proliferation of osteoclasts.

Note. Loss of oestrogen actions after menopause shifts the bone balance towards bone resorption, thus causing osteoporosis (see page 579).

2. Effects on metabolism

(i) Protein metabolism. Oestrogens cause positive nitrogen balance due to growth promoting effect.

(ii) Fat metabolism. Oestrogens cause fat deposition in subcutaneous tissues, in the breasts and the thighs.

3. Water and electrolyte balance. Oestrogens, like other steroids in general, cause salt and water retention in the body and produce pre-menstrual tension in some women.

4. Effects on vasculature. In general, oestrogens have vasodilator and antivasoconstrictor effects.

5. Effects on CNS. Oestrogens are responsible for oestrous behaviour in animals and also increase the libido in human females. Oestrogens also act on other areas of the brain and effect the neuronal discharge and thus effect the brain functioning. It has been observed that in mice oestrogen improves the memory and learning.

6. Effects on skin. Oestrogens make the skin soft and more vascular. It makes the sebaceous glands secretions thin. Therefore, synthetic oestrogens are used as a part of treatment in acne.

Mechanism of action of oestrogens

Oestrogens act by entering into the cell and then bind with cytoplasmic receptors.

Oestrogen receptors

Oestrogen receptors are of two types (ERα and ERβ) and are coded by two different genes located on separate chromosomes.

Note. Most of the actions of oestrogens are mediated via genomic receptors (ERα and ERβ). However, some of its effects are so rapid (e.g. effect on brain neuronal discharge and feedback effect on gonadotropins release) that they might be mediated through non-genomic receptors present on the plasma membrane.

Synthetic oestrogens

Types. Various types of synthetic preparations of oestrogen are available. *Ethinyl derivatives* of oestradiol, such as *diethylstilbestrol* and *ethinyloestradiol,* are the potent oestrogens when given orally (because these are not metabolized in the liver like natural oestrogens).

Therapeutic uses. The oestrogenic preparations are used in following conditions:

- To reduce menopausal symptoms like hot flushes.
- To prevent post-menopausal osteoporosis.
- To prevent progression to atherosclerosis and incidence of heart attacks and strokes.
- As contraceptive when used along with progesterone.

PROGESTERONE

Progesterone is C-21 steroid meant for maintenance of pregnancy and biologically called prostagen or gestagen.

Synthesis, plasma levels and transport of progesterone

Sites

In a normal adult non-pregnant woman, progesterone is mainly secreted by corpus luteum and during pregnancy by the placenta. A small amount is also secreted by the adrenal cortex and by testes in case of males (Fig. 9.3-6).

Biosynthesis

Progesterone is synthesized from cholesterol. Progesterone itself is an important intermediary compound formed during biosynthesis of steroids (oestrogens and androgens).

Plasma levels

In a normal adult woman, the plasma levels of progesterone vary with different phases of sexual cycle (Fig. 9.3-7):

In early follicular phase, plasma concentration of progesterone is very low (about 0.9 ng/mL or 9 ng/dL).

In midcycle (late follicular phase), its level starts rising due to secretion from the granulosa cells and it is mainly 17α-hydroxyprogesterone and

In luteal phase, it reaches to its peak value, i.e. 18 ng/mL and

At the end of cycle, its levels fall to its minimum value.

Note. Throughout the sexual cycle the levels of progesterone are higher than the oestrogens.

During pregnancy, the levels of progesterone further rise (see page 666).

After menopause. Progesterone levels fall to its minimum (0.2 ng/mL) or even not detectable.

Transport

In the plasma, progesterone is present in two forms:

Bound form. About 98% of progesterone in the blood is present in bound form with plasma proteins, with albumin (80%) and with corticoid-binding protein also known as *transcortin* (18%).

Unbound form or free form. Only 2% of circulating progesterone is present in this form.

Metabolism and excretion of progesterone

In the liver, progesterone is metabolized to form pregnanediol and 17α-hydroxyprogesterone to pregnanetriol. The metabolites then conjugated with glucuronic acid and sulphuric acid to form water-soluble substances, which are excreted by the kidney into the urine and small amount into the bile.

Functions of progesterone

The physiological actions of progesterone can be grouped as reproductive actions and other actions.

I. Reproductive actions

Reproductive actions are mainly on the reproductive organs primed by oestrogens and these include:

1. Action on uterus. The progesterone is responsible for the secretory phase of the endometrial cycle and prepares the endometrium to receive the zygote. It decreases the uterine motility.

Uterine motility. Progesterone decreases the uterine motility by following ways:

- It decreases the synthesis of voltage-dependent Ca^{2+} channel proteins, therefore Ca^{2+} uptake decreases.
- It decreases the number of oestrogen receptors on the myometrium.

2. Endocervix. The cervical secretions become thick and viscid, and ferning pattern disappears.

3. Vagina. Vaginal epithelium becomes thickened, cornified and infiltrated with leucocytes.

4. Fallopian tubes. Progesterone increases the epithelial cell secretions rich in nutritive materials to provide nutrition to shedded ovum, incoming sperm or to zygote if fertilization occurs.

5. Breast. Progesterone causes lobular and alveolar growth of breast.

6. During pregnancy, the main function of progesterone is to maintain the pregnancy (see page 667).

II. Other actions

The other systemic effects of progesterone are:

1. Thermogenic effect. Progesterone, known as a thermogenic steroid, increases the basal body temperature by 0.5°C in post-ovulatory phase.

2. Effect on CNS. Progesterone alters the secretion and release of various neurotransmitters in the hypothalamus and other areas of the brain and thereby decreases the appetite and produces *somnolence.*

3. Effect on respiration. Progesterone increases the sensitivity of the respiratory centre to carbon dioxide stimulation. Due to this fact, the $pACO_2$ is slightly less in woman during the luteal phase of sexual cycle.

4. Effect on fat metabolism. Progesterone (particularly C-19 progesterone) decreases the serum high-density lipoprotein (HDL). Thus it acts as a proatherogenic agent.

OTHER OVARIAN HORMONES

Besides female sex steroids (oestrogen and progesterone), ovaries also secrete peptide hormones as:

1. Inhibin. Structurally it is polypeptide, it inhibits the FSH release.

2. Activin. Structurally, it is also a polypeptide, its action is to activate FSH secretion from the anterior pituitary.

3. Relaxin is a polypeptide hormone produced by corpus luteum and other sites include: uterus, placenta and mammary glands and in males from the prostate gland. Its main role is during pregnancy as it relaxes pubic symphysis and pelvic joints, softens and dilates the uterine cervix and facilitates delivery.

In non-pregnant state, it releases from the corpus luteum and endometrium during secretory phase and its function is not known.

In males, relaxin is present in the semen and helps in sperm motility.

4. Ovarian androgens. A small amount of testosterone is also secreted by the ovaries during biosynthesis of oestrogen and progesterone, but the main source of androgens in female is adrenal cortex.

These androgens are responsible for acne vulgaris, libido and pubic hair.

FEMALE SEXUAL CYCLE

The sexual life span of a female can be divided into three periods:

1. Birth to puberty. During this period, primary and accessory female sex organs remain quiescent.

2. Puberty to menopause. With the onset of puberty the female sexual cycle starts, which repeats every 28 days. The occurrence of first menstrual cycle is called *menarche*. The permanent stoppage of menstrual cycle is called menopause, which occurs at the age of about 45–50 years. The period between menarche and menopause is called *reproductive period.* During this period, females have rhythmical sexual cycles.

3. Post-menopausal period extends after menopause (45–50 years) to rest of the life. During this period, the female sexual cycle ceases.

Female sexual cycle refers to the monthly rhythmic sexual cycle occurring in females during the normal reproductive period.

Components of human female sexual cycle. During each female sexual cycle, rhythmical changes occur in ovaries and accessory sex organs—uterus, cervix and vagina.

Duration of female sexual cycle is usually 28 days. But under physiological conditions, it may vary between 20 and 40 days. Traditionally, first day of the menstrual bleeding is taken as the 1st day of female sexual cycle.

OVARIAN CYCLE

Ovarian cycle refers to the rhythmic changes occurring in ovaries during each female sexual cycle of about 28 days. During each cycle, a single mature ovum is released from the ovary. Ovarian changes occurring during the female sexual life completely depend on the gonadotropic hormones (FSH and LH), which are secreted by the anterior pituitary. The ovarian cycle can be divided into three phases:

- Pre-ovulatory phase or follicular phase
- Ovulation
- Post-ovulatory phase or luteal phase

PRE-OVULATORY PHASE

Pre-ovulatory or follicular phase of the ovarian cycle extends from the fifth day of the cycle till the time of ovulation (which takes place at about 14th day of the cycle). Thus, this phase generally lasts for 8–9 days (but may vary from 10 to 25 days).

- Changes in the ovary during this phase are mostly under the influence of FSH from the anterior pituitary. Luteinizing hormone also helps in the maturation of the follicle in the latter part of follicular phase (for details see hormonal control of female sexual cycle; see page 658).
- During this phase of each cycle, some 10–15 primordial follicles start maturing, but only one follicle matures fully and the rest undergo atresia (atrophy) at different stages of development. A fully matured graafian follicle is characterized (Fig. 9.3-5E and see page 649).

OVULATION

Ovulation refers to the release of secondary oocyte from the ovary (following rupture of graafian follicle) into the peritoneal cavity. It usually occurs *14 days after the onset* of menstruation (Fig. 9.3-7B).

Process of ovulation involves following sequence of events:

- *LH surge and ovulatory peak of FSH.* The ovulation is caused by a LH surge at mid cycle in response to an elevation in plasma oestradiol concentration (150 pg/mL).
- *Ovulatory peak of FSH* (2–3 fold increase in secretion) occurring 2 days prior to ovulation is thought to be stimulated by progesterone. FSH increases the granulosa cell LH receptors.

Changes in graafian follicle. The LH and FSH produce following changes in the graafian follicle before ovulation:

- *Rapid swelling of the follicle.* There occurs a rapid growth of new blood vessels into the follicle wall and

prostaglandins are secreted into the follicular tissue. Both these cause diffusion of plasma into the follicular fluid and further swelling of the follicle.

- *Formation of stigma.* Due to rapid swelling of the follicle, its outer wall is stretched forming a very thin avascular area (stigma) over the most convex point of the follicle, which protrudes like a nipple in the peritoneal cavity.
- *Release of proteolytic* enzymes from the lysosomes in the theca externa cells is activated by the progesterone.
- *Dissolution of capsular* wall and its further weakening is caused by the proteolytic enzymes.
- *Rupture of graafian follicle.* The simultaneous stretching and enzymatic dissolution of the follicular wall leads to degeneration of *the stigma*. Within 30 min of protrusion, fluid begins to ooze from the stigma followed soon by rupture of follicle with release of ovum (secondary oocyte) surrounded by corona radiata into the peritoneal cavity near the fimbriated end of the fallopian tube. Thus, usually only one ovum is released from any one of two ovaries during each sexual cycle. The released ovum enters the fallopian tube through its fimbriated end.

Determination of ovulation time

The ovulation time can be determined by following indirect methods:

1. From basal body temperature. The basal body temperature falls slightly (0.3–0.5°C) just prior to ovulation and increases slightly after ovulation. Therefore, the time of ovulation can be determined by measuring the morning temperature from rectum or vagina (Fig. 9.3-7C).

2. From hormonal excretion in urine. The urinary excretion of end products of oestrogen like oestrone, oestradiol and 17β-oestradiol increases to the peak at the time of ovulation and that of end products of progesterone-like pregnanediol increases after ovulation. Therefore, time of ovulation can be determined by estimating their urinary levels for few days during mid period of menstrual cycle.

3. From hormonal levels in plasma. The plasma content of FSH, LH, oestrogen and progesterone is measured during mid period of menstrual cycle and time of ovulation is determined from following observations:

- LH and oestrogen levels are increased and FSH level is decreased at the time of ovulation.
- Progesterone level is increased after ovulation.

4. By ultrasound scanning, the process of ovulation can be recorded.

Note. The ovarian cycles during which ovulation does not occur are called anovulatory cycles. If LH surge occurring prior to ovulation is not of sufficient magnitude, ovulation does not occur. First few cycles after puberty may be anovulatory.

POST-OVULATORY PHASE

Post-ovulatory phase, also called luteal phase of ovarian cycle, is of remarkably constant period of about 14 days. This phase is characterized by following events (Fig. 9.3-4):

Formation of corpus haemorrhagicum. Following ovulation, the outer wall of the graafian follicle collapses and promptly fills with blood forming the so-called corpus haemorrhagicum. Minor bleeding from the follicle into the abdominal cavity may cause peritoneal irritation and fleeting lower abdominal pain (mittelschmerz).

Formation of corpus luteum. Soon, the granulosa cells and theca cells of the follicle lining begin to proliferate, and the clotted blood is rapidly replaced with yellowish lipid-rich *luteal cells*. This process is called luteinization and the total mass of the cells is now called *corpus luteum.* Luteinizing hormone is responsible for luteinization.

Formation of corpus albicans. If there is no fertilization and pregnancy does not occur, the corpus luteum begins to involute (regress) after 24th day of the sexual cycle and is eventually replaced by a whitish scar tissue called the corpus albicans. This involution occurs due to falling levels of FSH and LH and also the hormone *inhibin* secreted by the lutein cells. With the involution of corpus luteum, on 26th day of the normal female sexual cycle, levels of oestrogen, progesterone and inhibin fall. This removes feedback inhibition of the anterior pituitary; consequently, the FSH and within a few days LH secretion begins and the next ovarian cycle is initiated.

Corpus luteum of pregnancy. However, if the ovum released is fertilized and pregnancy occurs, then the corpus luteum formed during post-ovulatory phase persists and serves as the major source of oestrogen and progesterone till the third month of pregnancy when the placenta takes over its endocrine function.

ENDOMETRIAL CYCLE

Endometrial cycle refers to the cyclic changes occurring in the endometrium during active reproductive period (menarche to menopause) in females leading to recurrent monthly bleeding per vaginum (menstruation). These cyclic changes in the endometrium are brought about by the cyclic production of oestrogens and progesterone by the ovaries. Menstrual is a Latin word meaning mensis, i.e. lunar month of 28 days. Though the menstrual cycle for description purposes is considered to be of 28 days, but the cycle is by no means as regular as the name suggests. The menstrual cycles of 25 to 35 days are also regarded as normal cycles.

PHASES OF ENDOMETRIAL CYCLE

The endometrial cycle of 28 days can be divided into three phases (Fig. 9.3-7E):

- Menstrual phase (1st–5th day)
- Proliferative phase (6th–14th day)
- Secretory phase (15th–28th day)

For the purpose of better understanding, the menstrual phase is described last of all.

Proliferative phase

Extent of proliferative phase of endometrial cycle is from day 6th to 14th day. It follows the phase of menstruation, after which only a thin basal layer of original endometrium is left.

Hormone responsible for the changes in the endometrium during this phase is oestrogen secreted by the developing graafian follicle in the ovary. Thus, proliferative phase of the endometrial cycle coincides with the follicular phase of ovarian cycle.

Changes in endometrium, which occur during proliferative phase, are:

1. Thickness of endometrium, which is less than 1 mm at the end of menstrual phase, increases to 3–4 mm at the end of the proliferative phase.
2. Angiogenesis in the stratum functionale leads to proliferation of blood vessels, which become the spiral arterioles that profuse the stratum functionale.
3. Endometrial glands are stimulated to grow. The glands contain glycogen but they are non-secretory.

Secretory phase

Extent of secretory phase (also known as post-ovulatory phase of endometrial cycle) is from day 15th to 28th day.

Hormones responsible for changes in the endometrium during this phase are both oestrogen and progesterone secreted by the *corpus luteum* formed after ovulation. Thus, the secretory phase of endometrial cycle coincides with the luteal phase of ovarian cycle.

Changes in the endometrium, which occur during this phase, are:

- There is elongation and coiling of the endometrial mucous glands. These glands become secretory and secrete thick viscous fluid containing glycogen.
- *Blood supply* of endometrium further increases as progesterone promotes spiraling of blood vessels.
- *Two characteristic features of endometrium* in secretory phase thus are prominent corkscrew-shaped glands and increased vascularity.
- *Thickness* of endometrium increases to 5–6 mm at the end of secretory phase. Thus the thickened endometrium with large amounts of nutrients is ready to provide appropriate conditions for implantation of ovum during this phase.
- *If fertilization does not occur* and there is no pregnancy, the corpus luteum in the ovary involutes to form corpus albicans and on day 26th of the menstrual cycle the levels of oestrogen and progesterone fall suddenly and mark the end of secretory phase of endometrial cycle.

Menstrual phase

The menstrual phase of endometrial cycle is also called bleeding phase. The average duration of this phase is 3–5 days. About 24 h before the end of menstrual cycle, there is sharp decline in the plasma levels of oestrogen and progesterone, which is responsible for menstrual bleeding. The sequence of events is:

- *Intense spasm of spiral arteries* occurs leading to hypoxia and ischaemia. This effect is mediated via local production of *leukotrienes* and *prostaglandins*.
- *Necrosis of stratum functionale of the endometrium* and of the walls of the spiral arteries occurs as a result of ischaemia.
- *Blood vessels get open up* due to necrosis of their wall resulting in seepage of blood into the surrounding endometrial necrotic tissue.
- *Separation of necrotic tissue* starts gradually from the underlying basal viable tissue and ultimately it is sloughed off. The necrosis and sloughing does not occur simultaneously in whole of the uterus rather it occurs in patches and is completed in 3–5 days.
- *Endometrial debris* contains necrosed sloughed off tissue, blood, serous fluid and a large amount of prostaglandins and fibrolysins.
- *Average amount of blood loss during each* menstrual cycle is 30 mL.
- *Menstrual blood immediately gets clotted inside the uterine cavity but soon gets liquefied* by fibrolysins present in the endometrial debris.
- During menstrual phase, about two-thirds of the superficial endometrium is sloughed off and only a thin basal layer (2 mm thick) is left behind.

CYCLIC CHANGES IN CERVIX

The mucosal lining of cervix (endocervix) also shows certain cyclic changes during sexual cycle. These are:

During menstruation phase, the mucosa of cervix does not undergo desquamation (shedding off) like that of endometrium.

During proliferative phase (oestrogen phase), the secretions of the mucosal cells of endocervix become thin watery and alkaline. At the time of ovulation, the cervical mucus is thinnest and its elasticity is maximum. It can be stretched like a long, thin elastic thread up to 8–12 cm (spinnbarkeit effect). The mucus also produces a characteristic fern-like pattern when a drop of mucus is spread on the glass slide and allowed to dry *(Fern test)* (Fig. 9.3-8B).

This characteristic nature of cervical mucus favours the transport of sperms in the female genital tract and makes the conditions favourable for fertilization.

During secretory phase under the influence of progesterone, cervical secretions decrease in quantity and become thick, tenacious and cellular, and fern pattern is not seen (Fig. 9.3-8C). These changes make a plug and prevent the entry of sperm through cervical canal.

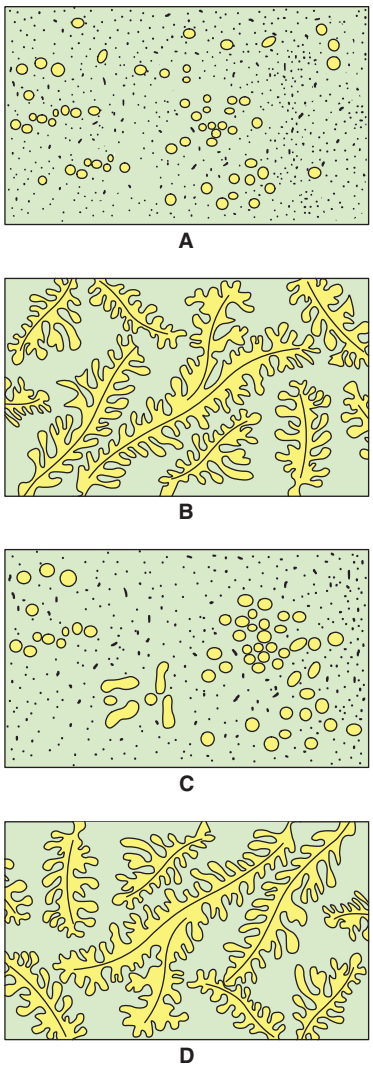


Fig. 9.3-8 Characteristics of cervical mucus as seen on smear examination during various phases of normal menstrual cycle: A, on 7th day (no fern pattern); B, on 14th day (typical fern pattern); C, on 21st day (fern pattern disappears) and D, 21st day of an anovulatory cycle (fern pattern persists).

Note. *Fern test.* The fern pattern of cervical mucus in the proliferative phase and its disappearance in the secretory phase is indicative of ovulatory cycle, whereas persistence of fern pattern (Fig. 9.3-8D) throughout the cycle indicates anovulatory cycle.

CYCLIC CHANGES IN VAGINA

Vaginal canal is lined by stratified squamous epithelium, which is highly sensitive to oestrogens (oestradiol). Vaginal epithelium undergoes following cyclic changes in the endometrial cycle:

In proliferative phase, vaginal epithelium becomes thickened (by adding up more and more layers of epithelium) and cornified.

In secretory phase under the influence of progesterone, vaginal epithelium proliferates and gets infiltrated with leucocytes and the vaginal secretions become thick and viscid.

OTHER CHANGES DURING SEXUAL CYCLE

Hormonal oscillations during sexual cycle though mainly effect ovaries, uterus, cervix and vagina but some changes have also been observed in the fallopian tubes, breast and in the body weight.

1. Changes in fallopian tubes are as follows:
(i) During follicular phase, there occurs an increase in the number of cilia of epithelial cells and their rate of beating.

(ii) At the time of ovulation, the motility of fallopian tubes increases.

(iii) During luteal phase, under the influence of progesterone, there occurs an increase in the secretion of epithelial cells. This provides nutrition to the ovum, incoming sperm and the zygote if fertilization occurs.

2. Changes in breast. Some women complain of feeling of fullness and tenderness in the breasts. These symptoms have been related to the proliferation of lobules and duct system under the influence of oestrogen and progesterone. All these symptoms regress during menstrual phase.

3. Pre-menstrual weight gain. Many women experience feeling of heaviness (pre-menstrual tension) near the end of cycle. This effect is due to salt and water retention caused by the oestrogen. The feeling of heaviness disappears during menstruation phase.

HORMONAL CONTROL OF FEMALE SEXUAL CYCLE

The hypothalamo-hypophyseal-gonadal axis regulates the cyclic changes occurring during the female sexual cycle. The role of each component of the axis is (Fig. 9.3-9):

ROLE OF HYPOTHALAMUS

Hypothalamus regulates the secretions of gonadotropins (both FSH and LH) through the gonadotropin-releasing hormone (GnRH). The GnRH is also known as luteinizing hormone releasing hormone (LHRH). It reaches the anterior pituitary through hypothalamo-hypophyseal portal system where it is stored as small granules. It stimulates the anterior pituitary cells to release gonadotropins. The release of GnRH is influenced by various factors, such as; dopamine, endorphins, ratio of FSH and LH, gonadal hormones (oestrogen and progesterone), dark and light cycles operating through melatonin (released from pineal gland).

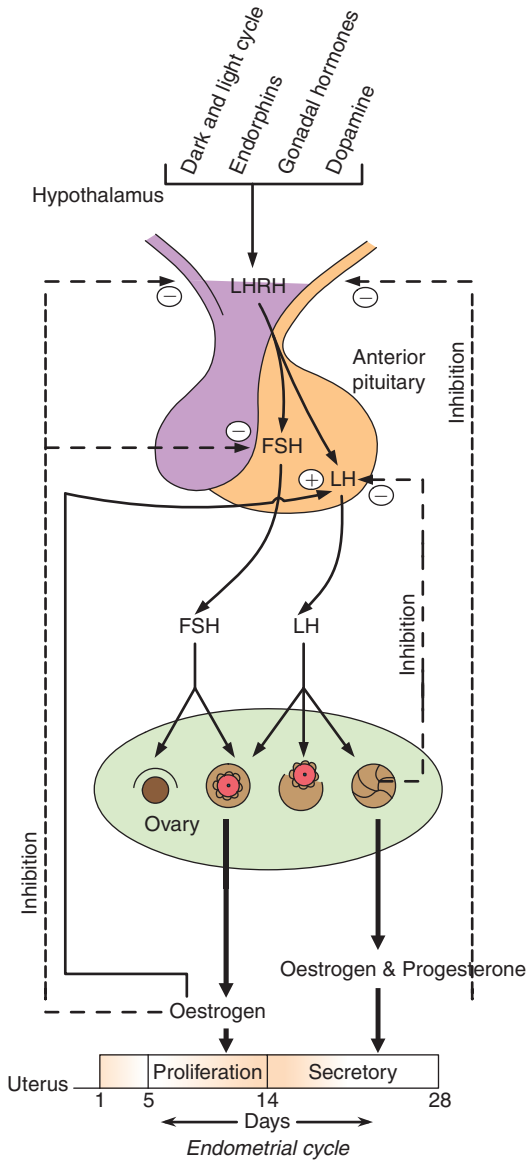


Fig. 9.3-9 Hypothalamo-hypophyseal-ovarian axis regulating the female sexual cycle through positive and negative feedback mechanisms.

Note. Some women are infertile due to hypothalamic disorders. In such cases normal ovulation and menstrual cycles can be restored by exogenous administration of GnRH.

ROLE OF ANTERIOR PITUITARY

The anterior pituitary plays its role in the female sexual cycle regulation by releasing gonadotropins (FSH and LH). Gonadotropins in turn regulate the ovarian cycle, i.e. formation of graafian follicles (folliculosis), ovulation and formation of corpus luteum.

Regulation of gonadotropins

The secretion of both FSH and LH is regulated by:

1. Gonadal hormones. The gonadal hormones (oestrogen and progesterone) regulate gonadotropin secretion by their feedback effect (Fig. 9.3-9). Depending on relative plasma level of these hormones, the effect may be positive or negative, or both positive and negative.

- *Oestrogen, in moderately high plasma concentration* (just before ovulation) inhibits the release of FSH by negative feedback effect) and promotes LH secretion (by positive feedback effect).
- *High levels of oestrogen and progesterone* in mid-luteal phase inhibit the secretion of FSH and LH (by negative feedback effect).
- *Low levels of gonadal hormones* (during menstruation phase) increase the secretion of both FSH and LH (by positive feedback effect).

The feedback effect (positive or negative) of ovarian hormones is brought about by its action either directly on anterior pituitary or through the hypothalamus (Fig. 9.3-9).

Note. Clomiphene (a synthetic drug) induces ovulation by acting on hypothalamus and thereby promotes LH release from anterior pituitary.

Oral contraceptives are the preparations containing high concentration of oestrogen and progesterone. These drugs inhibit gonadotropin release by negative feedback effect and prevent ovulation.

2. Human chorionic gonadotropin (HCG). It is a glycoprotein secreted by syncytiotrophoblasts during early pregnancy (12–16 weeks of gestation). Like LH, HCG also maintains the functional state of corpus luteum and thus elevates gonadal hormones resulting in inhibition of gonadotropin release.

3. Prolactin. It is a mammotropic hormone secreted from anterior pituitary during lactation. It inhibits GnRH release and thus lowers the basal secretion of FSH and LH (cause for lactation amenorrhoea).

4. Activin. It is structurally quite similar to inhibin (secreted from ovary). It is synthesized in the cells of the anterior pituitary. It stimulates the synthesis and release of FSH by autocrine and paracrine actions.

ROLE OF OVARIES

Ovaries play an important role in the regulation of ovarian cycle and endometrial cycle by secreting gonadal hormones (oestrogen and progesterone).

Oestrogen

In each sexual cycle, the plasma concentration of oestrogen starts rising from first day of the cycle and reaches to its peak just before ovulation (at 12–13th day) called *oestrogen surge.*

- Oestrogen through its positive feedback effect is responsible for *ovulation* due to LH surge.
- Oestrogen is responsible for the proliferative phase of the endometrial cycle.

Progesterone

- After ovulation, there occurs formation of corpus luteum and the progesterone concentration starts rising. Therefore, in the luteal phase of ovarian cycle level of both oestrogen and progesterone are high, but progesterone rises markedly.
- Progesterone prepares the oestrogen primed endometrium for implantation. Thus, it is responsible for the secretory phase of endometrial cycle.

ABNORMALITIES OF FEMALE SEXUAL CYCLE

The abnormalities of female sexual cycle are grouped as:

- Abnormalities of ovarian functions
- Abnormalities of menstruation

A. Abnormalities of ovarian functions

1. Hypogonadism (hyposecretion of ovarian hormones) means less than normal secretions by the ovaries. It occurs when the ovaries are poorly developed or absent since birth or genetically become abnormal and non-functional. Hypogonadism results in female eunuchoidism.

2. Ovariectomy. When ovaries are removed surgically in a sexually mature female, it leads to following effects:

- Atrophy of genital apparatus (i.e. uterus, vagina and external genitalia),
- Stoppage of menstruation.
- Vasomotor changes, such as flushing of skin of face, neck and chest (hot flushes)
- Emotional disturbances, such as irritability and depression.

3. Hypersecretion by the ovaries. The ovarian secretions are well regulated by hypothalamo-hypophyseal ovarian axis. Hence extreme hypersecretion by the ovaries is a rare condition. If it occurs in:

- *Pre-pubertal stage* then precocious puberty results (see page 632).
- *Post-menopausal stage,* then hypertrophy of the endometrium and irregular bleeding are the common features.

B. Abnormalities of menstruation

1. Anovulatory cycles means menstrual cycles occur at normal intervals, but ovulation does not occur. Anovulatory cycles are the normal entity up to 1–2 years after the menarche and few years before menopause. Anovulatory cycles in the fertile period of womanhood are the main cause of female infertility.

2. Amenorrhoea. The term amenorrhoea refers to the absence of menstrual bleeding or periods. It is of two types:

- *Primary amenorrhoea* means menstrual bleeding has never occurred and this condition is because of failure of sexual maturation.
- *Secondary amenorrhoea* means cessation of menstrual cycles in a woman who previously has normal and regular cycles. Pregnancy is the most common cause of secondary amenorrhoea. Other conditions which result in secondary amenorrhoea are: emotional disturbances, environmental changes, hypothalamic and pituitary disorders and certain systemic diseases.

3. Hypomenorrhoea. The term refers to scanty menstruation.

4. Menorrhoea. It refers to an abnormally profuse bleeding during normal regular cycles.

5. Metrorrhagia. This condition refers to the occurrence of uterine bleeding in between the periods.

6. Oligomenorrhoea means infrequent and reduced frequency of menstruation.

7. Dysmenorrhoea is the term related to discomfort menstruation (or painful menstruation).

8. Pre-menstrual syndrome. About 7–10 days before the end of cycle some women experience symptoms like irritability, lack of concentration, feeling of depression, heaviness, headache and constipation which is called pre-menstrual syndrome.

Chapter 9.4

Physiology of Coitus, Pregnancy and Parturition

PHYSIOLOGY OF COITUS
- **Male sexual act**
  - Erection of penis
  - Emission of seminal fluid
  - Ejaculation
- **Female sexual act**
  - Sexual excitement
  - Orgasm
  - Resolution phase

PHYSIOLOGY OF PREGNANCY
- **Fertilization and implantation**
  - Fertilization
    - Transport of gametes
    - Sperm capacitation
    - Fusion of gametes
  - Implantation
- **Placenta and pregnancy tests**
  - Placenta
    - Placental membrane
    - Functions
  - Pregnancy tests
- **Physiological changes in mother during pregnancy**
  - Changes in genital organs
  - Weight gain
  - Haematological changes
  - Cardiovascular system changes
  - Respiratory system changes
  - Urinary system changes
  - Gastrointestinal system changes
  - Metabolic changes
  - Endocrine system changes
  - Changes in the skin
  - Psychological changes

PHYSIOLOGY OF PARTURITION
- **Mechanics of parturition**
  - Uterine contractions
  - Cervical dilatation
- **Control of parturition**
  - Hormonal factors
  - Mechanical factors

PHYSIOLOGY OF COITUS

Coitus refers to the process of sexual intercourse by which sperms are deposited into the vagina. Physiologically, coitus involves both male and female sexual act or sexual arousal.

MALE SEXUAL ACT OR MALE SEXUAL AROUSAL

The male sexual act consists of following three sequence of events:

- Erection of penis,
- Emission of seminal fluid and
- Ejaculation.

I. Erection of penis or stage of excitement

In this stage, penis becomes hard, stiff and an elongated structure. Erection of penis is brought about by integrity of the reflex arc, which comprises:

- Sexual stimulation,
- Afferents to the integrating centres,
- Integrating centres,
- Efferents and
- Response.

1. Sexual stimulation. Sexual stimulation has two components:

Psychological component. It is in the form of thought, feeling, watching movie or book picture, etc. The psychological

component arises in the cerebral cortex or from limbic system. The impulses from psychic stimulations are carried or descend to the integrating centres located in the spinal cord. The psychic component can *reinforce* or inhibit the integrating centres.

Physical components of sexual sensations involve:

Sensations from genitalia (mainly from glans penis), *adjacent areas* (like scrotum, epithelium of anus and perineal structures) and *structures* like urethra, bladder, prostate, seminal vesicles, testes, vas deferens particularly when these structures are filled with the secretions.

2. Afferents to integrating centres. From genitalia and other structures, the afferent impulses are carried by *pudendal nerve* and *sacral plexus* to the sacral part of spinal cord.

3. Integrating centres. Integrating centres for reflexogenic erection are located in lumbar segments (L_2–L_3) and sacral segments (S_2–S_4) of the spinal cord. The lumbar centres are in turn connected to the sacral centres.

4. Efferent pathway for erection is carried through sacral parasympathetic fibres via nervi erigentes to the:

- Smooth muscle fibres of the penile arterioles.
- Erectile tissue (corpora cavernosa and corpora spongiosum of the penis).
- The bulbourethral glands.
- Some of the fibres end pre-synaptically on to noradrenergic neurons, where acetylcholine acts on muscarinic receptors to decrease the release of norepinephrine.
- Nervi erigentes also carry non-cholinergic and non-adrenergic fibres, which contain a large amount of enzyme *nitric oxide synthase* which causes formation of nitric oxide. Nitric oxide is another potent vasodilator.
- Therefore, the substances (neurotransmitters) released are ACh, VIP (vasoactive intestinal peptide) and nitric oxide.

5. Response. By acting on the effector structures, there occurs:

- Vasodilation of the penile arterioles leading to an increased blood flow under pressure resulting in filling of the erectile tissue with blood.
- The blood-filled erectile tissue compresses the central vein of the penis (blocking the venous flow from penis), which further increases the pressure within the sinusoids of erectile tissue, and thus the penis becomes hard rigid structure.
- *Lubrication:* Parasympathetic activity during sexual stimulation causes secretion of mucus from the urethral and bulbourethral glands.

II. Emission of seminal fluid

In this stage, semen moves into the urethra. Emission is a sympathetic response integrated at upper lumbar spinal segment centres (L_1 and L_2).

- When sexual stimulus becomes very strong then reflexly emission and ejaculation occur. The afferent fibres are carried by internal pudendal nerve to the spinal cord (lumbar segment L_1–L_2). Efferents are then carried along sympathetic fibres through hypogastric and pelvic sympathetic plexus to initiate emission.
- Emission is carried out by contraction of vas deferens.

III. Ejaculation

Ejaculation means deposition of seminal fluid into the vagina of female.

APPLIED ASPECTS

Impotence. Impotence refers to lack of power in male to copulate. It may be:

- *Primary impotence,* which is of rare occurrence.
- *Secondary impotence* may have organic causes or may be followed by drugs or alcohol. Stress and depression are also possible causes.

FEMALE SEXUAL ACT

Female sexual act, similar to male sexual act is reflexogenic and involves psychological and physical components. The three phases of female sexual act are:

- Sexual excitement,
- Orgasm and
- Resolution.

I. Phase of sexual excitement

The phase of sexual excitement is also called phase of female erection and lubrication. It corresponds to erection of penis in males. This phase is brought about by integrity of the reflex arc, which comprises following components:

1. Sexual stimulation. Sexual stimulation in females like that of males has two components, psychological and physical.

Psychological stimulation. The sexual desire in females is aroused by erotic thoughts, which originate from the cerebral cortex or limbic system (amygdala). The sexual desire is believed to be increased at the time of ovulation (may be because of high levels of oestrogen). It is also believed that sexual desire in female is produced partly by oestrogen.

Physical component of the sexual stimulation consists of sexual sensations aroused from massaging/irritation of external genitalia (vulva, clitoris, labia minora and labia majora) and perineal region. Clitoris is highly sensitive and is responsible for initiation of sexual sensations. Massaging of the breasts and even kissing enforce the sexual sensations.

2. Afferents to the integrating centres. The sensory signals are transmitted via pudendal nerve to spinal cord. The impulses are then transmitted to the cerebral cortex and also to the integrating centres for the local reflex responses.

3. Integrating centres. The local reflexes are integrated in sacral segments (S_2, S_3 and S_4) and lumbar segments (L_1 and L_2) of the spinal cord. These integrating centres are also influenced by the psychological components.

4. Efferent pathway. The parasympathetic signals for female erection and lubrication travel by nervi erigentes from sacral plexus to the arteries of external genitalia, and Bartholin glands and mucosal epithelial cells of vagina.

5. Response. During sexual intercourse, the erectile tissue (located around introitus and clitoris) is activated by parasympathetic impulses producing:

- Increase in blood flow and accumulation of blood in erectile tissue resulting in an increase in the size of external genitalia, and vaginal congestion. Congestion of vagina occurs due to transudation of fluid from the vaginal epithelium,
- Vaginal lubrication facilitates the penile insertion. Further vaginal congestion leads to tightening of the vaginal opening around the penis.
- Stimulation also results in copious secretion from the Bartholin glands (situated beneath labia minora) and mucous cells in vagina. These secretions further lubricate vagina and help in producing massaging effect on penis.
- With increasing excitement, blood flow further increases resulting in deepening of colour of labia majora. Along with local response, systemic effects (as increase in heart rate, respiratory rate and blood pressure) and in general increase in the muscle tone also occur.

II. Orgasm

The orgasm results when intensity of sexual stimulation reaches its peak. It is analogous to emission and ejaculation in males.

During orgasm, there occurs rhythmic contractions of peroneal muscles, uterus and vagina, and dilatation of the cervical canal. The intense sexual sensation perceived during orgasm is called climax. This stage lasts for about 15–30 s.

It has been observed that in lower animals oxytocin released from the posterior pituitary via amygdala (limbic system–hypothalamus–posterior pituitary stimulation) is responsible for the uterine contractions.

III. Resolution phase

Orgasm is immediately followed by the resolution phase. This phase is characterized by a sense of satisfaction followed by relaxed state of mental peacefulness called resolution. The heart rate, blood pressure, respiration and all other parameters come to their normal level and there occurs relaxation of the muscles.

PHYSIOLOGY OF PREGNANCY

Physiology of pregnancy is mainly concerned with maternal adaptations to provide ideal atmosphere for fertilization, nutrition to the growing fetus and safe child birth. The physiology of pregnancy can be discussed under following headings:

- Fertilization and implantation,
- Formation of placenta and its functions,
- Physiological changes during pregnancy and
- Applied aspects.

FERTILIZATION AND IMPLANTATION

FERTILIZATION

Fertilization refers to the fusion of male and female gametes (i.e. spermatozoon and ovum). It takes place in the middle segment (ampulla) of the fallopian tube. It involves following events:

1. Transport of gametes

Before fertilization, the ovum and sperms reach the ampulla for fertilization.

Transport of ovum. At the time of ovulation, the ovum is directly expelled into the peritoneal cavity and then enters into the fallopian tube. When ovulation occurs, the fimbriae of infundibulum encircle the surface of ovary, rub it and pick up the ovum and then direct it towards the ostium by continuous beating of cilia. The contractions of smooth muscle fibres present in the wall of fallopian tube also help in transport of the ovum.

Structure of ovum. The released mature ovum (Fig. 9.4-1) consists of oocyte (containing 23 unpaired chromosomes) surrounded by the inner membranous layer called zona pellucida consisting of glycoproteins and on its outer side surrounded by corona radiata consisting of granulosa cells

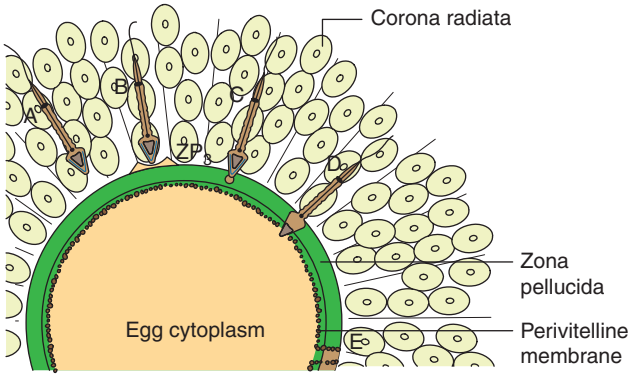


Fig. 9.4-1 Sequential events of fertilization of an ovum by the sperm: A, penetration of corona radiata; B, binding of sperm to zona pellucida; C, acrosomal reaction; D, fusion of sperm with oocyte and E, discharge of cortical granules into the perivitelline space producing vitelline block to polyspermy.

arranged in multilayers. These cells are held together by matrix composed of hyaluronic acid.

Fate of ovum. The ovum is held up at ampulla isthmic junction for 2–3 days. It remains viable for 6–24 hours after ovulation. During this period if viable sperm penetrates it then fertilization takes place and leading on to pregnancy. On the other hand, if fertilization does not occur, the ovum dies out and degenerates.

Transport of sperms in the female genital tract. After ejaculation, several million sperms (average–200 million sperms per ejaculation) get deposited in the vagina. After ejaculation, normal sperm shows flagellar movements in the fluid medium at a rate of 1–4 mm/min. Therefore, in 30–60 min, they are able to reach the fallopian tube. The motility of sperms in turn depends on:

(i) pH of medium. Neutral and alkaline pH enhances the activity of sperms. The vaginal fluid is acidic but the alkaline semen (pH 7.5) neutralizes the vaginal acidic fluid. Therefore, motility increases in cervix and in the body of uterus for next 25–40 h.

(ii) Cervical secretion. During proliferative phase of menstrual cycle and at the time of ovulation (under the influence of high level of oestrogen), cervical secretion becomes thin and watery, which favours the passage of the sperms.

(iii) Hormones. Local release of hormones as well as high concentration of certain hormones in the blood affects sperm transport. These include:

- *Oxytocin.* During coitus, stimulation of female genitalia leads to reflex release of oxytocin from the neurohypophysis; oxytocin causes propulsive movements of uterus, which help to aspirate seminal fluid from vagina into the fallopian tube.
- *Oestrogen.* It makes the cervical secretion thin and watery, thus favours transport of sperms.
- *Prostaglandins.* Prostaglandins present in the semen (contributed by seminal vesicle fluid) also increase female genital tract movements.
- *Progesterone.* After ovulation, progesterone present in the follicular fluid is released which further stimulates sperm motility.

2. Sperm capacitation

Sperm capacitation refers to the process that makes a sperm to fertilize an ovum. It takes about 1–10 h (capacitation period). Sperm capacitation occurs due to removal of certain factors, which normally remain quiescent in male genital tract. These are:

- *Cholesterol contents of acrosomal membrane.* In the female genital tract, the cholesterol contents of acrosomal membrane decreases and it becomes weak leading to easy release of enzymes from the head.
- *Calcium ions.* The membrane of sperm becomes permeable to calcium ions. The influx of Ca^{2+} acts by two ways: it makes the flagellar movements of the sperms more strong and whipish (hyperactivation of sperms) and secondly, it triggers the release of enzymes from the acrosome.

3. Fusion of gametes

The fusion of ovum and sperm involves the following steps (Fig. 9.4-1):

Chemoattraction. Chemoattraction of the sperms to ovum occurs by substances produced by the ovum.

Penetration of sperm through ovum coverings. It is made possible by the release of enzyme hyaluronidase and other proteolytic enzymes present on the acrosome of the sperm. The binding of sperm to zona pellucida glycoprotein (ZP_3) triggers acrosomal reaction (Fig. 9.4-1A & B).

- *Acrosomal reaction* (Fig. 9.4-1C). It involves release of acrosin (protease enzyme) from anterior membrane of acrosome of the sperm. Acrosin opens the penetrating pathway for passage of sperm head into the perivitelline space (space between zona pellucida and oocyte membrane) (Fig. 9.4-1D).
- *Fertilin* is a protein present on acrosomal reacted sperm which interacts with the protein present on vitelline membrane and within 30 minutes the membranes of sperm and oocyte fuse, and genetic material of sperm enters into the oocyte.
- Only one sperm can enter into the oocyte and further entry of sperms is prevented by the activation of ovum.

Ovum activation

Fusion of membranes of the gametes leads to ovum activation, which involves following events:

- *The membrane potential of the ovum decreases* (depolarization), which is followed by some structural changes in the zona pellucida.
- *Release of calcium* from intracellular egg reserve leads to exocytosis of the cortical granules (situated near the oocyte membrane) into the perivitelline space (Fig. 9.4-1E).
- *Vitelline block to polyspermy.* The spread of cortical granules along the perivitelline membrane prevents further entry of sperm into the ovum. This is called vitelline block to polyspermy.
- *Zona blockade to polyspermy.* The cortical granules contain certain substances like glycosidases and proteases. Glycosidases cause alterations in the ZP_3 receptor protein of the zona pellucida and proteases degrade the ZP_3. Both of these mechanisms cause loss of affinity of sperm for zona pellucida and thus prevent polyspermy. This is called zona block to polyspermy.

IMPLANTATION

Implantation of a fertilized ovum involves following steps:

1. Formation of blastocyst. The fertilized ovum starts dividing immediately and is called morula (16-cell stage) and blastocyst (100-cell stage). Blastocyst on cut section shows inner cell mass surrounded by a layer of cells called *trophoblast*, which is covered by zona pellucida.

The trophoblast layer consists of an inner layer (cytotrophoblast) and an outer layer (syncytiotrophoblast). The syncytiotrophoblasts secrete proteolytic enzyme that digest and liquify the endometrial cells.

2. Transportation of blastocyst in uterine cavity. In next 3–4 days, blastocyst is transported into the cavity of the uterus.

3. Implantation of blastocyst in the endometrium. The blastocyst then erodes and burrows into the endometrium (Fig. 9.4-2). Then blastocyst goes deeper and deeper into the uterus mucosa till whole of it lies within the endometrium.

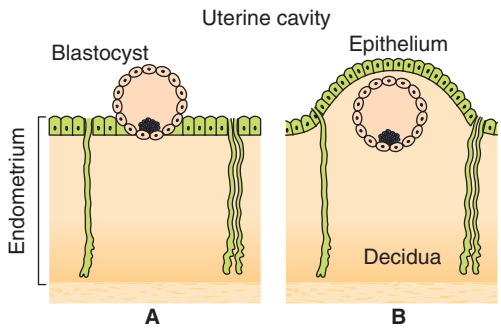


Fig. 9.4-2 A, Implantation of blastocyst in the endometrium and B, decidual reaction.

4. Decidual reaction. After implantation, the endometrium is called decidua. The stroma cells of the endometrium get enlarged, become vacuolated and filled with glycogen and lipids. These cells are called decidual cells. The stored glycogen and lipids are the source of nutrition for the embryo till placenta takes up this function. Therefore, this change in stroma cell is called decidual reaction.

IMPORTANT NOTE

Test tube baby: Fallopian tubes obstruction rendered the woman infertile. In such cases in-vitro fertilization (IVF) is made possible by adding spermatozoa to an isolated ovum. After few days, the blastocyst so formed is inserted into progesterone primed uterus for further growth. The baby born due to in-vitro fertilization is called as test tube baby.

PLACENTA AND PREGNANCY TESTS

PLACENTA

Placenta is a temporary organ formed during pregnancy. It is an important link between the mother and the fetus.

- When fully formed, the placenta (Fig. 9.4-3) is a disc-shaped structure, has a diameter of 15–20 cm and weighs about 500 g.
- After birth of the baby, the placenta is shed off along with the decidua.

The placental membrane

The maternal and the fetal blood do not mix with each other. They are separated by a placental membrane, made up of the layers of the wall of the villus. From the fetal side, these are (Fig. 9.4-4):

- Endothelium of fetal blood vessels and its basement membrane,
- Surrounding mesenchymal tissue (connective tissue),
- Cytotrophoblast and its basement membrane and
- Syncytiotrophoblast,

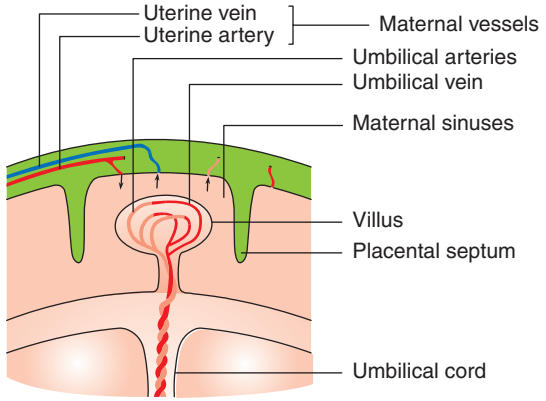


Fig. 9.4-3 Schematic diagram showing structure of placenta.

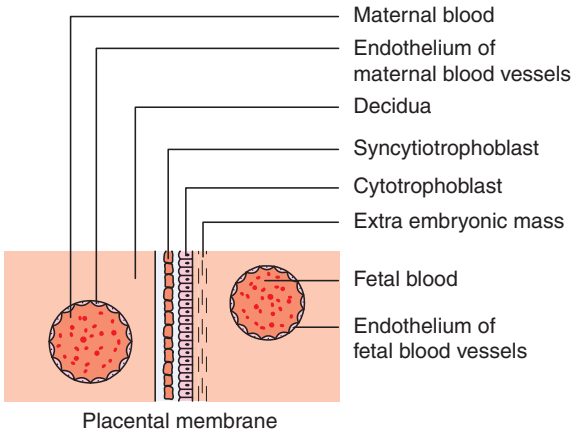


Fig. 9.4-4 Structure of the placental membrane.

- The total area of the membrane varies from 4 to 14 m^2. Its thickness is 0.025 mm in the beginning and in the later part of pregnancy it is reduced to 0.002 mm.

Functions of placenta

The fully functional placenta develops by the end of third month (12 weeks) of pregnancy. Placenta serves mainly three functions:

- Hormone secretion (endocrinal functions of placenta),
- Transport of substances between mother and fetus and
- Protection of the fetus.

A. Hormone secretion

The syncytiotrophoblast of the placenta serves as an endocrine gland. The hormones secreted by the placenta are:

- Human chorionic gonadotropins (HCG),
- Human chorionic somatomammotropins (HCS),
- Human chorionic thyrotropin,
- Placental progesterone,
- Placental oestrogens, and
- Relaxin.

1. Human chorionic gonadotropins. Human chorionic gonadotropin is a polypeptide (largest active peptide). It is secreted by syncytiotrophoblast, soon after fertilization it is detected in the maternal blood as early as 6–8 days after conception, and reaches its peak between 60 and 90 days of gestation. After this the concentration falls to a very low level and just before labour its level falls to zero (Fig. 9.4-5A). Its approximate peak value in human maternal blood during normal pregnancy is 100 IU/mL.

Physiological effects of HCG are:

- Human chorionic gonadotropin is a luteotropic hormone. Its actions are similar to luteinizing hormone (LH) of anterior pituitary hence also called second *luteotropic hormone.* It maintains the functions of the corpus luteum up to 7 weeks after conception until fetoplacental unit is able to synthesize its own oestrogen and progesterone.
- Human chorionic gonadotropin stimulates fetal testes in male fetus to secrete testosterone prior to fetal pituitary LH secretion. This testosterone and Mullerian regression factor secreted by the fetal testes is responsible for development of male genital organs and descent of testes during intrauterine life.

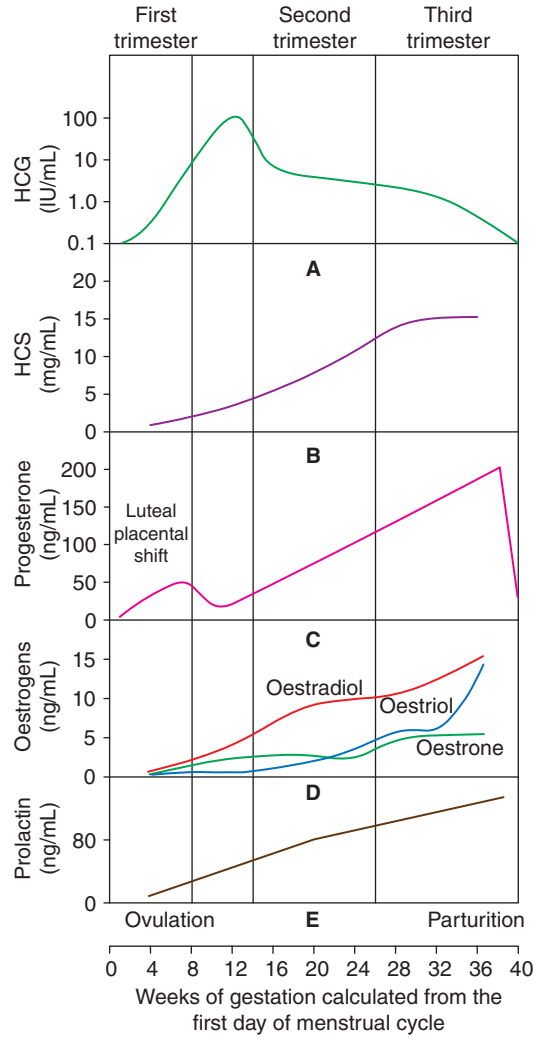


Fig. 9.4-5 Profile of plasma concentration of hormones during normal pregnancy: A, human chorionic gonadotropin (HCG); B, human chorionic somatomammotropin (HCS); C, progesterone; D, oestrogens and E, prolactin.

APPLIED ASPECTS

Clinical importance (application) of HCG is the presence of HCG in the urine, which forms the basis of all the pregnancy tests. Human chorionic gonadotropin appears in the urine as early as 10 days after gestation with 99% accuracy.

If fetus dies early then HCG disappears from the blood as well as from the urine.

2. Human chorionic somatomammotropin. The syncytiotrophoblast cells of placenta also secrete a large amount of HCS. Human chorionic somatomammotropin is protein in nature and structurally resembles to growth hormone.

Plasma concentration. The secretion of HCS begins at fifth week of pregnancy. It increases gradually throughout pregnancy and its plasma concentration is directly proportional to the weight of placenta. Its peak reaches at term and peak value is 15 mg/mL (Fig. 9.4-5B).

It functions as maternal growth hormone of pregnancy and causes deposition of protein in the tissues and brings about nitrogen, calcium and potassium retention.

3. Human chorionic thyrotropin. Human chorionic thyrotropin secreted by the placenta has properties quite similar to that of thyroid stimulating hormones. The physiological role of this substance is not very clear.

4. Placental progesterone. Progesterone is C-21 steroid hormone.

Synthesis. During early pregnancy, it is synthesized by corpus luteum of pregnancy and then by syncytiotrophoblasts of placenta (85% of total contribution). The various facts regarding synthesis of progesterone in placenta are:

- Placental syncytiotrophoblasts do not synthesize cholesterol. Therefore, cholesterol is mainly derived from maternal circulation and very little is contributed by the fetus.
- The fetus, placenta and mother, though they are independent, but constitute a functional unit called *fetoplacental maternal unit.*
- The pathways of progesterone synthesis in the fetoplacental maternal unit are shown in Fig. 9.4-6.
- In the syncytiotrophoblasts of placenta, pregnenolone is formed from the maternal cholesterol.

Pregnenolone is then oxidized by 3β hydroxysteroid dehydrogenase (3β-HSD) to progesterone.

Plasma concentration. During pregnancy, plasma concentration of progesterone rises steadily throughout gestation (Fig. 9.4-5C) reaching a maximum plateau at 30–40 weeks of gestation and its level does not fall to zero like other placental hormones. Just before the onset of labour, its level decreases.

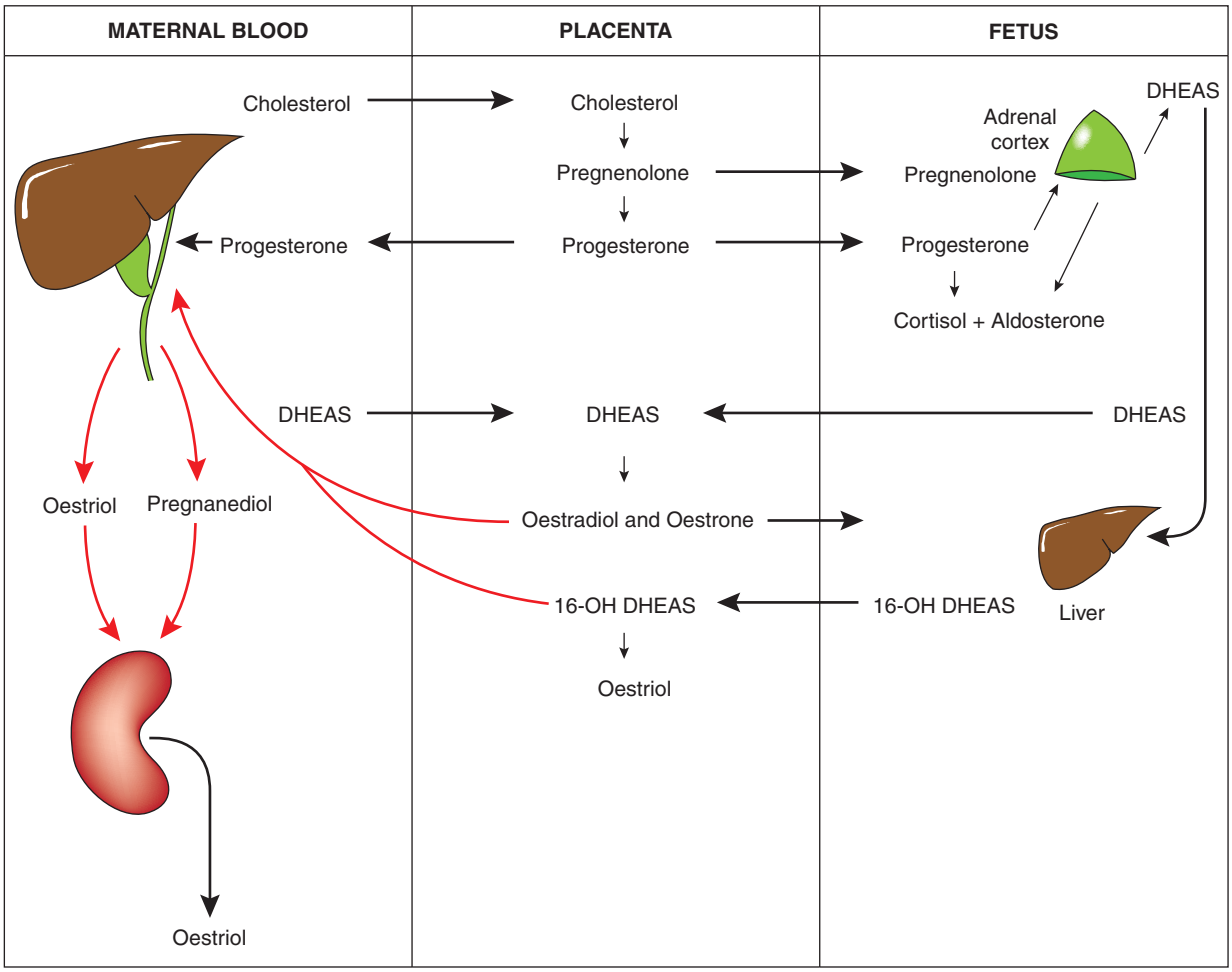


Fig. 9.4-6 Fetoplacental maternal unit for steroid hormone synthesis.

Fate and metabolism of progesterone. Progesterone synthesized by placenta diffuses back into the maternal circulation and also in the fetal circulation.

In the maternal circulation. The progesterone exerts its physiological effects and is then metabolized in the liver. The principal metabolite of progesterone is *pregnanediol*, which is glucuronised and secreted by kidneys into the urine.

In the fetal circulation. Up to 10th week of gestation, the fetal adrenal cortex (inner zone or fetal zone) cannot synthesize its corticosteroids (cortisol) because 3β-hydroxysteroid dehydrogenase (3β-HSD) enzyme system is blocked. Therefore, fetal adrenal cortex requires placental progesterone, which circulates into the fetal adrenal cortex and get hydroxylated at C-17, C-21 and C-11 positions to form aldosterone, cortisol and corticosterone (Fig. 9.4-6).

After 10th week of gestation, fetal adrenal cortex no longer depends on the placental progesterone for synthesis of corticosteroids.

Some of pregnenolone entering in fetus from the placenta along with the pregnenolone synthesized in the fetal liver is a substrate for the formation of dehydroepiandrosterone sulphate (DHEAS) and 16-hydroxy dehydroepiandrosterone (16-OH DHEAS).

Physiological effects of placental progesterone are:

(i) It helps to preserve the pregnancy by promoting the growth of endometrium. It converts secretory endometrium of luteal phase of menstrual cycle to decidual during pregnancy.
(ii) Progesterone has a marked inhibitory effect on the contractions of uterus.
(iii) It causes development of alveolar system of mother's breast. Its synergic action with oestrogen prepares the breast for lactation after the birth of the baby.
(iv) Progesterone has an immunosuppressive role in protecting the fetus.
(v) By acting as a precursor for the corticosteroid synthesis by the fetal adrenal cortex, it helps in growth and development of the fetus.

5. Placental oestrogens. Placental oestrogens are C-21 steroid hormones quantitatively oestriol is the major oestrogen of pregnancy with smaller amount of oestradiol and oestrone.

Synthesis. The steps involved in biosynthesis are shown in Fig 9.4-6.

Plasma concentration. Like progesterone, plasma oestrogen (oestriol) concentration rises throughout the gestation. Its peak value (14 ng/mL) and secretory curve parallels as that of the progesterone and maximum *plateau* is reached at 30–40 weeks of gestation.

Note. The plasma concentration of oestriol reflects the functional status of fetoplacental maternal unit activity.

Physiological effects. It mediates following effects:

(i) It causes growth and development of maternal reproductive organs (uterus increases in size, weight, length and volume both by hypertrophy and stretching of myometrium).
(ii) Oestrogen stimulates development of lactiferous ductal system in mammary glands.
(iii) It stimulates hepatic synthesis of thyroxine binding globulins, steroid-binding globulins and angiotensinogens. It also stimulates renin secretion.
(iv) Just before term, oestrogen to progesterone ratio increases and uterus is dominated by oestrogen.

6. Other placental hormones. A number of other substances which are secreted from placenta are:

- Corticotropin-releasing hormone (CRH),
- β-endorphins,
- α MSH,
- Dynorphin A,
- Gonadotropin-releasing hormones (GnRH),
- Inhibin,
- Leptin,
- Prolactin and
- Prorenin.

Exact role of the above substances during pregnancy is not yet clear. However, substances like GnRH and inhibin act in paracrine fashion to regulate HCG.

B. Transport of substances between the mother and the fetus

1. Transport of nutrients. The major function of placenta is to provide foodstuffs from mother's blood into the fetus.

By the 4th week of pregnancy, the placenta takes up the nutritive functions. The nutritive materials which are transported from mother's blood into the fetus are:

- *Glucose,*
- *Fats,*
- *Amino acids,*
- *Calcium and inorganic phosphates* and
- *Potassium,* sodium and chloride ions and substances with molecular weight less than 1000 can cross readily by simple diffusion.

2. Excretion of waste products through placenta. Excretory products, especially urea, uric acid and creatinines, etc. formed in the fetus are transported into the mother's blood and then excreted by mother's kidneys. Thus placenta also acts as a fetal kidney.

3. Diffusion of respiratory gases

(i) Oxygen transport

- Dissolved oxygen from the maternal sinuses of placenta diffuses into the fetal blood along the pressure gradient (mean pO_2 in mother's blood is 50 mm Hg whereas in fetal blood mean pO_2 is 30 mm Hg),
- The low pO_2 of the fetal arterial blood would have been a serious problem but presence of fetal haemoglobin in the RBCs which has higher affinity for oxygen than adult haemoglobin, and higher haemoglobin concentration of fetal blood (50% greater than mother), shifts the oxygen–haemoglobin dissociation curve to left (Fig. 9.4-7) and
- The fetal blood coming to placenta carries more of CO_2, which is released into the maternal blood. Therefore, the pH of maternal blood is slightly acidic as compared to the fetal blood. The haemoglobin–oxygen dissociation curve of fetal blood shifts to left and of maternal blood to the right (double Bohr's effect). All the above factors help the fetus to receive sufficient oxygen. For details page 970.

(ii) Transport of CO_2. Transport of CO_2 from fetus occurs by diffusion along the pressure gradient. The pCO_2 of fetal blood is 2–3 mm Hg higher than that of the maternal blood (as CO_2 is continuously being formed in the fetus and eliminated only through placenta). Thus placenta acts as fetal lungs.

4. Transport of antibodies. Maternal immunoglobulins are transferred into the fetus and are responsible for innate immunity.

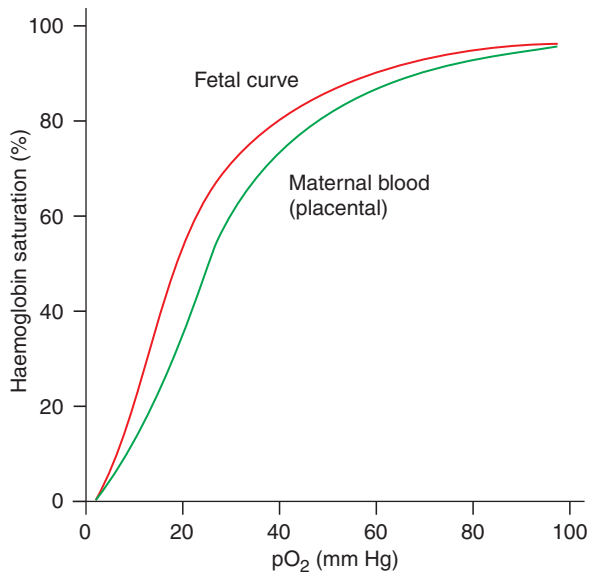


Fig. 9.4-7 Oxygen-haemoglobin dissociation curve for maternal and fetal blood.

Rh agglutinins are easily transported as compared to ABO agglutinins, that is, why the effects of Rh incompatibility are more severe.

5. Transport of harmful substances. Certain viruses and many drugs (like nicotine and barbiturates) can easily cross the placental barrier and may produce harmful effect on the fetus. Therefore, as far as possible one should avoid these drugs and smoking during pregnancy. One should also try to remain away from viral infections. HIV virus is also transmitted from mother to fetus.

C. Protection of the fetus

Placenta protects the fetus in many ways:

- It acts as a barrier for certain harmful substances,
- It provides nutrition to the fetus,
- Its hormonal secretion is responsible for the proper growth of the fetus and
- Placental progesterone decreases uterine contractions and thus protects the fetus from being expelled.

IMPORTANT NOTE

It is important to note that the fetus and the mother are two genetically different individuals and fetus is like a foreign tissue (transplant) in the mother. However, the transplant is well tolerated and not rejected. The possible reasons are:

(i) Placental trophoblast, which separates maternal and fetal tissues, does not express polymorphic MHC class I and II genes, rather it expresses HLA-G (monomorphic) genes. Therefore antibodies against fetal proteins do not develop (see page 137).

(ii) Further, production of maternal antibodies during pregnancy is reduced in general.

PREGNANCY TESTS

In an adult healthy woman, amenorrhoea is the first sign of pregnancy, but it occurs in many other conditions as well. Therefore, detection of early pregnancy is made possible by certain pregnancy tests. The pregnancy detection tests are based on presence of HCG in the urine of pregnant lady.

Gravindex test

Immunological pregnancy tests are based on the antigenic properties of HCG.

The kit for this test consists of:

- Gravindex antigen (latex particles coated with HCG), Gravindex antibodies (serum containing antibodies against HCG) and a dark coloured slide.

Procedure. This test is performed on the control and test samples of urine.

Control sample. A drop of urine sample from non-pregnant subject (containing no HCG) is mixed with a drop of antiserum containing HCG antibodies. Then it is mixed with HCG-coated latex particles. There will be agglutination because urine of non-pregnant subject does not contain antigen therefore, antibodies are not neutralized. Thus occurrence of agglutination indicates no pregnancy or pregnancy test is negative.

Test sample. A drop of urine of suspected pregnant lady (containing HCG) is mixed with a drop of antiserum (containing HCG antibodies). Then it is mixed with HCG-coated latex particles. There will be no agglutination because antibodies have been neutralised by the HCG present in the urine. Therefore, occurrence of no agglutination means positive pregnancy test.

PHYSIOLOGICAL CHANGES IN MOTHER DURING PREGNANCY

The normal average duration of pregnancy in human beings is 280 days (40 weeks) and is calculated from the first day of the last menstrual period or 256–270 days from the time of ovulation. As the pregnancy progresses, various types of extra demands are imposed on the mother's body by the growing fetus, which are met with by certain adaptations in almost all the organ systems of the body. These physiological changes include.

I. CHANGES IN GENITAL ORGANS

1. Uterus. To accommodate the growing fetus, a marked increase in the size of uterus takes place. The enlargement is mainly due to hypertrophy and to some extent hyperplasia of the myometrial smooth muscle fibres.

2. Ovaries. The follicular changes and ovulation do not occur because FSH and LH of anterior pituitary are inhibited.

3. Cervix. Endocervix gets hypertrophied, the cervical glands increase in number and their secretions form a plug that closes the cervical canal and the tough cervix becomes soft.

4. Mammary glands. Under the influence of various hormones breast enlarges in early pregnancy. Hyperplasia of ductal and alveolar tissue occurs, the areola becomes pigmented and many sebaceous glands become prominent in the areola. Nipples also become larger and pigmented.

II. WEIGHT GAIN

A woman may gain total of 10–12 kg of weight during normal pregnancy, which is contributed by:

- Fetus: 3 kg,
- Placenta and amniotic fluid: 1.5 kg,
- Uterus and breast enlargement: 1.0 kg,
- Increase in blood volume and interstitial fluid: 1.5 kg and
- Fat deposition: 3.5–4 kg.

III. HAEMATOLOGICAL CHANGES

1. Blood volume. The total blood volume increases by 30%. The plasma volume increases relatively more than that of the red cell volume, which causes haemodilution thus there is physiological anaemia of pregnancy.

2. The haematological indices show following changes:
- RBC count decreases,
- Hb concentration decreases,
- PCV decreases,
- Erythrocyte sedimentation rate increases, and
- Reticulocyte count increases.

3. Plasma proteins. The total concentration of plasma proteins decreases from 7.5 to 6 g/dL due to haemodilution. The fibrinogen level increases, but serum albumin markedly decreases.

4. Leucocytes. Total leucocyte count increases and may reach up to 20,000 µL.

5. Platelets. There occurs a slight decrease in the platelet count.

6. Coagulation factors. Pregnancy seems to be a hypercoagulable state due to an increase in following: fibrinogen, and factors VII, VIII, IX and X. The hypercoagulability of the blood plays an significant role of haemostasis at the time of separation of placenta during delivery.

IV. CARDIOVASCULAR SYSTEM CHANGES

1. Position of heart. The gravid uterus pushes the diaphragm upwards resulting change in the position of heart.

2. Heart rate. Heart rate also increases by 10–12 beats/min.

3. Cardiac output. Cardiac output increases from 5–7 L/min at 20 weeks of gestation.

4. Blood pressure may show following changes:
- *Systolic blood pressure.* In normal pregnancy, there is either no change in systolic pressure or some fall may occur.
- *Diastolic pressure* decreases and by 16–20 weeks of pregnancy its value is lowest. Then it starts rising and comes back to normal.

5. Blood flow. Blood flow to skin, uterus and kidneys increases to meet the demands.

6. Venous pressure. The gravid uterus exerts pressure on the pelvic veins, abdominal veins and femoral veins, thus

increasing the venous pressure. The rise in femoral venous pressure results in oedema in feet (common occurrence).

V. RESPIRATORY SYSTEM CHANGES

1. Hyperventilation. High levels of plasma progesterone during pregnancy increase the sensitivity of respiratory neurons to CO_2 resulting in hyperventilation.

2. Gas exchange. Gas exchange across the alveoli is greatly enhanced due to a marked increase in the pulmonary blood flow.

3. Oxygen consumption. Oxygen consumption of body increases by 15% to meet the demands of growing fetus and for the extra work of heart, uterus and other tissue.

VI. URINARY SYSTEM CHANGES

Kidney functions show following changes:

1. Renal blood flow. There is a marked increase in renal blood flow due to increase in the cardiac output and vasodilatation.

2. Glomerular filtration rate (GFR) increases by 50% due to an increase in renal blood flow and solute load.

3. Renal tubular absorptive capacity for sodium and chloride ions also increases by 50% due to high level of steroid hormones secreted by the placenta and the adrenal cortex.

4. Glycosuria is a common physiological phenomenon during pregnancy.

5. Proteinuria occurs due to increase in excretion of proteins.

6. Water balance. During later months of pregnancy, excess of water is retained due to:

- Decreased protein concentration and
- Retention of sodium.

7. Acid–base balance. Hyperventilation during pregnancy results in respiratory alkalosis. Kidneys, therefore, compensate for it by excreting more HCO_3^- ions in the urine.

VII. GASTROINTESTINAL SYSTEM CHANGES

1. GIT secretions. Hypochlorhydria is very common due to decreased gastric secretion.

2. GIT motility decreases under the influence of hormones resulting in delayed gastric emptying.

3. Gall bladder functions. Gall bladder increases in size and empties its contents at a very slow rate.

4. Liver functions are also altered during pregnancy. The fibrinogen synthesis increases and albumin decreases thus plasma A:G ratio is also altered.

5. Morning sickness. Anorexia, nausea and vomiting are very common in early pregnancy (first trimester) especially in the morning hours hence known as morning sickness. The cause for the morning sickness is not known.

6. Glucose tolerance curve also shows disturbances. It becomes diabetic type due to glucose being rapidly absorbed from the intestine.

VIII. METABOLIC CHANGES

1. The basal metabolic rate (BMR) of the pregnant woman increases by about 15% during later half of the pregnancy.

2. Protein metabolism. When the diet is balanced and adequate then there is nitrogen retention and positive nitrogen balance. The proteins are deposited in the uterus, breast, in the fetus and in the placenta.

3. Carbohydrate metabolism shows following changes:

- Blood glucose level increases due to rapid absorption from the gut.
- *Glycosuria* is of common occurrence due to the increase in GFR and decrease in renal threshold for glucose.
- Ketosis may occur due to anorexia and excessive vomiting.

4. Fat metabolism. About 3–4 kg of fat is deposited in the body during pregnancy. There occurs an increase in plasma concentration of cholesterol, phospholipids and triglycerides.

5. Mineral metabolism depicts following changes:

- *Calcium and phosphorus.* In normal pregnancy, mother retains about 50 g of extra calcium and 30–40 g of phosphorus. These are deposited in the fetus and also retained in the mother stores (skeleton).
- *Iron metabolism.* Iron requirement tremendously increases during pregnancy and lactation.

IX. ENDOCRINE SYSTEM CHANGES

Almost all the endocrine glands of the mother react substantially during pregnancy. Firstly, due to increased metabolic load on the mother and secondly in response to the hormones produced by the placenta and fetus.

X. CHANGES IN THE SKIN

1. Hyperpigmentation occurs on the face (butterfly pattern known as chloasma), areola, nipple and midline of abdomen (linea alba) extending from pubic symphysis to xiphisternum. The hyperpigmentation is related to increased

secretion of adrenocorticotropic hormone (ACTH) and melanocyte-stimulating hormone (MSH) during pregnancy.

2. Stria gravidarum. These are linear scars present on the lower abdomen due to stretching of skin.

XI. PSYCHOLOGICAL CHANGES

The nervous system shows mild changes in the form of craving for particular types of food item, alterations in the behaviour, emotions and mood. In few cases, true psychosis may also develop but cause is not known.

PHYSIOLOGY OF PARTURITION

Parturition is the process by which baby is born. It involves preparation for child birth, act of child birth and recovery from child birth. It is difficult to predict the exact date of onset of labour. It may occur any time between 37th and 40th weeks of gestation. The uterine myometrium and cervix play an important role for this process.

MECHANICS OF PARTURITION

From the functional point of view, mechanics of parturition mainly involves:

- Uterine contractions and
- Cervical dilatation.

UTERINE CONTRACTIONS

The uterus, which remains quiescent during the period of pregnancy, becomes progressively more and more excitable towards the end of pregnancy, until finally it begins strong rhythmical contractions with such a force that expels the fetus.

After 30th week, the activity gradually increases. The characteristics of these contractions are:

- *Start of contraction.* The uterine contractions during labour start at the top of the fundus of the uterus and spread on to the body part.
- *Force of contractions* is high in the fundus and body part of the uterus, and comparatively weak in the lower segment near the cervix.
- *Frequency of contractions* in early labour is only once in 30 min (1/30 min) and as labour progresses it increases to one in 3 min (1/3 min).
- *Pressure during contractions* may rise up to 30–35 mm Hg.
- *Period of relaxation* follows each contraction.

The uterine contractions are intermittent and are beneficial for the fetus otherwise strong and continuous contractions sometimes impedes blood supply through placenta and would cause fetal death.

Control or regulation of uterine contractility

The exact cause of increased uterine activity is not known, but few factors which lead towards parturition are:

- An increase in number of oxytocin receptors on the cells of the uterine smooth muscle during the final weeks of pregnancy.
- Increased synthesis of contractile proteins in the myometrial cells.

CERVICAL DILATATION

Throughout pregnancy, cervix remains as a rigid structure, but at the time of parturition certain structural and biochemical changes occur and the cervix becomes soft. This is known as cervical ripening. It allows the cervix to stretch when uterine contractions start.

The changes in the cervix which are responsible for its dilatation are:

- Breakdown of collagen fibres,
- Increase in the amount of hyaluronic acid, having high water retaining capacity,
- Decrease in the amount of dermatan sulphate and

These changes are mainly under the influence of prostaglandins.

CONTROL OF PARTURITION

The mechanisms responsible for onset of labour in human are still not understood exactly. The control of parturition includes the role of (Fig. 9.4-8):

- Hormonal factors and
- Mechanical factors.

A. HORMONAL FACTORS

The hormonal changes that initiate the parturition and that cause increased excitability of uterine musculature are:

1. Activation of fetal hypothalamic-pituitary-adrenal axis. Recently, it has been hypothesized that the initial signals for the onset of labour comes from the fetus only and due to some unknown factors, CRH secretion in the fetus resulting in an increase in ACTH secretion few days before parturition. ACTH causes fetal adrenal cortex to secrete large amount of androgens, which are converted to oestrogen in the placenta, DHEAS and cortisol.

The above changes lead to an altered oestrogen–progesterone ratio.

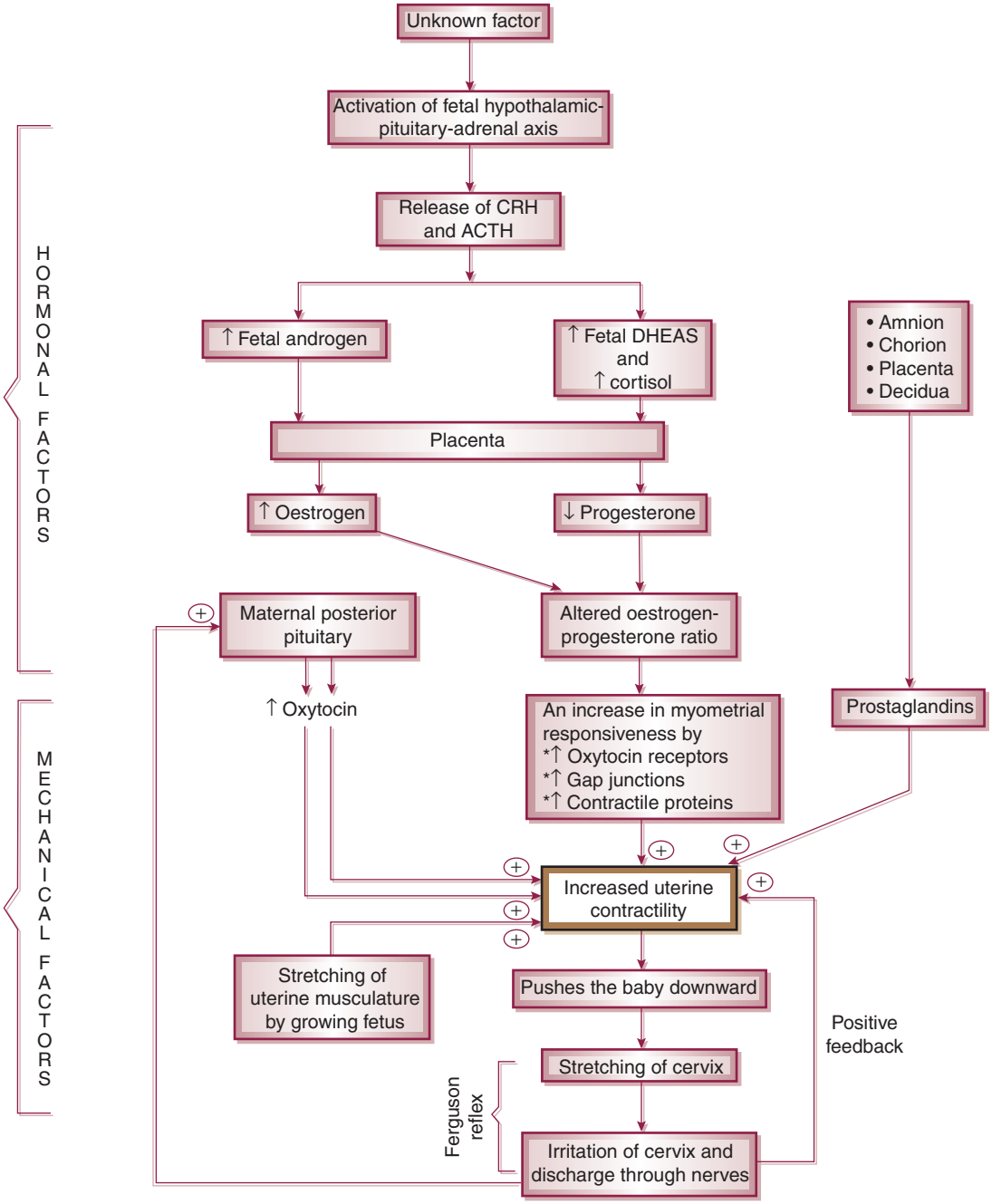


Fig. 9.4-8 Summary of control of parturition depicting role of hormonal and mechanical factors.

2. Role of altered oestrogen–progesterone ratio. The altered oestrogen–progesterone ratio also causes:

- An increase in release of oxytocin from maternal posterior pituitary,
- An increase in number of oxytocin receptors in myometrium,
- An increase in prostaglandin synthesis and
- An increase in synthesis of myometrial contractile proteins.

3. Role of oxytocin and prostaglandins. It has been suggested that the altered oestrogen–progesterone ratio leads to prostaglandin synthesis in human pregnancy from the placenta, amnion, chorion and decidua. The prostaglandins enhance the force of oxytocin-induced uterine contractions.

B. MECHANICAL FACTORS

Mechanical factors that increase the contractility of uterus include:

1. Stretch of uterine musculature usually increases their own contractility (myogenic theory). During pregnancy, movements of the fetus cause stretching. As the pregnancy advances with growing fetus stretch further increases leading on to uterine contractility.

2. Positive feedback effect. Stretching and irritation of cervix is particularly important because of positive feedback effect through initiation of the reflex increase in uterine contractility.

The positive feedback mechanism continues until the baby is expelled.

3. Role of Ferguson reflex. Once labour is started, the uterine contractions dilate the ripened cervix. The cervical dilatation in turn sets of signals in afferent nerves that increase oxytocin secretion from the posterior pituitary. This is called *Ferguson reflex.*

Clinical application. The obstetricians frequently induce labour by rupturing the membrane so that the head of the baby stretches the cervix more forcefully than usual and thus initiate it, leading to initiation of positive feedback effect and Ferguson reflex.

The control of parturition is summarized in Fig. 9.4-8.

Chapter 9.5

Physiology of Lactation

FUNCTIONAL ANATOMY OF BREAST
- Control of breast development and growth

PHYSIOLOGY OF LACTATION
- **Phases of lactation**
  - Mammogenesis
  - Lactogenesis
  - Galactokinesis
  - Galactopoiesis
- **Importance of lactation**
  - Advantages of breastfeeding to the baby
  - Advantages of breastfeeding to the mother

FUNCTIONAL ANATOMY OF BREAST

Breastfeeding is the characteristic feature of all the mammals including human beings. It has evolved as the best method of nourishing the newborn. The mammary glands (the secondary sex organs) play an important role in the lactation process.

Mammary glands are present in both the sexes; in males they remain rudimentary but in females they are well developed after puberty.

Gross anatomy. The fully developed breast is a soft, rounded, elevated structure present over the pectoral region having central dark pigmented area (areola). The central part of areola, projected above the surface, is called nipple.

Histological structure. Each mammary gland is covered by an overlying skin and underlying it discrete masses of glandular tissue is present in the connective tissue consisting of stroma and adipose tissue (Fig. 9.5-1).

- *The mammary glands* consist of 15–20 lobes and each lobe has a number of lobules.
- *The glandular tissue* mainly consists of alveoli having secretory cells.
- *The secretions* from these cells are poured by apocrine manner and by exocytosis into the ducts (lactiferous ducts). About 15–20 ducts open at the summit of nipple, and just before opening lactiferous ducts show a dilatation called lactiferous sinus.

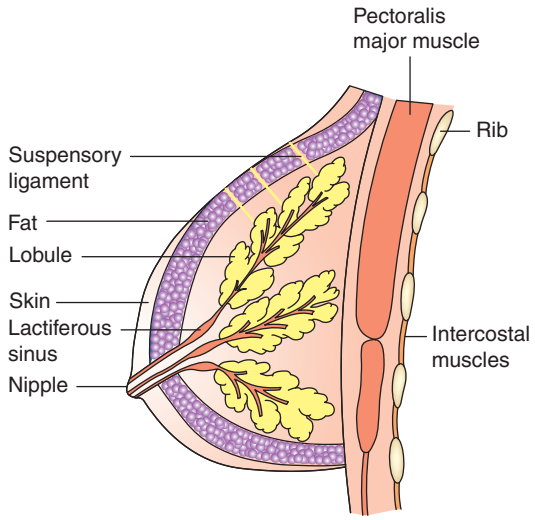


Fig. 9.5-1 Structure of mammary gland.

- *The smaller ductules* are lined by a single layer of columnar epithelial cells whereas large ducts are lined by one or two layers of cells and near the opening at the nipple these are lined by squamous cells.
- *Around the alveoli,* ductules and lobules are present in the myoepithelial cells. They squeeze the contents and pour their secretions into the ductules.
- *Electron microscopically,* the secretory cells are seen to contain both rough and smooth surface endoplasmic reticulum, numerous mitochondria, prominent Golgi apparatus and lysosomes. The proteins are present in

the cytoplasm as membrane bound vesicles and fat is stored as large globules.

Breasts at birth. At birth, the mammary glands are rudimentary consisting of tiny nipple and few ducts radiating from it.

Breasts at puberty. From birth to puberty, the mammary glands remain quiescent. During puberty, following changes occur:

At thelarche, i.e. at the time of puberty (9–11 years of age), before the start of menses. The breast starts developing and get enlarged. During this stage, only duct system proliferates and shows branching.

At menarche, i.e. after the onset of menses, cyclic growth of mammary glands (period of growth followed by quiescence) occurs in each menstrual cycle. The growth period further corresponds to phases of menstrual cycle.

- *In proliferative phase* (or oestrogen phase), the duct cells proliferate and continue throughout rest of the cycle.
- *In luteal phase* (progestational phase), progesterone stimulates the proliferation of terminal ductules, so there is formation of glandular tissue.

Thus there is a progressive growth of breast in successive cycles, along with modelling of the breast by fat deposition in the adipose tissue.

Breasts in pregnancy. During pregnancy, remarkable growth of both ductal and glandular systems occurs. It is only during first pregnancy that glandular tissue develops fully.

In first half of pregnancy, the duct system proliferates and shows extensive sprouting and branching along with growth of stroma and deposition of fat.

In second half of pregnancy, there is enormous growth of glandular tissue.

The extensive growth of mammary glands during pregnancy is known as *mammogenesis* or preparation of breast for lactation.

Breasts during lactation. After child birth, the alveolar cells get enlarged and distended and start forming milk (lactogenesis).

Involution of breast. After a normal period of lactation (7–9 months), the alveolar epithelium undergoes apoptosis and glands revert back to pre-pregnant stage.

CONTROL OF BREAST DEVELOPMENT AND GROWTH

Various hormones necessary for full growth and development of mammary glands at various stages are (Fig. 9.5-2):

1. Oestrogen. It is primarily responsible for the ductal growth and fat deposition. It also causes thickening of nipples.

2. Progesterone. The development of glandular tissue mainly depends on progesterone. Both oestrogen and progesterone work best with co-operation of hypothalamo-pituitary-adrenal cortex axis.

3. Other hormones including growth hormone, thyroxine, cortisol and insulin enhance overall growth and development of mammary glands at all stages.

4. Corpus luteal and placental hormones, particularly oestrogen, progesterone, human chorionic somatomammotropic hormone (HCS, or HPL), are essential for further growth of breast during pregnancy.

5. Prolactin. It is another very important hormone for the development of breasts during pregnancy and lactation. It acts on mammary gland tissue which has already grown under the influence of oestrogen and progesterone.

PHYSIOLOGY OF LACTATION

PHASES OF LACTATION

The physiology of lactation can be divided into four phases:

- Preparation of breast for milk secretion (mammogenesis),
- Synthesis and secretion of milk (lactogenesis),
- Expulsion of milk (galactokinesis) and
- Maintenance of lactation (galactopoiesis).

MAMMOGENESIS

During pregnancy. The breast develops fully and is prepared for milk secretion after delivery.

LACTOGENESIS

Stages of lactogenesis

The process of milk secretion occurs in two stages:

Stage I. In later few weeks of pregnancy, a small amount of fluid is secreted in the alveolar cells. Its rate of secretion is only 1/100th as that of milk secretion in postpartum period. The stage I secretion occurs due to high plasma levels of prolactin and placental HCS. But due to suppressive lactogenic action of oestrogen and progesterone, free flow of milk never occurs during pregnancy (Fig. 9.5-3).

Stage II. It is the initiation of lactation after child birth. Immediately after the baby is born, sudden loss of oestrogen

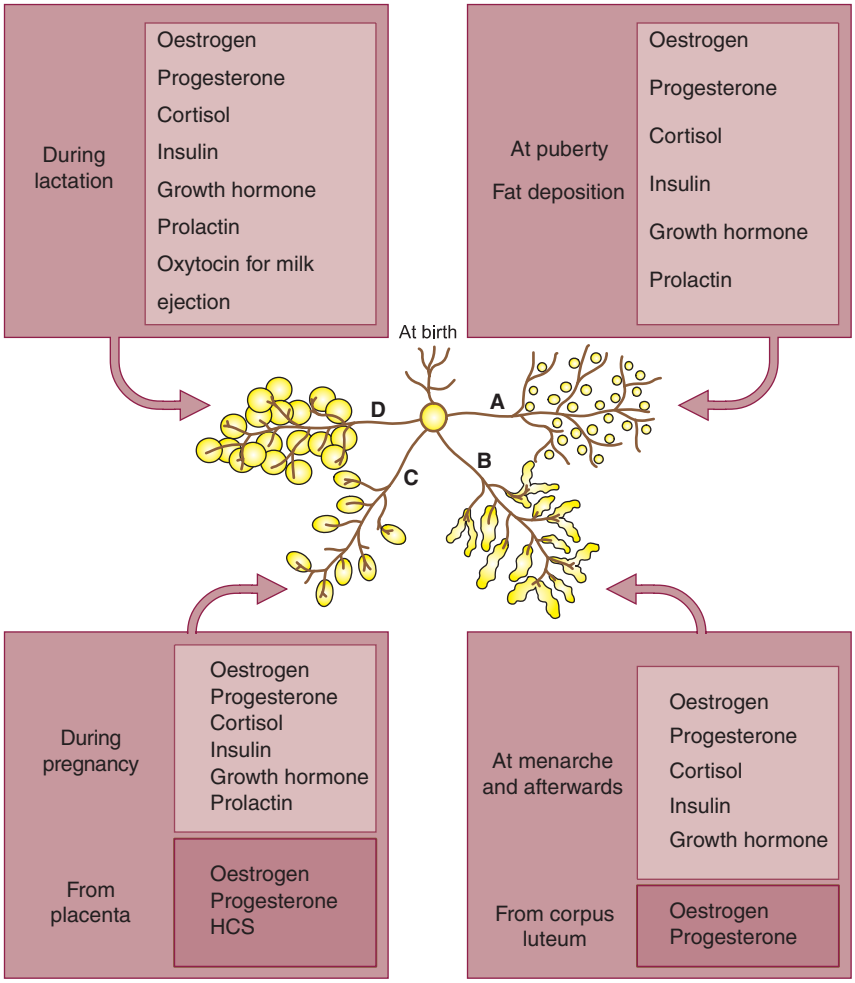


Fig. 9.5-2 Schematic diagram of hormonal control of breast development during different stages: A, at puberty; B, at menarche and afterwards; C, during pregnancy and D, during lactation.

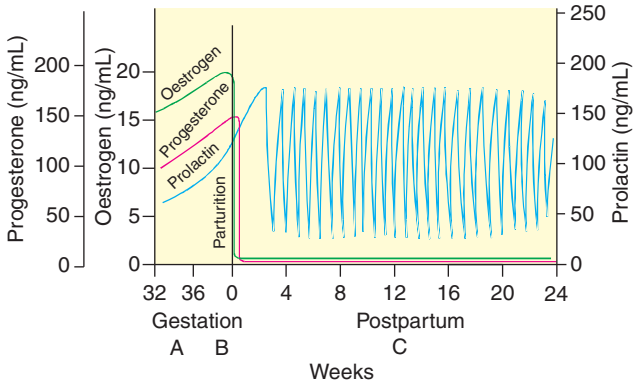


Fig. 9.5-3 Changes in rate of secretion of oestrogen, progesterone and prolactin: A, in late pregnancy; B, before parturition and C, during postpartum period.

and progesterone secretion by the placenta allows the lactogenic effect of prolactin. In this stage:

- The secretion rate increases to 500–750 mL/day and
- In next 1–7 days, the breasts begin to secrete milk instead of colostrum.

Human milk

Types of human milk. The nature and composition varies with postpartum period. Therefore, the human milk is of three types:

1. Colostrum is deep yellow coloured fluid secreted by the mammary glands during first few days of postpartum period, it contains:

- High protein contents (8.5 g/dL), rich in immunoglobulins and lactoferrin and
- Granular bodies (colostrum corpuscles)—consisting of alveolar cells and leucocytes loaded with fats.

The colostrum is easily coagulated into solid masses.

2. Transition milk or intermediate milk. It is secreted from 6th to 15th day of postpartum period. The nature and composition of the secretion changes from colostrum to mature milk. Hence it is called transition milk.

3. Mature milk is formed from 15th day of postpartum onwards and continues during the whole lactation period (7–9 months).

Composition of human milk. Human milk contains 88.5% water and about 11.5% solids. The solids include both organic and inorganic constituents. The composition of mature human milk, colostrum and cow's milk is shown in Table 9.5-1.

Note. Human milk is balanced diet as it contains first class proteins (caseinogen and lactalbumin), carbohydrates fat, mineral salts and vitamins. Therefore it is an ideal food for the baby.

EXPULSION OF MILK OR GALACTOKINESIS

Though milk is secreted continuously into the alveoli of the breast, but it does not flow continuously from alveoli into the duct system. It depends upon the suckling reflex and some local mechanisms acting within the breast.

Suckling reflex

It is a neuroendocrinal reflex. The characteristic features and mechanism of suckling reflex (Fig. 9.5-4) are:

- When baby suckles, the sensory nerve endings or receptors located in skin of areola and nipple get stimulated.
- The sensory impulses are transmitted to the hypothalamus through somatic nerves (from nipple and areola to spinal cord and then to hypothalamus). The activation of hypothalamus causes release of oxytocin and prolactin from the pituitary gland.
- The oxytocin is carried to the breasts through blood where it causes contraction of myoepithelial cells that surround the outer wall of the alveoli. This process is called *milk ejection* or *milk expulsion* or *milk let down.*
- Another important observation is that suckling of one breast causes milk flow in the other breast also.

Table 9.5-1 Composition of human, colostrum, mature milk and cow's milk

Content	Human milk		Cow's milk (g/dL)
	Colostrum (g/dL)	Mature milk (g/dL)	
1. Proteins	8.5	1.0–2.0	3.5
2. Carbohydrates (lactose)	3.5	6.5–8.5	4.7
3. Fat	2.5	3.0–5.0	3.5
4. Minerals Na^+, K^+, P^+ and Cl^-	0.35	0.18–0.25	0.75
5. Calcium	–	0.03	0.14

- Even stimuli, such as sight, sound or crying of infant and thought of their infants also cause milk ejection, indicating the psychological component in the neuroendocrine reflex.

Note. In case of engorgement of breasts after delivery (which is a very painful condition, suckling becomes difficult). Oxytocin administration leads to free flow of the milk.

Inhibition of milk ejection

- Many psychogenic factors in the form of psychological stress, pain etc. inhibit milk ejection by inhibiting oxytocin release.
- Alcohol is also a potent inhibitor of oxytocin.

MAINTENANCE OF MILK SECRETION OR GALACTOPOIESIS

Maintenance of milk secretion or *galactopoiesis* depends upon the surge in prolactin secretion. After few weeks of child birth, prolactin level falls to its basal value; however, in nursing mothers neuroendocrine reflex causes 10–20 fold surge in prolactin secretion that lasts for 1 h. Each time when baby suckles, the impulses from nipple and areolar receptors are transmitted by somatic nerves up to the hypothalamus, which cause 1–20 fold surge in the prolactin secretion.

The amount of milk production is related to infant's demand.

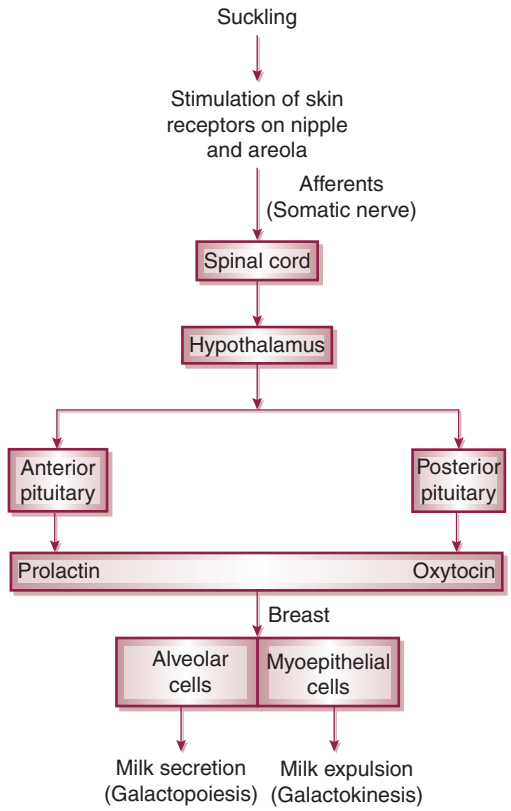


Fig. 9.5-4 Mechanism of suckling reflex.

IMPORTANCE OF LACTATION

Breastfeeding is being advocated all over the world because of its advantages both for the baby as well as for the mother.

ADVANTAGES OF BREASTFEEDING TO THE BABY

1. Balanced diet. Human milk contains proteins, carbohydrates, fat, mineral salts (calcium and phosphorus) and vitamins. So, it is a natural balanced food for the newborn.

2. Protection against infections. Human milk has high count of lymphocytes, neutrophils and macrophages and high content of lysozymes and immunoglobulins. All these substances due to anti-infection property confer non-specific as well as specific immunity.

3. Easily digestible. Human milk because of its following digestive properties can be easily digested by the newborn babies:

- Casein is easily digestible,
- Lactoferrin prevents iron overloading,
- Folate and cobalamin binding proteins assist in absorption of corresponding vitamins,
- Higher concentration of lactose promotes calcium absorption,
- Lipases assist in lipid digestion because lipid digestion is poor in newborn babies.

4. Growth factors. Growth promoting factors, like epidermal growth factor, insulin and somatomedin-C are present in the human milk.

5. Other advantages of breastfeeding to baby are:

- It is sterile,
- It is convenient to give at right temperature,
- It is inexpensive and
- Chances of allergy to breast milk are rare.

ADVANTAGES OF BREASTFEEDING TO THE MOTHER

1. Lactational amenorrhoea. Due to high plasma level of prolactin during lactation, there occurs suppression of FSH and LH. Therefore, after birth of the baby menstruation and ovulation do not start.

2. Involution of uterus. The oxytocin released during each session of breastfeeding also acts on the uterine myometrium, and helps it to involute during postpartum period. The proper involution of uterus protects it against infections.

3. Protection against breast engorgement. Breastfeeding does not allow the milk to stagnate, thus preventing the breast engorgement, which is highly painful condition. The stagnant milk acts as a favourable medium for bacterial growth. Therefore breastfeeding protects against infection.

4. Protection against obesity. Body fat is used for milk synthesis, therefore, there are less chances of becoming obese after pregnancy.

5. Emotional bonding and psychological satisfaction is enhanced by breastfeeding.

6. Protection against cancer. The chances of breast cancer are more in those women who have never borne children. This is related to the hormone oestrogen, which is responsible for aetiology of breast cancer. Therefore, prolonged lactation provides protection against breast cancer.

Chapter 9.6

Physiology of Contraception

INTRODUCTION

Aims of contraception. Contraception refers to prevention of pregnancy. The aims of contraception are:

- The main aim of contraception is family planning to check the enormous increase in population growth, which is the root cause of socioeconomic problems of poor and developing countries, like India.
- Certain contraceptive measures are important to prevent the sexually transmitted diseases like AIDS.
- Contraceptives are also recommended on medical grounds to control the stress of pregnancy, labour and lactation in women suffering from heart diseases, etc.

Methods of contraceptions can be broadly grouped as:

- Spacing methods and
- Terminal methods.

Both types of contraceptive measures are available for use by females as well as males; therefore, these can be described as:

- Contraceptive methods in females and
- Contraceptive methods in males.

CONTRACEPTIVE METHODS IN FEMALES

SPACING METHODS

The spacing methods increase the gap between two pregnancies. These include:

- Rhythm method,
- Barrier methods,
- Chemical methods and
- Intrauterine contraceptive devices.

RHYTHM METHOD

Rhythm method is also known as *calender method* or *safe period method* or *natural method.* This method of contraception depends on the time of ovulation. In a woman having regular menstrual cycle, ovulation occurs on 14th day of the cycle. After ovulation, ovum remains viable for 48–72 h. Similarly, after ejaculation sperms remain alive for 24–48 h. Thus pregnancy occurs only if coitus is performed during this period. This is the period of high fertility and is called as *dangerous period.*

Therefore, to avoid pregnancy intercourse should be avoided in the dangerous period. Rest of the cycle, i.e. 5–6 days

after bleeding phase of menstrual cycle and 5–6 days before the next cycle is the *safe period* (period of least fertility). This method of contraception is successful only if menstrual cycles are regular and woman knows the exact time of ovulation by keeping a record of basal body temperature.

Disadvantage of this method is that it is the most unreliable method when the menstrual cycles are irregular and time of ovulation is variable.

BARRIER METHODS

Barrier methods of contraception prevent the meeting of ovum and sperms after coitus. These include:

1. Mechanical barriers

The mechanical barriers used as contraceptive are: diaphragm and cervical caps (Fig. 9.6-1A and B).

Advantages. These devices are inexpensive and usually do not require any medical consultation.

Disadvantages of mechanical barriers include:

- Failures are quite common because chances of displacement of the device are very high.
- Some women get cervicitis (inflammation of cervix) and local irritation.

2. Chemical barriers

Chemical barriers refer to spermicidal agents, which can destroy the sperms when applied in the female genital tract before coitus. The common spermicidal agents used are:

- Ricinoleic acid (oldest)
- Nonoxynol-9
- Octoxynol-3

These spermicidal agents are available in various forms such as: foam tablets, pastes, creams, jellies and vaginal sponges. Vaginal sponge is a polyurethane sponge impregnated with nonoxynol-9. It is available by the trade name 'TODAY'.

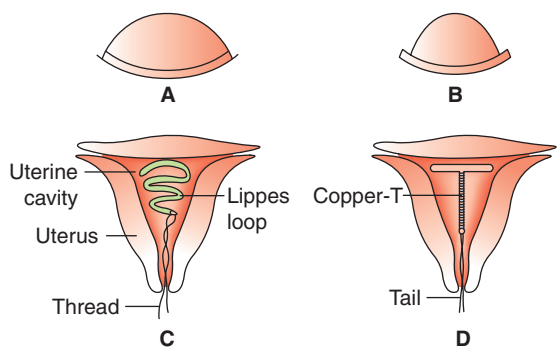


Fig. 9.6-1 Female contraceptive devices: A, vaginal diaphragm; B, cervical cap; C, Lippe's loop and D, Copper-T.

3. Combined methods

As mentioned above, mechanical barriers (diaphragm and cervical caps) along with the spermicidal agents give good protection.

CHEMICAL METHODS

Chemical methods for contraceptions are used in various forms like locally applied chemicals (in the form of cream, jellies etc.) and taken as drugs (either orally or in injectable form or as implants).

(a) Oral contraceptives

Oral contraceptives/steroidal drugs are most widely used contraceptive measure by the women all over the globe. These are recommended in women of younger age group (up to 35 years).

Mechanism of action. In general, oral contraceptives contain synthetic preparation of oestrogen and progesterone and when taken orally, the plasma concentration of these hormones rises. The raised levels of these hormones by their negative feedback effect, act on the anterior pituitary to inhibit the release of gonadotropins (FSH and LH) and thus inhibit ovulation.

Types of pills. The oral contraceptives are available in different types of pills:

- Combined pill (classical pill)
- Sequential pill
- Minipill
- Post-coital (morning after) pill

1. Combined pill or classical pill

Composition. It contains both oestrogen and progesterone. Oestrogen (20–50 µg) in combined pills is usually *ethyl estradiol* or mestranol (methoxy derivative of ethyl oestradiol) and progesterone (0.5–2 mg) like norethisterone, norgestrel or levonorgestrel.

Availability. The combined pills are available under two brand names; MALA-N (packet of 21 tablets) and MALA-D (packet of 28 pills, out of which 21 are white coloured of hormones and 7 are brown coloured containing *ferrous fumarate*).

Dosage. The combined pills are taken orally every day at fixed time (preferably at night before going to bed) for 21 days, starting from fifth day of menstrual cycle to 25th day followed by a gap of 7 days in case of MALA-N. During this gap period bleeding occurs. This bleeding is not a menstrual bleeding as it occurs due to withdrawal of the hormones; therefore, it is called *withdrawal bleeding.*

Mechanism of action. Combined pill acts by three ways:

(i) It prevents ovulation,

(ii) It prevents implantation, if ovulation occurs and ovum is fertilized by a sperm and

(iii) It makes the cervical secretion thick and viscid and thus prevents entry of sperms in the female genital tract.

Note. In combined pills, these days *phase regimens* (biphasic or triphasic) are preferred, because they are more physiological than fixed dose preparations.

2. Sequential pill

Composition. These pills contain high dose of oestrogen along with a moderate dose of progesterone.

Dosage. Only oestrogen is given starting from fifth day of menstrual cycle to 15th day and then followed by both (oestrogen + progesterone) for next 5 days.

Note. Nowadays the sequential pill is not used because of high incidence of *endometrial carcinoma.*

3. Minipill

Minipill (progesterone only) or *micropill.*

Composition. These preparations contain low doses of progesterone (norethisterone—0.35 mg or norgestrel—0.075 mg).

Dosage. The regimen of these pills is that the pill should be taken daily through whole of the menstrual cycle.

Mechanism of action. Minipill prevents fertility without inhibiting ovulation. It acts on the cervical mucosa (makes it thick), and also decreases motility of fallopian tubes.

4. Post-coital pill (morning after pill)

As the name indicates, it is recommended within 72 h of the unprotected intercourse.

Dosage. Double dose of combined pill (2 pills) should be taken immediately followed by another double dose (two pills) after 12 h.

Indications. This method of contraception should be used only in emergency cases, like rape, contraceptive failure and unprotected sex.

Mechanism of action. Possible mechanisms involved are:

- It causes hypermotility of the fallopian tubes and of uterus and thus prevents fertilization and implantation.
- If ovulation and fertilization has occurred, then it prevents implantation of the blastocyst.

Disadvantages. Routinely, this method of contraception is not practised because of various side effects like nausea and vomiting.

Advantages and disadvantages of oral contraceptives

Advantages. Oral contraceptives have 100% effectivity.

Disadvantages. Although oral contraceptives are extensively used, but its prolong use leads to certain adverse effects as:

- Hypertension,
- Risk of thromboembolism,
- Metabolic effects like diabetes and obesity, and
- Carcinogenic effects (carcinoma breast and carcinoma cervix).

Contraindications

Absolute contraindications for the use of oral contraceptives are:

- Woman having carcinoma of breast or of uterus,
- Liver diseases and
- Hyperlipidaemia.

Relative contraindications. Oral contraceptive should not be given to woman of age group above 35 years.

(b) Depot preparations

Depot preparations are long-acting drugs and are highly effective. These are available in three forms:

- Injectable preparations
- Subdermal implants
- Vaginal rings

Advantages and disadvantages of depot preparations

Advantages. As depot preparations are long-acting drugs, therefore to avoid daily intake of oral pill, these preparations are preferred and also the contraceptive effectivity lasts for longer period.

Disadvantages of depot preparation are that sometimes they lead to sterility and alterations in menstrual bleeding pattern.

INTRAUTERINE CONTRACEPTIVE DEVICES

Intrauterine contraceptive devices (IUCDs) are inserted into the uterine cavity for long-term contraception. The devices are usually made up of inert materials like plastic, polythene and metal.

Lippes loop (Fig. 9.6-1C). It is a serpentine or S-shaped device made up of plastic to which is attached a fine nylon tail. The plastic used is non-toxic and non-tissue reactive. A small amount of *barium sulphate* is also present in the plastic material to allow its radiographic observation. Lippes loop is available in different sizes.

Copper-T. Copper-T is the most commonly used IUCD in India. As the name indicates it is made up of copper and its shape resembles the letter T. Like Lippes loop it is also attached with a nylon thread (tail) (Fig. 9.6-1D).

Insertion. Most ideal time for its insertion is during menstruation or within 10 days of the beginning of menstruation, because the diameter of cervical cavity at this time is greater. It can also be inserted during first week after the delivery.

Mechanism of action. Copper-T acts by following ways:

Prevents implantation and growth of fertilized ovum by evoking aseptic inflammation and thus making endometrium unsuitable for implantation.

Advantages of IUCDs are:

- This method of contraception is quite safe, effective and reversible. Intrauterine contraceptive devices can be easily pulled out or removed when contraception is not required.
- Provides long-term contraception without adverse effects.

Disadvantages of IUCDs are:

- In some cases may cause heavy bleeding,
- The IUCD may come out accidentally, when not inserted properly and
- Risks of ectopic pregnancy are there.

TERMINAL METHODS

Terminal method of contraception means permanent sterilization, which can be achieved either surgically or laparoscopically. Following methods have been employed:

SURGICAL METHODS

Tubectomy. Tubectomy is the permanent method of sterilization in female and is recommended only when the family is completed. In tubectomy operation, fallopian tubes are cut and then cut ends are ligated and buried as shown in Fig. 9.6-2.

Laparoscopic occlusion. In this procedure, the fallopian tubes are occluded using silicon rubber bands, Falope rings or Hulka-Clemens clips. This method is much quicker and simple and hospitalization is not required.

MEDICAL TERMINATION OF PREGNANCY

Medical termination of pregnancy (MTP or abortion) is allowed under MTP Act 1971. Medical pregnancy act has laid down following criteria:

- Conditions in which pregnancy can be terminated the person who can do termination and
- Place, where it should be performed.

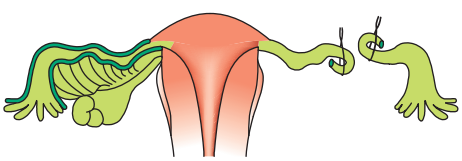


Fig. 9.6-2 Procedure of tubectomy (female sterilization).

Indications

Conditions in which pregnancy can be terminated are:

- *Medical.* When continuation of pregnancy is hazardous to the mother.
- *Eugenic.* When there is substantial risk to the child if born from that pregnancy.
- *Humanitarian grounds.* When pregnancy is the result of rape.
- *Failure of contraceptive measure.*

Methods

Medical termination of pregnancy is possible only in first few months of pregnancy (from 7th week to beginning of second trimester). Following procedures have been employed depending upon the duration of pregnancy:

1. Dilatation and curettage (D and C). In this procedure, cervix is dilated with dilators and implanted ovum is removed by doing curettage of the endometrium.

2. Vacuum aspiration. Like D and C, in this procedure cervix is dilated and then implanted ovum is removed (aspirated) by applying suction. This method is employed only up to 12 weeks of gestation.

3. Administration of prostaglandins. In this method, prostaglandins are administered into the vagina (intravaginally) which causes uterine contractions resulting in expulsion of the products of conception.

PREGNANCY VACCINES

Pregnancy vaccines are under experimental trial. These have not yet been tried in women.

CONTRACEPTIVE METHODS IN MALES

SPACING METHODS

The spacing methods of contraception used in males are:

- Natural method
- Barrier methods
- Chemical methods

NATURAL METHOD OR COITUS INTERRUPTUS

It is the oldest method of voluntary fertility control. In this method, male withdraws the penis before ejaculation into the vagina and tries to prevent deposition of semen into the vagina. This method needs practice and discipline. The failure rate is high because of following reasons:

- Pre-coital secretions of the male may contain sperms and even a drop of semen is sufficient to cause pregnancy.

- Slightest mistake in timings of withdrawal may lead to deposition of certain amount of semen.

BARRIER METHODS

Condom. Condom is the most widely used barrier by the males all around the world. In India, it is known by its trade name *Nirodh* (Fig. 9.6-3). It consists of a fine latex sheath and is electronically tested.

Mechanism of action. Condom prevents deposition of semen into the vagina thus does not allow the sperms and the ovum to meet.

Advantages
- They are easily available, safe and inexpensive.
- Their use does not require any medical supervision.
- They also provide protection against sexually transmitted diseases.

Disadvantages
- It may slip off or tear off during coitus due to its incorrect use.
- It interferes with sexual sensations.

CHEMICAL METHODS

Antispermatogenic drugs

Few drugs which inhibit spermatogenesis have been available. These include:

1. Male pill (Gossypol)

Composition. Male pill contains Gossypol, a phenolic derivative of cottonseed oil.

Mechanism of action. Gossypol acts as an effective azoospermic agent. Its exact mechanism of action is not yet known. In 99.9% of cases sperm count decreases to 4 million/mL.

2. Hormonal preparations

Various hormonal preparations which can be used as contraceptive measures in males are:

Testosterone. Testosterone (400 mg) when given orally produces azoospermia.

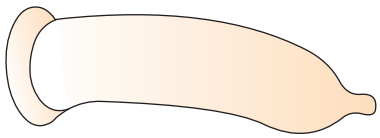


Fig. 9.6-3 Condom.

Testosterone with danazol (17α-ethyl testosterone). This preparation is better tolerated and is more effective.

Cyproterone acetate. Chemically, this drug is related to progesterone. It acts as a potent antiandrogenic agent. It produces oligospermia but also causes loss of libido.

3. Tripterygium wilfordii

This is a special type of wine (prepared from a plant) used in Chinese medicine which reduces the sperm count, but mechanism of action is not yet known.

4. Calcium channel blockers

Calcium channel blockers (e.g. nifedipine) block the Ca^{2+} channels on the cell membrane of the sperms. As a result, the sperm membrane becomes rigid and loaded with cholesterol. The rigid membrane of sperm prevents its binding to the zona pellucida of the ovum.

TERMINAL METHODS

The permanent methods employed for sterilization in males are:

- Vasectomy
- Vas occlusion using no scalpel technique

1. Vasectomy

Vasectomy is a simple operation in which about 1 cm piece of vas deferens is removed after clamping. Then both the ends are

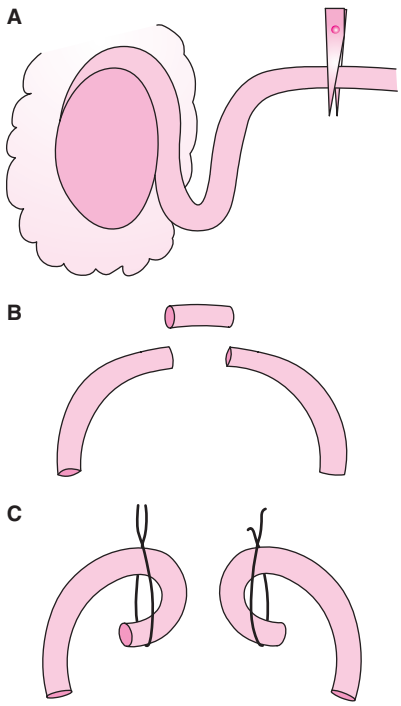


Fig. 9.6-4 Procedure of vasectomy (male sterilization).

ligated and sutured so that they face away from each other (Fig. 9.6-4). This procedure reduces the risk of recanalization later on.

Advantages of vasectomy are that it is simpler, faster, less expensive procedure and no hospitalization is required. It is 100% effective.

Disadvantages. The failure rate of vasectomy is only 0.15% and that too because of wrong identification of vas.

2. No scalpel vas occlusion

No scalpel vas occlusion is a newer technique, which is quite safe, convenient and is acceptable to males.

Principle of occlusion. An elastomer is injected into the vas deferens, it get hardened in situ within 20 min and plug the vas (occlude it).

Advantages. It is an easy procedure and reversal is possible with 100% efficacy.

SECTION 10

Nervous System

SUBSECTION-10A: PHYSIOLOGICAL ANATOMY AND FUNCTIONS OF NERVOUS SYSTEM

SUBSECTION-10B: NEUROPHYSIOLOGY

ORGANIZATION OF NERVOUS SYSTEM

Nervous system, through sophisticated signalling, acts as a control network within the body. The specialised cells that constitute the functional units of the nervous system are called *neurons*. The neurons are responsible for the reception and response to changes in the internal and external environment. It is estimated that the human nervous system is composed of more than 100 billion neurons, which are linked together in a highly intricate manner. Thus, the various parts of the nervous system are interconnected, but for convenience of description the nervous system can be divided anatomically and functionally into different divisions.

ANATOMICAL DIVISIONS OF THE NERVOUS SYSTEM

The nervous system is broadly classified into two anatomical divisions: the central nervous system and the peripheral nervous system.

1. **Central nervous system (CNS)**, which occupies the central axis of the body includes brain and spinal cord. Brain has three parts:

- *Forebrain* comprises telencephalon, i.e. central hemispheres (cerebrum) or anterior part of the forebrain, and diencephalon or posterior part of the forebrain. The upper 2/3rd of diencephalon is called thalamus and lower 1/3rd is called hypothalamus.
- *Mid brain* or mesencephalon and
- *Hind brain* or rhombencephalon comprises pons, medulla oblongata and cerebellum.

2. **Peripheral nervous system (PNS)** is the part of nervous system which lies outside the CNS. The PNS consists of peripheral nerves and the ganglia associated with them. Peripheral nerves attached to the brain are called *cranial nerves* (12 pairs) and those attached with spinal cord are called *spinal nerves* (31 pairs).

FUNCTIONAL DIVISIONS OF THE NERVOUS SYSTEM

Functionally, nervous system can be divided into two parts:

- Somatic nervous system and
- Autonomic nervous system.

Both somatic and autonomic nervous systems have two divisions:

- Sensory division (for collecting information) and
- Motor division (for executing the action).

1. Somatic nervous system

Sensory division of the somatic nervous system collects the information about the changes that take place in the external environment and interprets the meaning of these changes. The sensory division of the somatic nervous system consists of:

- *Sensory receptors* that receive stimulus from the external environment. A stimulus is a change of environment of sufficient intensity to evoke a response in an organism. The stimulus may be mechanical, chemical, thermal, auditory or visual.
- *Afferent neurons* that carry impulses from the receptors to the brain and spinal cord.
- *Parts of the brain* that primarily deal with the processing of information.

Motor division of the somatic nervous system executes appropriate actions with the help of the skeletal muscles in response to changes in the external environment detected by *the sensory* division. It also co-ordinates the actions of different skeletal muscles of the body. Thus skeletal muscles are the *effector organs* of somatic nervous system.

Motor division of the somatic nervous system consists of neurons that carry signals away from the brain and spinal cord to the skeletal muscles. A single motor neuron arising in the CNS traverses directly to the skeletal muscle without the mediation of ganglia. Somatic nervous system is under voluntary control.

2. Autonomic nervous system

The autonomic nervous system (ANS) collects the information about the changes that take place in the internal environment (i.e. internal viscera), interprets these changes and guides the action and gets the plan executed with the help of smooth muscles of viscera, *cardiac muscles* and *secretory epithelium* of glandular *tissues* (which are effector organs of ANS). In other words, the ANS is responsible for the *activities* of the organs of digestion, circulation, excretion, respiration and *reproduction*, as well as of *adrenal medulla, sweat, salivary* and *lacrimal glands*. It also controls the activities of smooth muscles of *iris, ciliary body* and *arrectores pilorum*.

The word autonomous is taken from the Greek words, the '*autos*' meaning self and the '*nomos*' meaning control. Thus ANS is an involuntary system.

Divisions of ANS. The ANS has two main divisions: sympathetic and parasympathetic, each having a central and a peripheral component.

- *Sympathetic division,* also called *thoracolumbar division*, consists of thoracic and lumbar chains of sympathetic ganglia.
- *Parasympathetic division*, also called *craniosacral division*, consists of the ganglia associated with 3rd, 7th, 9th and 10th cranial nerves.

UNDERSTANDING THE NERVOUS SYSTEM

For the purpose of understanding, this section on nervous system has been divided into two subsections:

Subsection-10A: Physiological anatomy and functions of nervous system. This subsection deals with the anatomy and functions of various parts of nervous system.

Subsection-10B: Neurophysiology. Neurophysiology or neurophysiological processes include the study of sensory, motor, autonomic and higher functions of the nervous system. Neurophysiology, though a very complex subject, primarily involves following processes:

Reception of changes in internal and external environment. It is a function of receptors which are of various types.

Transmission of impulses from the receptors to the brain is the function of afferent neurons. Specialized junctions called synapses are there to transmit impulses from one nerve cell to another. This process is known as *synaptic transmission*. The sensations ascend along the *sensory tracts*.

Relay of sensory impulses occurs in the thalamus, which is a large cluster of nuclei that serves as a relay station.

Processing of the sensation occurs in the cerebral cortex. The part of the brain that deals primarily with the processing of *somatic* sensations is called the *somatosensory cortex*. Similarly, the part of the brain, which is involved in processing visual sensation is called *visual cortex*.

Initiation of the response to a sensation occurs from the concerned area of cerebral cortex, For example, the somatic motor commands are initiated from the motor cortex.

Modulation and co-ordination of the response occurs in the subcortical centres. For example, the commands from the motor cortex are co-ordinated and refined by the *basal ganglia loop system* as well as by the cerebellum.

Execution of response. The response to a sensation is ultimately conveyed to the effector organs by the efferent nerves. The *efferent nerves* to the skeletal muscles are called *somatic motor nerves*. The response invoked may be involuntary, i.e. in the form of reflexes or voluntary movements.

Storage of information occurs in the nervous system in the form of memory for future plans.

Emotional and instinctual behaviour is controlled by the hypothalamus and limbic system.

Consciousness and sleep are special functions of brain. These include activity of reticular activating system and other systems.

Higher functions of the nervous system include learning, memory, judgement, language and other functions of the mind.

“This page intentionally left blank"

Chapter 10.1

Physiological Anatomy, Functions and Lesions of Spinal Cord

PHYSIOLOGICAL ANATOMY AND FUNCTIONS OF SPINAL CORD

GROSS ANATOMY

- The spinal cord (Fig. 10.1-1) extends from the upper border of the first cervical vertebra to the lower border of the first lumbar vertebra.
- Its upper end becomes continuous with medulla oblongata and its lower end called *conus medullaris* becomes continuous with a fibrous cord called *filum terminale*.
- The spinal cord is cylindrical in shape and presents two fusiform-shaped enlargements: the *cervical enlargement* for innervation of upper limbs and *lumbar enlargement* for innervation of lower limbs.
- The cord possesses in the midline anteriorly a deep longitudinal fissure, the *anterior median fissure* and on the posterior surface a shallow furrow, the *posterior median sulcus*.
- The spinal cord, like the brain, is surrounded by three meninges: the dura mater, the arachnoid mater and the pia mater.

INTERNAL STRUCTURE

As seen on cross-section (Fig. 10.1-2), the neural tissue of spinal cord presents inner grey matter and outer white matter. Grey matter is constituted by the nerve cell bodies, dendrites and parts of axons, while white matter is formed by the myelinated and unmyelinated nerve fibres.

A. SPINAL GREY MATTER

In transverse section, the grey matter of spinal cord forms an H-shaped mass in the centre of which is present a canal called the spinal canal. The *spinal grey* matter exhibits following parts:

Dorsal horn or posterior grey column refers to the posterior horn-like projection of the H-shaped grey matter in each lateral half of the cord. The dorsal grey column has been subdivided (from anterior to posterior side) into a base, a neck and a head.

Ventral horn or anterior grey column refers to the anterior projection of the grey matter in each lateral half of the cord.

The ventral grey column has been subdivided into an anterior part the head and a posterior part the base.

Lateral horn or intermediate horn or lateral column refers to a small lateral projection between the ventral and dorsal grey columns, present in the thoracic segments and first two lumbar segments only.

Grey commissure is the part of the grey matter, which connects the two (right and left) symmetrical halves of spinal grey matter across the midline. It is traversed by the central canal.

Neurons in spinal grey matter

Neurons in ventral horn

The ventral horn neurons of spinal grey matter are involved in the motor functions and send motor nerve fibres to the muscles and other effector organs. The neurons of ventral horn are arranged in three mediolateral columns:

1. Medial group,
2. Lateral group and
3. Central group (for details see Chapter 10.9, page 820).

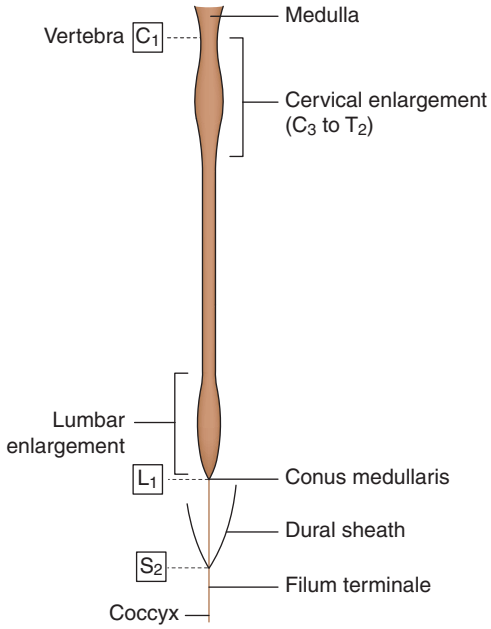


Fig. 10.1-1 Gross appearance of the spinal cord and its relation with vertebrae.

Neurons in dorsal horn

The dorsal horn neurons of spinal grey matter are involved in sensory functions. The dorsal horn neurons are of two types:

1. Internuncial neurons. These are located between the sensory fibres terminating in the dorsal horn and the motor neurons originating in the ventral horn.

2. The tract cells. These cells receive impulses from the various receptors of the body through dorsal nerve root fibres. Axons of these cells enter the white matter of the spinal cord on the same or opposite side and constitute either intersegmental tracts or ascending tracts which terminate in various masses of grey matter in the brain. These tracts form a considerable part of white matter of spinal cord.

The above described neurons of dorsal horn are arranged in four sets of longitudinal neuronal columns. From apex to base of dorsal horn these groups are (Fig. 10.1-3):

1. Substantia gelatinosa of Rolando. It is a column of small cells which caps the apex of dorsal horn as gelatinous material along the entire length of spinal cord. The SG cell has a role in the 'gate control' of pain.

2. Nucleus proprius. It extends along the entire length of the spinal cord and is composed of internuncial cells and tract cells whose axons form the ascending tracts which occupy the anterolateral white funiculi of white matter of spinal cord.

3. Dorsal nucleus. Dorsal nucleus also called the *thoracic nucleus or Clarke's column* extends from C_8 to L_2 segments of spinal cord. It is composed of tract cells which receive

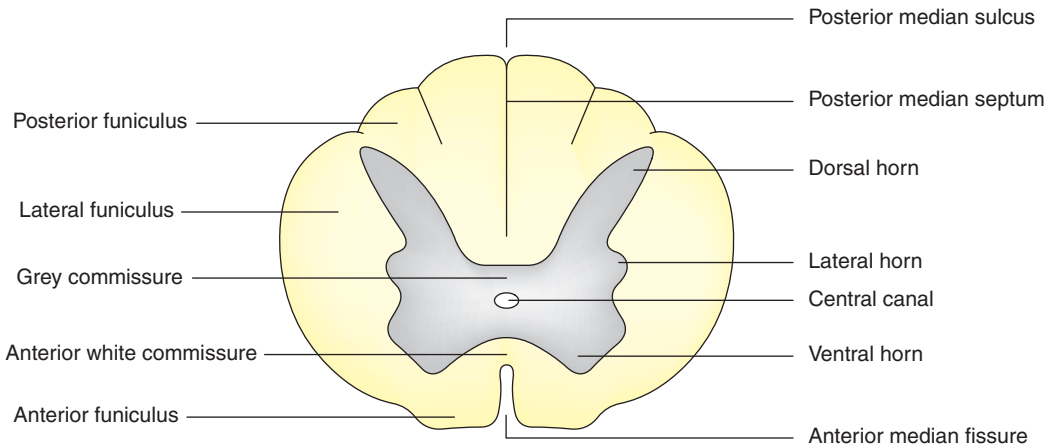


Fig. 10.1-2 Cross-section of thoracic segment of the spinal cord.

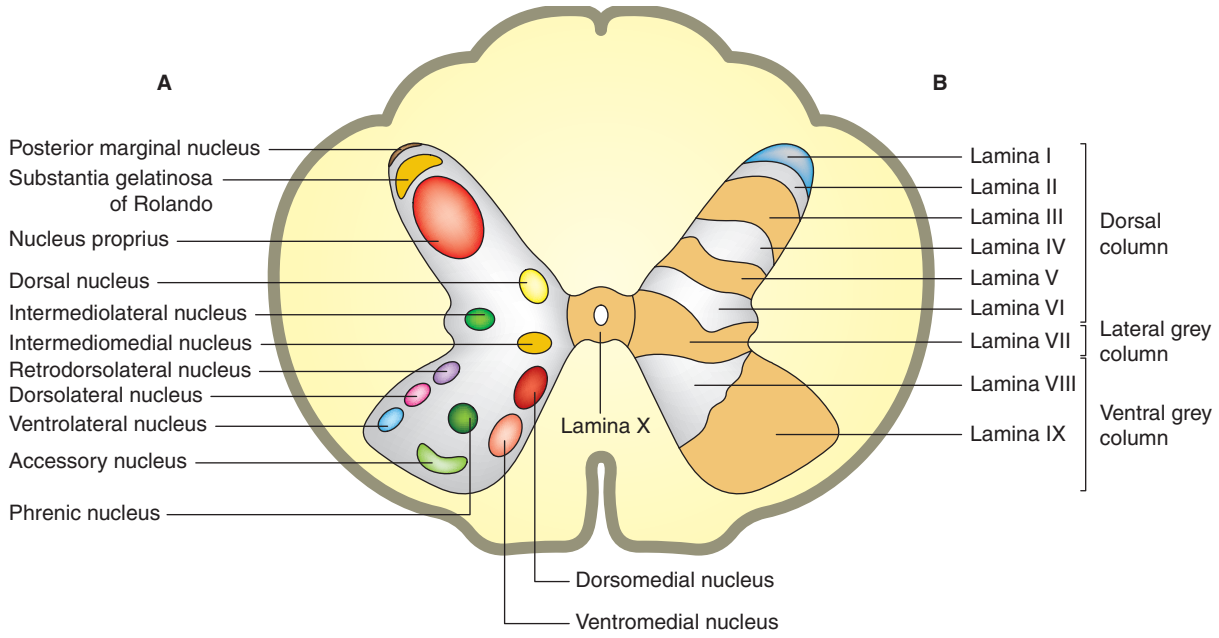


Fig. 10.1-3 Subdivisions of the grey matter of the spinal cord: A, into nuclei and B, into laminae.

proprioceptive, touch and pressure sensations from the trunk and lower limbs. Axons of these cells form the ipsilateral posterior spinocerebellar tract.

4. Posteromarginal nucleus. It is formed by the marginal cells, which cover the substantia gelatinosa at the very tip of the dorsal horn.

Neurons in the lateral horn

The lateral horn cells of the spinal grey matter also called neurons of the intermediolateral group of visceral efferent neurons which extends from T_1 to L_2 segments and from S_2 to S_4 segments of the spinal cord.

Divisions of spinal grey matter into laminae

From the point of view of neuronal connections the spinal grey matter (which has been divided into ventral, lateral and dorsal columns and grey commissure as described above) can be divided into ten laminae (I to X) called Rexed laminae (Fig. 10.1-3).

Laminae I to VI are confined to dorsal grey column.

- *Lamina I* corresponds to posteromarginal nucleus,
- *Lamina II* corresponds to the substantia gelatinosa,
- *Laminae III and IV* correspond to nucleus proprius,
- *Lamina V* corresponds to neck of dorsal grey column and
- *Lamina VI* corresponds to dorsal nucleus in base of the dorsal grey columns.

Note. Afferent fibres carrying cutaneous sensations end predominantly in laminae I to VI. Proprioceptive impulses reach laminae V and VI. These also receive numerous fibres from the cerebral cortex.

Lamina VII is confined to lateral grey column (lateral or intermediate horns). It is composed of autonomic preganglionic neurons.

Laminae VIII and IX are confined to ventral grey horn. Lamina VIII occupies most of the ventral horn in the thoracic region. It is made of interneurons that receive terminals of vestibulospinal and reticulospinal tracts. Its efferent fibres are projected to lamina IX.

Lamina IX. It contains alpha and gamma motor neurons (that give off efferent fibres to skeletal muscles) and several internuncial neurons.

Lamina X. It forms the grey matter around the central canal and consists mostly of neuroglial cells.

B. WHITE MATTER OF SPINAL CORD

White matter is formed by the nerve fibres which are arranged as ascending and descending tracts (described later). In general, the white matter of spinal cord is divided into right and left halves, in front by a deep *anterior median fissure*, and behind by the *posterior median septum* (Fig. 10.1-2). In each half the spinal white matter exhibits following parts:

- *Posterior funiculus or posterior white column* is formed by the white matter present medial to the dorsal grey horn.
- *Anterior funiculus or anterior white column* refers to the white matter present anterior and medial to the ventral grey horn.
- *Lateral funiculus* is formed by the white matter present lateral to the ventral and dorsal grey columns.
- *Anterolateral funiculus* refers to the anterior and lateral funiculi collectively.

- *Ventral (anterior) white commissure refers* to the white matter which is present anterior to anterior grey commissure and joins the right and left halves of white matter.
- *Dorsal (posterior) white commissure* refers to some myelinated fibres running transversely in the grey commissure, posterior to the central canal.

Note. Tracts of spinal cord are described along with the tracts of the brain stem (see page 697).

SPINAL SEGMENTS AND SPINAL NERVES

SPINAL SEGMENTS

Spinal cord, though a continuous structure, can be considered to consist of 31 spinal segments, each giving attachment to rootlets of the ventral and dorsal root, of each spinal nerve. The 31 segments of spinal cord correspond symmetrically to 31 spinal nerves and are named as:

- *8 Cervical segments* give attachment to 8 cervical nerves,
- *12 Thoracic segments* give attachment to 12 thoracic nerves,
- *5 Lumbar segments* give attachment to 5 lumbar nerves,
- *5 Sacral segments* give attachment to 5 sacral nerves and
- *1 Coccygeal segment* gives attachment to 1 coccygeal nerve.

SPINAL NERVES

Each spinal nerve is a mixed nerve formed by the union of two roots: a dorsal (sensory) root and a ventral (motor) root (Fig. 10.1-4).

1. Dorsal nerve root. The dorsal nerve root is formed by several rootlets which are attached to the surface of the spinal cord along a vertical groove called the posterolateral sulcus. All sensory fibres reach the spinal cord through the dorsal nerve roots. Each dorsal nerve root is marked by a swelling called dorsal nerve root ganglion or spinal ganglion. Dorsal root ganglion is composed of T-shaped unipolar neurons with peripheral and central processes.

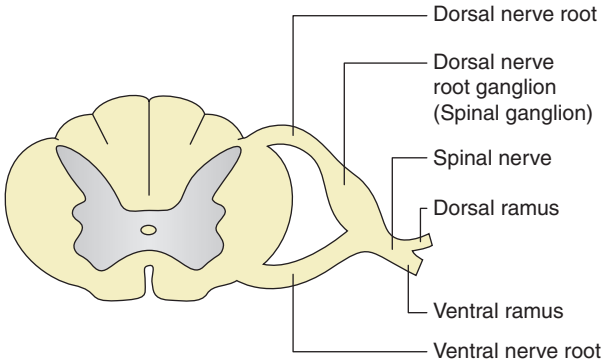


Fig. 10.1-4 Beginning of spinal nerve and its roots.

Peripheral processes of dorsal root ganglion cells extend up to sensory receptors in the skin. The area of the skin supplied by a spinal nerve is called dermatome.

Central processes of dorsal root ganglion cells constitute the dorsal nerve root, which is attached to spinal cord through various rootlets. Each rootlet just before entering the spinal cord divides into medial and lateral divisions.

Medial division of each rootlet consists of myelinated group I and II fibres which include:

- Proprioceptive fibres from muscles and
- Sensory fibres conveying touch, pressure and vibratory sensations.

Lateral division of each rootlet is composed of thinly myelinated group III fibres and unmyelinated group IV fibres:

- Fast and discriminative pain and temperature sensations are conveyed by group III fibres, and
- Slow pain and visceral sensations are conveyed by group IV fibres.

2. Ventral nerve root. Ventral nerve root is formed by various rootlets which are attached to the anterolateral aspect of spinal cord opposite the ventral grey column. The ventral nerve root is composed of axons of motor neurons present in the ventral grey horn. The ventral root also contains the autonomic fibres originating from the lateral (intermediate) horn of the spinal grey matter (lamina VII).

FUNCTIONS OF SPINAL CORD

Spinal cord serves three groups of functions:

- Sensory functions,
- Motor functions and
- Autonomic functions.

1. SENSORY FUNCTIONS

Entry of somatic sensations in spinal cord. All the somatic afferent impulses enter the spinal cord through the dorsal nerve root.

Onward transmission of somatic sensations. After entering the spinal cord, all the somatic sensations are conveyed to the brain (post-central gyrus) by following ascending tracts:

Spinothalamic tracts convey sensations from the opposite side as:

- *Ventral spinothalamic tract* conveys gross (crude) touch and tactile sensations and
- *Lateral spinothalamic tract* conveys pain and temperature.

Dorsal column tract sensations. These occupy the dorsal column of the white matter of cord. These are upward continuation of the fibres of the medial division of the dorsal nerve roots of the same side. These tracts mediate sensations of fine touch, tactile localization and discrimination, pressure, vibration sense, sense of position and sense of movement.

2. MOTOR FUNCTIONS

Spinal cord performs motor functions through the

Pyramidal tracts, which include corticospinal (ventral and lateral) tracts and

Extrapyramidal tracts, which include vestibulospinal tract, tectospinal tract, rubrospinal tract, olivospinal tract and reticulospinal tract.

Motor functions served by spinal cord are:

- Control of tone and power of muscles,
- Control of movement of muscles and joints,
- Control of deep (tendon) reflexes and
- Control of superficial reflexes.

3. AUTONOMIC FUNCTIONS

Visceral afferent impulses in spinal cord travel through dorsal nerve roots to lateral horns of T_1 to L_2 and S_2 to S_4 spinal segments.

Autonomic efferents travelling through the spinal cord supply the visceral organs and control the activity of smooth muscles, heart, glands of gastrointestinal tract (GIT), sweat glands and adrenals. The spinal cord also regulates the body temperature. In other words, spinal cord helps in maintaining the optimal internal environment of the body through its autonomic function.

PHYSIOLOGICAL ANATOMY AND FUNCTIONS OF BRAIN STEM

The brain stem consists (from below upwards) of the medulla oblongata, pons and mid brain (Fig. 10.1-5).

MEDULLA OBLONGATA

GROSS ANATOMY

The medulla oblongata is conical in shape and connects the pons above to the spinal cord below.

Surface of medulla (Fig. 10.1-5) exhibits:

- *Median fissure*, present in the centre of anterior surface of medulla, is continuous below with the anterior median fissure of spinal cord.
- *Pyramids* are two swellings, one each present on the either side of median fissure. These are composed of bundles of nerve fibres that originate in large nerve cells in the precentral gyrus of the cerebral cortex. The pyramids taper below and have most of the descending fibres which cross over to the opposite side, forming the *decussation of pyramids*.
- *Olives* are oval-shaped elevations, present one on each side just posterior to the pyramids. These are produced by the underlying olivary nuclei.
- *Inferior cerebellar peduncles*, which connect the medulla to cerebellum, are present behind the olives.
- *Gracilis and cuneatus tubercles* (produced by the medially placed underlying nucleus gracilis and the laterally placed underlying nucleus cuneatus, respectively) are present on the posterior surface of the inferior part of the medulla oblongata.
- *Cranial nerves*, 9th, 10th, 11th and 12th emerge from the surface of medulla.

INTERNAL STRUCTURE

The main features of the internal structure of the medulla oblongata are most easily reviewed by examining cross-sections at following levels:

- Transverse section at the level of pyramidal decussation (Fig. 10.1-6A).
- Transverse section at the level of sensory decussation (Fig. 10.1-6B) and
- Transverse section at the level of olive (Fig. 10.1-6C).

FUNCTIONS

1. Pathway for ascending and descending tract. The medulla oblongata forms the main pathway for the ascending and descending tracts of spinal cord.

2. House of vital centres. The medulla oblongata houses many important centres, which control the vital functions of the body:

- *Respiratory centres* (inspiratory and expiratory) control the normal rhythmic respiration (page 336).
- *Vasomotor and cardiac centres* control the blood pressure and functions of heart and vascular system (page 250).
- *Deglutition centre* controls the pharyngeal and oesophageal phase of deglutition (page 461).
- *Vomiting centre* is responsible for inducing vomiting in disorders of gastrointestinal tract (see page 477).
- *Superior and inferior salivary nuclei*, located in the medulla, control the salivary secretion (page 459).

3. Cranial nerve nuclei located in the medulla control following functions:

- *Twelfth cranial (hypoglossal) nerve* controls the movements of tongue,

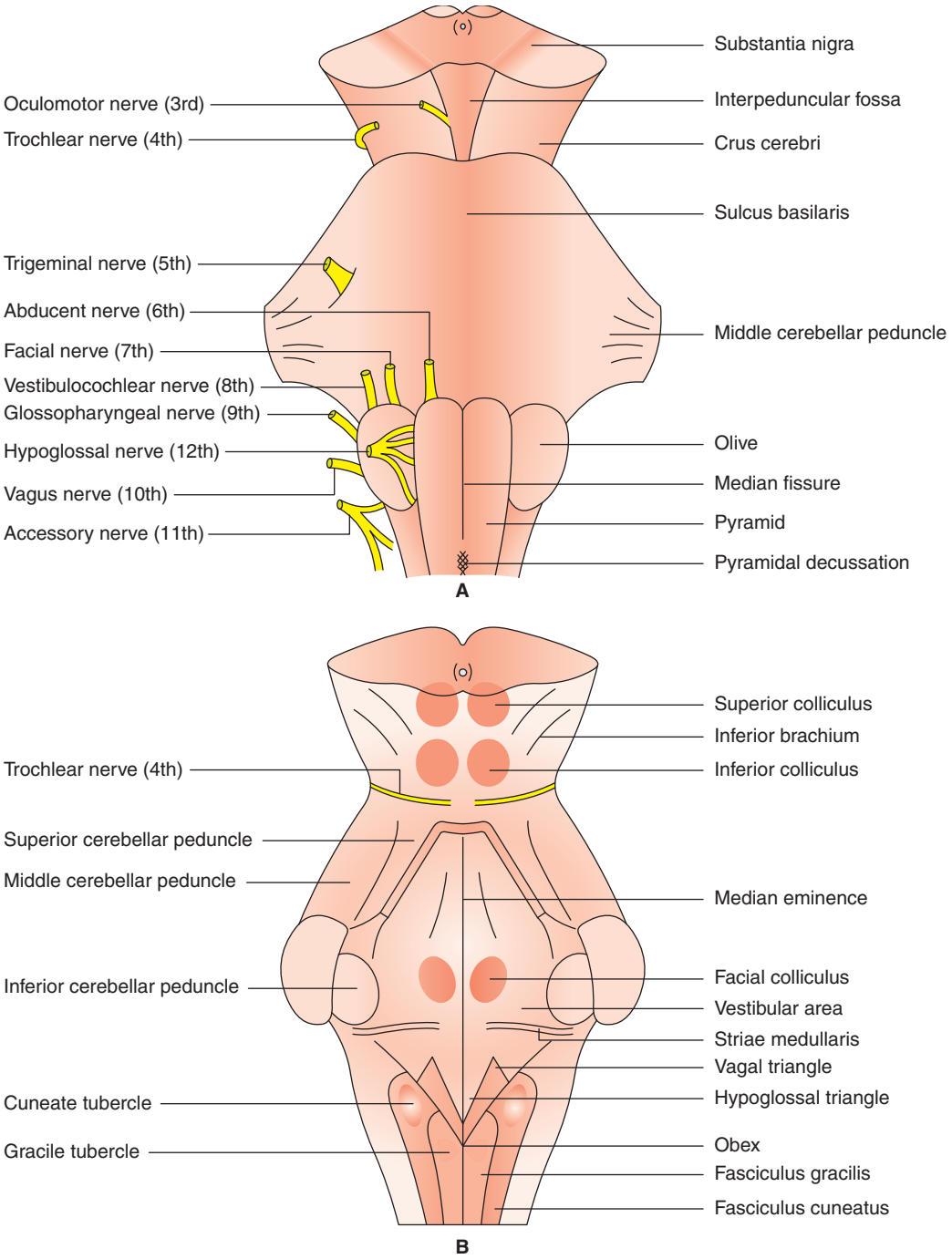


Fig. 10.1-5 Gross anatomy of brain stem: A, ventral aspect and B, dorsal aspect.

- *Eleventh cranial (accessory) nerve* controls the movements of shoulder.
- *Tenth cranial (vagus) nerve* controls the functions of important viscera viz heart, lungs and GIT.
- *Eighth cranial nerve* controls the auditory function (cochlear division of the nerve has the relay in medulla oblongata) and vestibular function (medial and inferior vestibular nuclei extend through much of the medulla).

PONS

GROSS ANATOMY

The pons is situated on the anterior surface of the cerebellum below the mid brain and above the medulla oblongata. Gross features of pons (Fig. 10.1-5) are:

- *Anterior surface* of pons is convex, exhibits prominent transversely running fibres and is marked in the midline

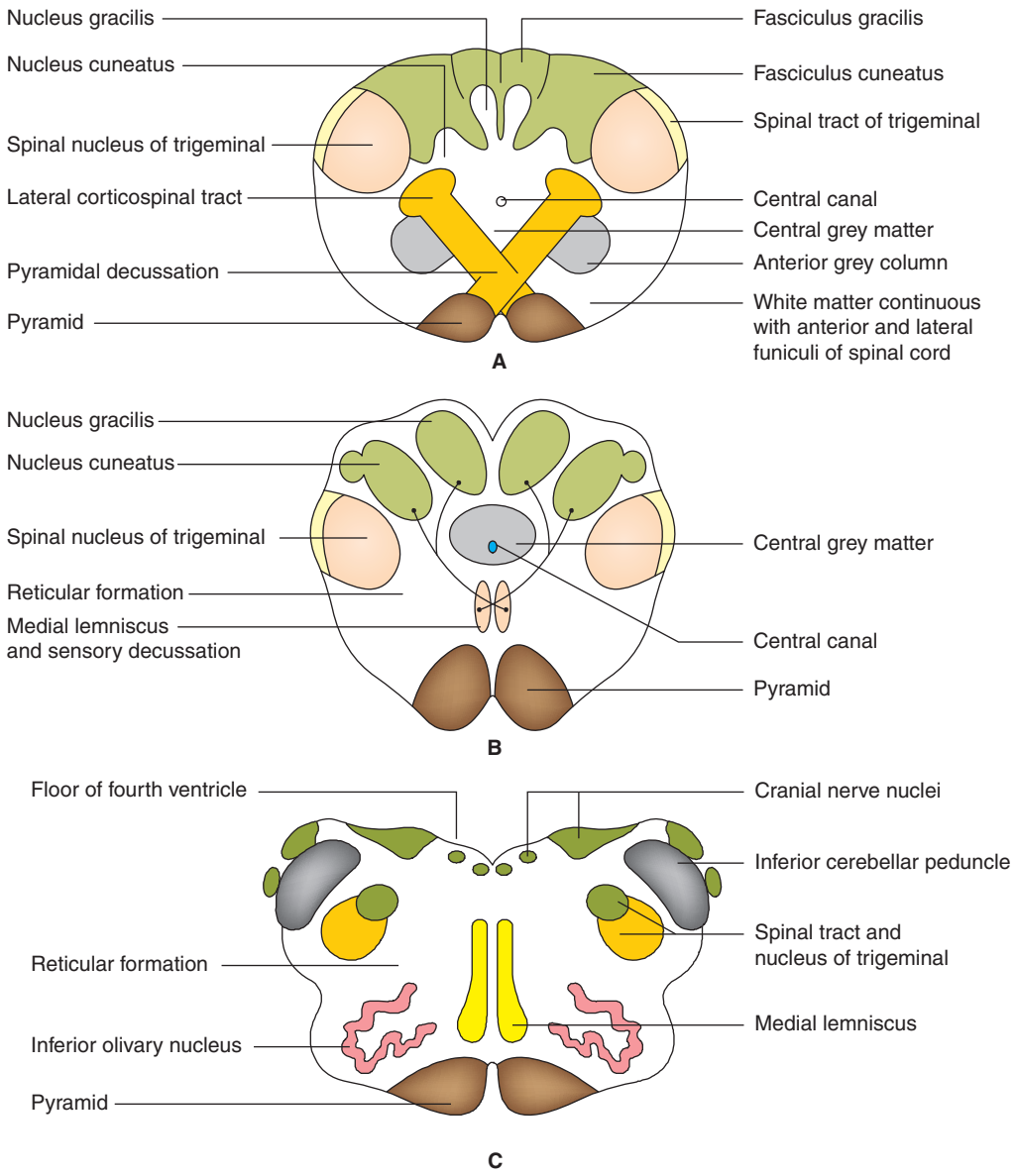


Fig. 10.1-6 Main features of internal structure of medulla oblongata exhibited by transverse sections at the level of: A, pyramidal decussation; B, sensory decussation and C, olive.

by a shallow groove, the *sulcus basilaris*, which lodges the basilar artery.

- *Middle cerebellar peduncles,* present laterally, connect the pons with the cerebellum.
- *Junction between pons and medulla* is marked by a groove through which emerge the 6th, 7th and 8th cranial nerves.
- *Posterior aspect of pons* forms the upper part of the floor of the fourth ventricle.

INTERNAL STRUCTURE

The main features of the internal structure of pons can be best studied by the transverse sections at the level of facial colliculus, i.e. through lower part of pons (Fig. 10.1-7A) and through the upper part of pons (Fig. 10.1-7B). Internally, pons is divisible into two parts:

1. Ventral (basilar) part of the pons contains:

- *Transverse fibres*, which when traced laterally form the middle cerebellar peduncle.
- *Vertical fibres*, present in the pons are of two types:
  - Corticopontine fibres, which end in pontine nuclei and
  - Corticospinal fibres that descend through the pons into the medulla where they form pyramids.
- *Pontine nuclei* are the groups of neurons which are scattered amongst the nerve fibres.

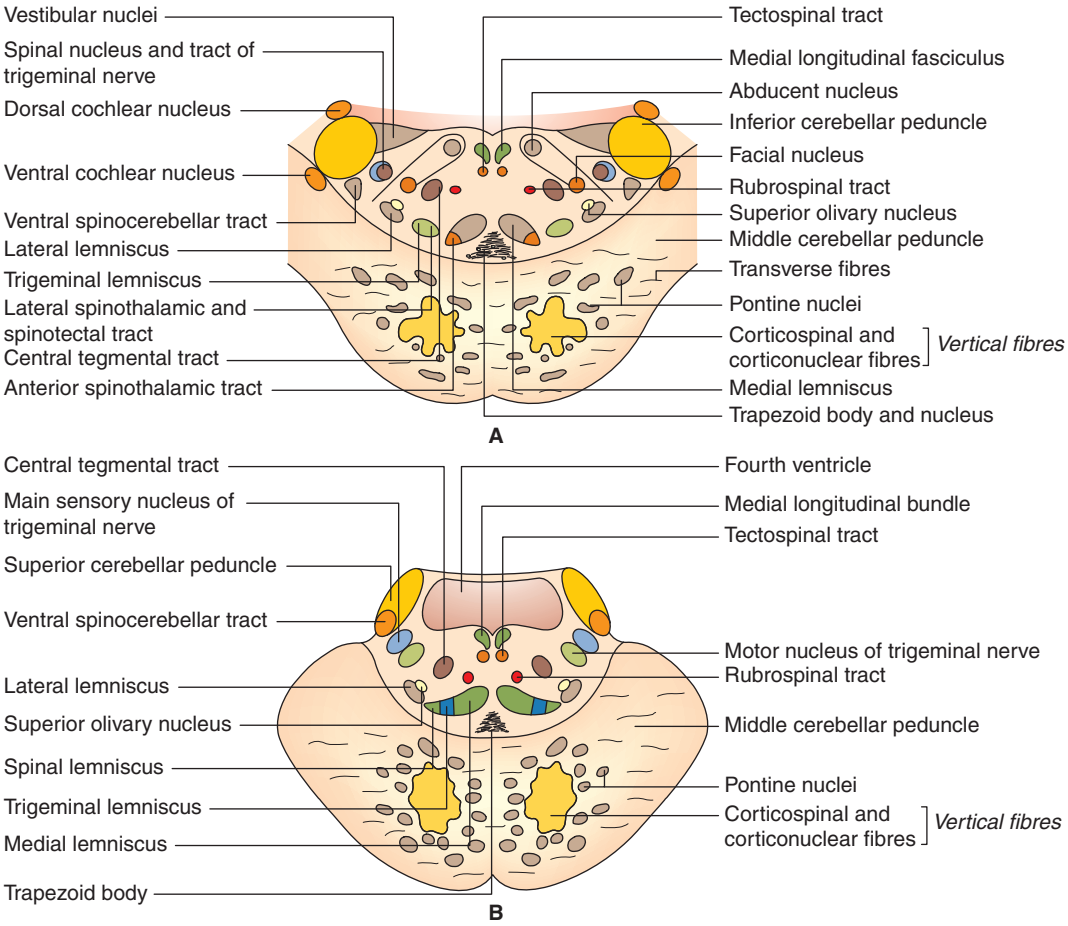


Fig. 10.1-7 Main features of internal structure of pons exhibited by the transverse sections at the level of: A, lower part of pons and B, upper part of pons.

2. Dorsal (tegmental) part of pons contains:
- Decussations of trapezoid body,
- Nuclei of 5th, 6th, 7th and 8th cranial nerves,
- Pontine reticular formation and
- A number of descending and ascending tracts.

The most prominent ascending tracts are the four lemnisci: medial, trigeminal, spinal and lateral.

FUNCTIONS

Pons subserves following functions:

1. *Connecting pathway between cerebral cortex and cerebellum*. The pontine nuclei receive corticopontine fibres and their axons from the middle cerebellar peduncles, which serve as a connecting pathway between the cerebral cortex and the cerebellum.
2. *Pathway for ascending and descending tracts* of spinal cord and medulla oblongata.
3. *Houses the nuclei* of 5th, 6th, 7th and 8th cranial nerves.
4. *Joining station* for medial lemniscus with fibres of 5th, 7th, 9th and 10th cranial nerves.
5. *Contains pneumotaxic and apneustic centres* for regulation of respiration (page 338).

MID BRAIN

GROSS ANATOMY

The mid brain is a narrow part of the brain that connects forebrain to hind brain. Gross anatomical features of mid brain are (Fig. 10.1-5):

Anterior surface of mid brain exhibits (Fig. 10.1-5A):
- *Crura cerebri*. These are two large bundles of fibres, one on each side of the middle line.
- *Interpeduncular fossa* in the triangular space between the two crura.
- *Oculomotor nerve* emerges from the medial aspect of crus of the same side.

Posterior surface of mid brain exhibits (Fig. 10.1-5B):
- *Superior colliculi* are two rounded swellings, one on each side of the midline. Each superior colliculus acts as a subcortical centre for visual reflexes and is connected

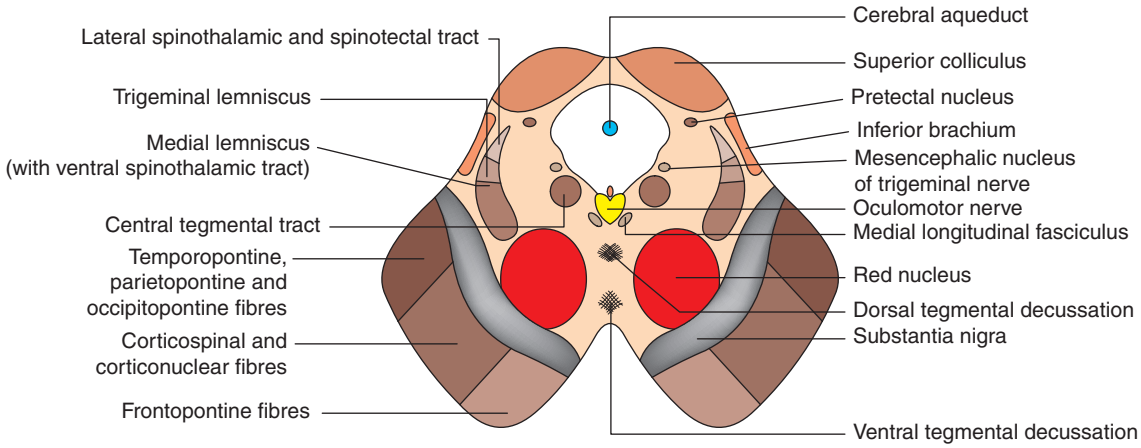


Fig. 10.1-8 Main features of internal structure of mid brain exhibited by transverse section at the level of superior colliculi.

to the lateral geniculate body by a raised band known as a superior brachium.

- *Inferior colliculi* are two rounded swellings one on each side of the midline located below the superior colliculi. Each inferior colliculus acts as a subcortical centre for auditory reflexes and is connected to medial geniculate body by an elevated band known as the *inferior brachium*.

INTERNAL STRUCTURE

The main features of the internal structure of mid brain can be studied by making two transverse sections—one at the rostral level through the superior colliculi and the other at the caudal level through the inferior colliculi.

Internally, for convenience of description the mid brain can be divided into two parts (Fig. 10.1-8):

(i) Tectum. Tectum refers to the part of mid brain lying behind a transverse line drawn through the cerebral aqueduct. It consists of superior and inferior colliculi of two sides.

(ii) Cerebral peduncles. Cerebral peduncles (right and left) constitute the part of mid brain lying in front of the line passing through the cerebral aqueduct. Each cerebral peduncle, in turn consists of three parts, which from anterior to posterior side are:

Crus cerebri (or basis pedunculi). It consists of a large mass of vertically running descending fibres from the cerebral cortex, which include frontopontine fibres (occupying medial one-sixth of crus), corticospinal and corticonuclear fibres (occupying intermediate 2/3rd of crus) and temporopontine, parietopontine and occipitopontine fibres (occupying the lateral one-sixth of crus).

Substantia nigra is a mass of pigmented grey matter (therefore appears dark in colour). Physically, it is considered part of basal ganglia. The connection and functions of substantia nigra are discussed on page 727.

Tegmentum of the two sides is continuous across the midline. It contains following important masses of grey matter and nerve fibres:

- *Red nucleus* is the biggest nucleus present in the upper part of mid brain. Physiologically, the red nucleus is a part of basal ganglia and is involved in regulation of posture (see page 833).
- *Nuclei* of 3rd, 4th and 5th cranial nerves.
- *Reticular formation* of mid brain is continuous below with the reticular formation of pons and medulla.
- *Fibre bundles* of the tegmentum include medial lemniscus, trigeminal lemniscus, lateral lemniscus, tectospinal and rubrospinal tracts.
- *Three decussations* take place in the tegmentum due to: crossing of fibres of superior cerebellar peduncle, rubrospinal tracts (Forel's decussation) and fibres of medial longitudinal bundle (Meynert's decussation).

TRACTS OF SPINAL CORD AND BRAIN STEM

The tracts that transmit sensory impulses to the brain are termed *ascending tracts* and the tracts which are responsible for transmission of motor impulses from the brain to motor neurons reaching muscles and glands are termed *descending tracts*. There are numerous ascending and descending tracts in the spinal cord and brain stem.

ASCENDING TRACTS

Ascending tracts convey impulses arising in various parts of the body to different parts of the brain. The ascending

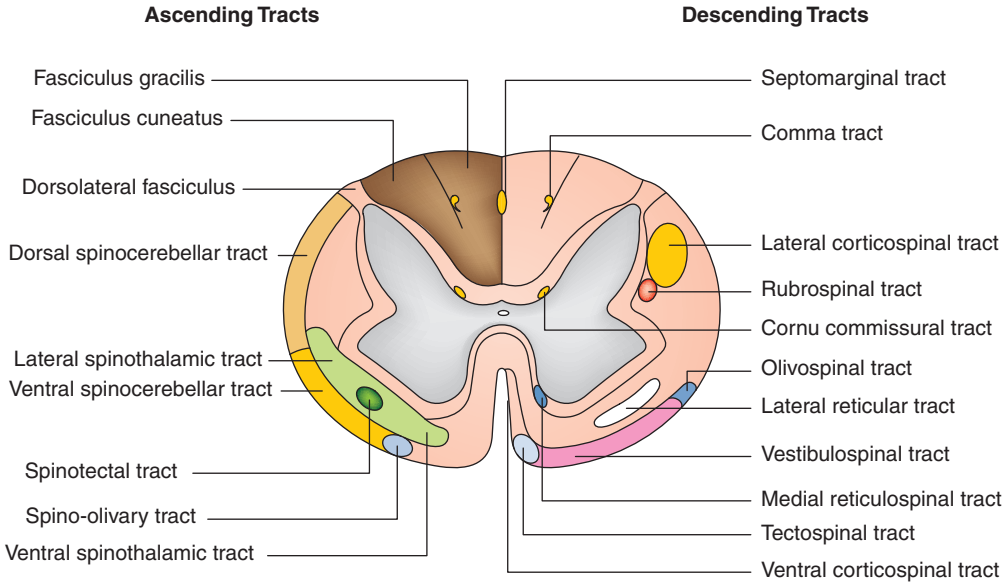


Fig. 10.1-9 Schematic diagram to show the position of the main ascending and descending tracts of the spinal cord and brain stem.

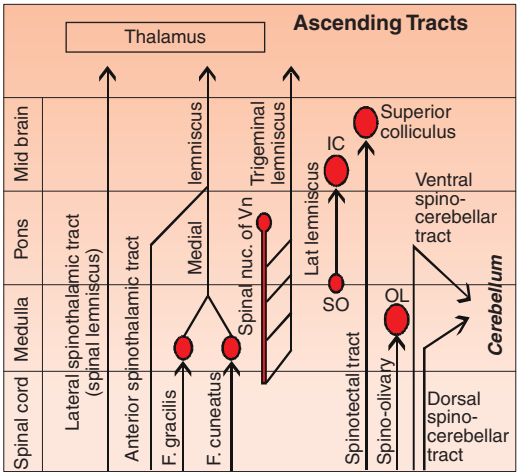


Fig. 10.1-10 Schematic diagram to show the various ascending tracts of spinal cord and brain stem. (SC = Superior colliculus; SO = superior olivary nucleus; Vn = vestibular nucleus; OL = olivary nucleus; IC = inferior colliculus.)

tracts present in the spinal cord (Figs 10.1-9 and 10.1-10) can be grouped as:

- Ascending tracts connecting spinal cord with cerebral cortex,
- Ascending tracts ending in the brain stem and
- Spinocerebellar pathways.

I. ASCENDING TRACTS CONNECTING SPINAL CORD WITH CEREBRAL CORTEX

1. Posterior column-medial lemniscus pathway

Fasciculus gracilis and fasciculus cuneatus

Location. Fasciculus gracilis and fasciculus cuneatus occupy the posterior white funiculus of the spinal cord, and are, therefore, often referred to as the *posterior column tracts*. Fasciculus gracilis is situated medial to fasciculus cuneatus. (Fig. 10.1-9).

Origin. These tracts are unique in that they are formed predominantly by the axons of the first-order sensory neurons located in the dorsal root ganglia (Fig. 10.1-11).

- Recently, it has been shown that these tracts also contain some fibres that originate in the dorsal grey column (laminae III and IV), i.e. sensory neurons of second order.

Arrangement of fibres. The fibres derived from the lowest ganglia are situated most medially; while those from the highest ganglia are most lateral. Therefore:

- *Fasciculus gracilis (tract of Gall)*, which lies medially, is composed of fibres from the coccygeal, sacral, lumbar and lower thoracic ganglia and
- *Fasciculus cuneatus (tract of Burdach)*, which lies laterally, consists of fibres from upper thoracic and cervical ganglia.

Course. After entering the spinal cord, the fibres ascend through the posterior white funiculus and reach the medulla (Fig. 10.1-11).

Termination. After reaching the medulla, the fibres of gracilis and cuneatus fasciculi terminate by synapsing with neurons in the nucleus gracilis and nucleus cuneatus, respectively (Fig. 10.1-11).

Medial lemniscus

The neurons of nucleus gracilis and nucleus cuneatus form the *second-order sensory neurons*. Their axons form the *internal arcuate fibres,* which run forward and medially to cross the midline. The crossing fibres of the two sides

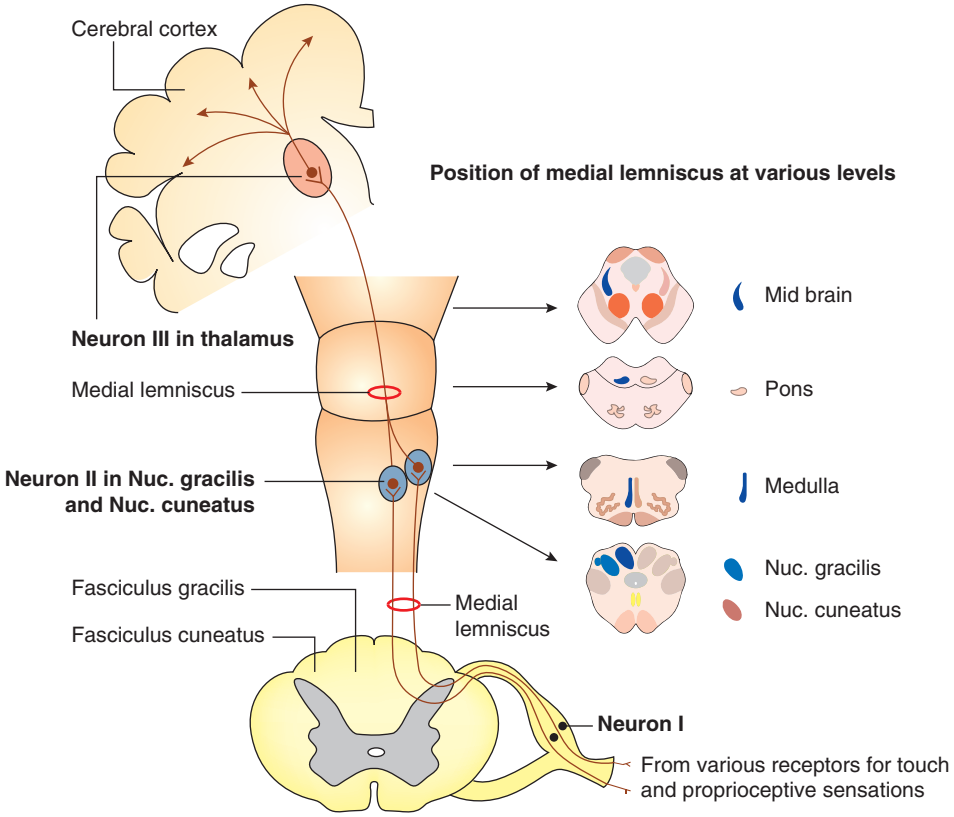


Fig. 10.1-11 Course of posterior column-medial lemniscus pathway. Note the sensory decussation and position of medial lemniscus at various levels of brain stem.

constitute the *sensory decussation* and then ascend through the medulla, pons and mid brain as *medial lemniscus* (Fig. 10.1-11). The fibres of medial lemniscus terminate in a ventral posterolateral nucleus of thalamus.

Third-order sensory neurons

The fibres of the medial lemniscus synapse with the third-order sensory neurons located in the thalamus. Axons of the third-order neurons pass through the internal capsule and corona radiata to reach the somatosensory areas of the cerebral cortex.

Functions of posterior column-medial lemniscus pathway. These fibres carry sensations of some components of touch, vibration and proprioception to the cortex and thus help in following functions:

1. Components of sense of touch include:

- Deep touch and pressure,
- Fine touch, i.e. epicritic tactile sensations,
- Tactile localization, i.e. ability to localize, exactly the part of skin touched,
- Tactile discrimination, i.e. the ability to recognize as separate two points on the skin that are touched simultaneously.
- Stereognosis, i.e. the ability to recognize the shape of known objects by touch with closed eyes.

2. Proprioceptive impulses help in conscious kinaesthetic sensations, i.e. the sense of position of different parts of the body under static conditions as well as rate of change of movement of different parts during body movements.

3. Sense of vibrations, i.e. ability to detect rapidly changing peripheral conditions. This is the ability to perceive the vibrations conducted to deep tissues through the skin.

2. Spinothalamic pathways

Anterior and lateral spinothalamic tracts (Fig. 10.3-12)

Location. The anterior spinothalamic tract is located in the anterior white funiculus near the periphery, while lateral spinothalamic tract is located in the lateral funiculus towards medial side, i.e. near the grey matter (Fig. 10.1-9).

Origin. The spinothalamic tracts are formed by the axons of the chief sensory cells of posterior grey horn, which form the second-order sensory neurons. The first-order neurons of this pathway are located in the spinal ganglia. These neurons receive the impulses from the cutaneous receptors. Central processes of these neurons enter the spinal cord and terminate in relation to the chief sensory cells of spinal grey matter.

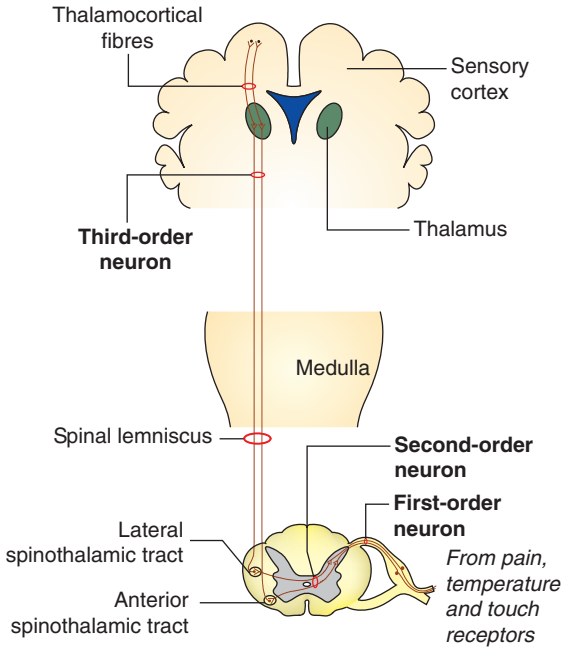


Fig. 10.1-12 Course of anterior and lateral spinothalamic tracts.

Course. After taking origin from the chief sensory cells, the fibres of anterior spinothalamic tract ascend in posterior grey horn for 2–3 segments in the same side. Then, they cross obliquely to the opposite side of the spinal cord in the white commissure (but some fibres may remain uncrossed).

The fibres of lateral spinothalamic tract cross within the same segment of spinal cord and reach the lateral column of the same segment.

The two tracts also carry about 10% uncrossed fibres and run up in the spinal cord, medulla, pons and mid brain and reach the thalamus. In the brain stem, they form the so-called *spinal lemniscus*.

Termination. All the spinothalamic fibres running in the spinal lemniscus terminate in the ventral posterolateral nucleus of thalamus. The neurons of this thalamic nucleus form the *third-order neurons* of this sensory pathway and relay the impulses to the somaesthetic area of the cerebral cortex.

Functions. Traditionally, it has been said that the anterior spinothalamic tracts carry sensations for *crude touch and pressure*, while the lateral tracts carry sensations of *pain and temperature*.

Dorsolateral spinothalamic tract

The dorsolateral spinothalamic tract also called *fasciculus dorsolateralis* or *tract of Lissauer*.

Origin. This is formed by the fibres arising from the neurons of the posterior root ganglia, which form the *first-order neurons*.

Location. It is located in the lateral white column between the periphery of spinal cord and tip of posterior grey horn (Fig. 10.1-9).

Course. These fibres enter the spinal cord through the lateral division of the posterior nerve root and pass upwards or downwards for few segments on the same side and synapse with cells of substantia gelatinosa of Ronaldo situated in the posterior grey column. The processes of these cells (*second-order neurons*) cross to the opposite side and ascend in the dorsolateral fasciculus to reach the ventral posterolateral nucleus of thalamus where they synapse.

Functions. This tract carries impulses arising in skin (mainly pain and temperature).

APPLIED ASPECT

Relief of pain after dorsolateral cordotomy may be a result of the cutting of these fibres.

Spino-cervico-thalamic pathway

Origin. This tract is formed by the axons arising from neurons located in laminae III to V of spinal grey matter.

Course and termination. After origin, the fibres ascend through the dorsolateral fasciculus and end in the *lateral cervical nucleus* (which is small collection of neurons lying amongst the fibres of the lateral funiculus in spinal segments C_1 and C_2. New fibres arising here project to the ventral posterolateral nucleus of the thalamus.

Functions. It is another pathway through which cutaneous sensations (touch, pressure, pain and temperature) reach the thalamus.

II. ASCENDING TRACTS ENDING IN BRAIN STEM

The ascending tracts arising in the spinal grey matter and ending in masses of grey matter in different parts of the brain stem are (Fig. 10.1-10) as follows:

1. Spinoreticular tract

Location. This tract is located in the anterolateral white funiculus (Fig. 10.1-9).

Origin. The spinoreticular fibres begin from the spinal neurons mainly in lamina VII (also V and VIII).

Course. The fibres are partly crossed and partly uncrossed and ascend in the ventrolateral part of the spinal cord, intermingling with the spinothalamic tracts.

Termination. These fibres end in reticular formation of medulla and pons:

- In the medulla, the fibres end in nucleus reticularis gigantocellularis and lateral reticular nucleus of same side; some fibres terminate in the opposite side.
- In the pons, these fibres terminate in the nucleus reticularis pontis caudalis of the same side or opposite side.
- Very few fibres terminate in the mid brain.

Functions. The fibres of the spinoreticular tract are the components of ascending reticular activating system and are concerned with arousing consciousness or alertness.

2. Spinotectal tract

Location. This tract is located in the lateral side of lateral white funiculus anterior to the lateral spinothalamic tract. It is bounded anteriorly by an anterior nerve root (Fig. 10.1-9).

Origin. Fibres of this tract arise from the chief sensory cells of the posterior grey column. First appearance of the fibres is in upper lumbar segments. This tract is very prominent.

Course and termination. After origin from the spinal grey matter, the fibres cross to the opposite side through anterior white commissure to the lateral funiculus. Then the fibres ascend up to the mid brain along with anterior spinothalamic tract and end in the superior colliculus and mid brain reticular nuclei.

Functions. These fibres form alternate route for conduction of *slow pain* and are also concerned with *spinovisual reflexes*.

3. Spino-olivary tract

Location. This tract is located in the anterolateral part of white funiculus and occupies mostly the anterior white funiculus (Fig. 10.1-9).

Origin, course and termination. The origin of the fibres of this tract is not specific. It is also a crossed tract. Its fibres terminate into olivary nucleus of medulla oblongata, from where the neurons project into the cerebellum.

Function. This tract is concerned with proprioception.

III. SPINOCEREBELLAR TRACTS

The spinocerebellar tracts carry proprioceptive impulses arising in the lower part of the body to the cerebellum. Recent investigations have shown that some exteroceptive sensations (e.g. touch) may reach the cerebellum through these pathways. Thus, the spinocerebellar tracts are constituted by the fibres of second-order neurons of the pathway for subconscious kinaesthetic sensation. The spinocerebellar pathway is organized into the ventral and dorsal tracts.

1. Ventral spinocerebellar tract

Location. It is located in the lateral white funiculus of the spinal cord along the lateral periphery (Fig. 10.1-9).

Origin. The ventral (anterior) spinocerebellar tract also known as Gower's tract is constituted by the second-order neurons (of proprioceptive pathway) located in the junctional area between the ventral and dorsal grey column (laminae V, VI, VII) in the lumbar and sacral segments of the cord. These neurons receive impulses from the first-order neurons located in the posterior root ganglia. The peripheral processes of first-order neurons receive impulses from muscle spindles, Golgi tendon organs and other proprioceptive receptors. Some fibres are related to end organs concerned with exteroceptive sensations (touch, pressure).

Course. After origin from the junctional (marginal) cells, the majority of fibres of ventral spinocerebellar tract cross to the opposite side and ascend in the lateral funiculus, anterior to the fibres of the dorsal spinocerebellar tract (Fig. 10.1-9) (some fibres ascend in lateral funiculus of the same side also). Then these fibres ascend through spinal cord, medulla, pons and mid brain. Finally, these fibres reach the cerebellum through the superior cerebellar peduncle.

Termination. These fibres terminate in the lower limb area of the cerebellar cortex.

2. Dorsal (posterior) spinocerebellar tract

Location. This tract is located in the lateral funiculus along the posterolateral periphery of spinal cord. It is situated posterior to the ventral spinocerebellar tract and anterior to the entry of posterior nerve root (Fig. 10.1-9).

Origin. The first-order neurons are located in the posterior nerve root ganglia. The peripheral processes of *first-order neurons* receive impulses from the muscle spindles, Golgi tendon organs and other proprioceptive receptors.

Course. Unlike ventral spinocerebellar tract, the dorsal spinocerebellar tract is *uncrossed*. The fibres of this tract after origins reach the lateral funiculus of same side and ascends through other spinal segments and reach the medulla oblongata. From here the fibres reach the cerebellum through the inferior cerebellar peduncle.

Termination. Most of the fibres of this tract terminate in the cortex of anterior lobe of cerebellum.

3. Cuneocerebellar tract

The central processes of some first-order neurons (related to cervical segments) reach the *accessory cuneate nucleus* in the medulla. The central processes of the second-order neurons located in the accessory cuneate nucleus form the cuneocerebellar tract (posterior external arcuate fibres),

Table 10.1-1 Major ascending tracts in the spinal cord

Tract	Location	Origin*	Termination	Functions
Fasciculus gracilis and fasciculus cuneatus (tracts of Gall and Burdach)	Posterior white column of spinal cord.	Dorsal root ganglia of spinal nerves of the same side.	Nucleus gracilis and nucleus cuneatus in medulla of the same side.	Joint sense, vibration sense, two point discrimination, stereognosis, conscious kinaesthesia.
Spinothalamic tracts • Lateral spinothalamic tract • Anterior spinothalamic tract	Lateral white column. Anterior white column.	Posterior horn cells of spinal cord of opposite side. Posterior horn cells of spinal cord of opposite side.	Ventral posterolateral (VPL) nucleus of thalamus. Ventral posterolateral (VPL) nucleus of thalamus.	Carry pain and temperature from opposite side of the body. Carry light touch, pressure, tickle and itch sensation from opposite side of the body.
Spinotectal tract	Lateral white column.	Posterior horn cells of spinal cord of opposite side.	Superior colliculus of tectum of mid brain Cerebellum.	Visuomotor reflexes viz head and eye movements towards the source of stimulation.
Spinocerebellar (anterior and posterior) tracts	Lateral white column (superficially).	Posterior horn cells of spinal cord of same side.		Unconscious kinaesthesia (proprioception).

*Location of cell bodies of neurons from which the axons of tract arise.

which enter the inferior cerebellar peduncle of same side to reach the cerebellum.

Functions. This tract brings the *conscious proprioception* impulses from the upper limb. Thus, it may be regarded as the forelimb equivalent of the dorsal spinocerebellar tract.

4. Rostral spinocerebellar tract

Origin, course and termination. This tract is believed to arise from the spinal grey matter in lower four cervical segments (lamina VII) from the neurons which constitute the *nucleus centrobasalis*.

Most of the fibres of this tract are uncrossed. They reach the cerebellum through the inferior and superior cerebellar peduncles.

Functions. This pathway is regarded, functionally, as the forelimb equivalent of the ventral spinocerebellar tract.

The major ascending tracts in the spinal cord are summarized in Table 10.1-1.

DESCENDING TRACTS

The descending tracts concerned with the various motor activities of the body, and formed by the motor nerve fibres arising from the brain and descending into the spinal cord and brain stem (Fig. 10.1-13).

DESCENDING TRACTS ENDING IN SPINAL CORD

Traditionally, the descending tracts ending in the spinal cord have been divided into two groups:

- Pyramidal tracts and
- Extrapyramidal tracts.

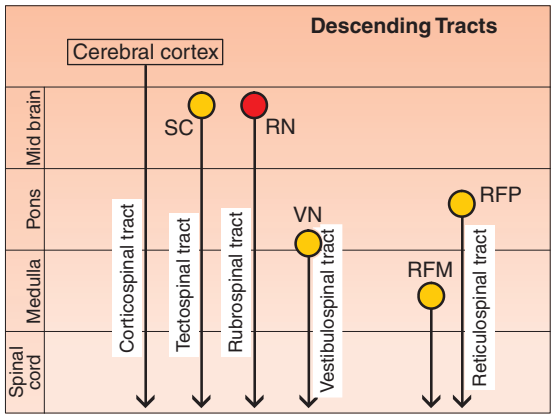


Fig. 10.1-13 Schematic drawing to show the various descending tracts ending in the spinal cord and brain stem. (SC = Superior colliculus; RN = red nucleus; VN = vestibular nucleus, RFP = reticular formation of pons; RFM = reticular formation of medulla.)

I. Pyramidal tracts

The pyramidal tracts refer to the corticospinal tracts, which are constituted by the axons that transmit motor signals directly from the cortex to the spinal cord (Fig. 10.1-14).

Origin. Corticospinal tract fibres originate from the following nerve cells in the cerebral cortex:

- Primary motor cortex (area 4) –30%,
- Premotor area (area 8) and supplementary motor area –30% and
- Somatic sensory areas (areas 3, 1, 2) –40%.

All the above fibres form the fibres of upper motor neurons of the motor pathway.

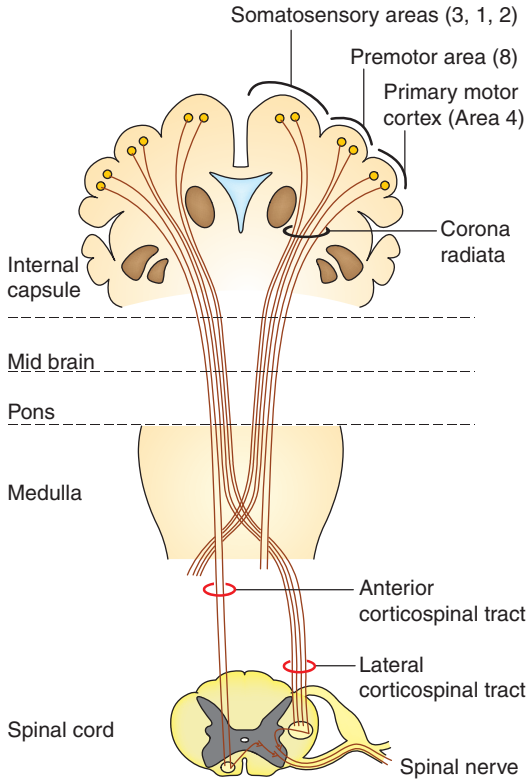


Fig. 10.1-14 Pathway of corticospinal tracts.

Course and termination. After originating from the cerebral cortex, the corticospinal tract fibres descend as a part of corona radiata and then pass through the posterior limb of the internal capsule and then downwards through the brain stem forming pyramids in the medulla (hence the name pyramidal tracts).

In the lower part of medulla about 90% fibres of each pyramid decussate in the mid line to reach opposite side. From here downward the fibres of corticospinal tracts are divided into two separate tracts:

1. Lateral corticospinal tract is constituted by 80% of fibres which have crossed to opposite side. The lateral corticospinal tract fibres descend the full length of spinal cord through the posterior part of lateral white funiculus (Figs 10.1-9 and 10.1.14). Most of these fibres terminate in the internuncial neurons of the spinal grey matter. The internuncial neurons carry the impulses to the motor neurons situated in the ventral grey horn. Some fibres of the tract terminate directly on the ventral horn cells. The axons of the ventral motor neurons supply the skeletal muscles directly by passing through the ventral nerve root. The neurons giving origin to the fibres of pyramidal tract along with their axons constitute the *upper motor neurons*. The ventral motor neurons in the spinal cord along with their axons constitute the *lower motor* neurons.

2. Anterior corticospinal tract is formed by 20% uncrossed pyramidal fibres. These fibres descend down through the anterior white funiculus of the same side. The anterior corticospinal tract fibres do not reach further than the mid-thoracic region. On reaching the appropriate level of the spinal cord, the fibres of this tract cross the midline (through the anterior white commissure) to reach grey matter on the opposite side of the cord and terminate in a manner similar to that of the fibres of the lateral corticospinal tract. Thus, the corticospinal fibres of both the lateral as well as the anterior tracts ultimately connect the cerebral cortex of one side with ventral horn cells in opposite half of the spinal cord.

Salient features of nerve fibres of corticospinal tracts

- Fibres of the corticospinal tract are unmyelinated at birth. Myelination begins in the second post-natal week and is completed by 2 years.
- The large fibres of pyramidal tracts have the tendency to disappear at old age causing automatic shaking movements of old age.

Functions

- The cerebral cortex controls voluntary fine skilled movements of the body through the corticospinal tracts. Interruption of the tract anywhere in its course leads to paralysis of the muscles concerned.

  Note. As the fibres are closely packed in their course through the internal capsule and brain stem, small lesions here can cause widespread paralysis.

- The pyramidal tract fibres also send collaterals to other areas of the motor control systems thus communicating motor command to the basal ganglia, cerebellum and the brain stem.
- In their course through the brain stem, some of the fibres (corticonuclear fibres) terminate directly on the motor nuclei of cranial neurons controlling facial muscles. Since these fibres perform the same function as pyramidal tracts, they are also considered part of the pyramidal system.

II. Extrapyramidal tracts

The descending tracts of spinal cord other than the pyramidal tracts are collectively called extrapyramidal tracts. These include:

- Rubrospinal tract,
- Vestibulospinal tract,
- Reticulospinal tract,
- Tectospinal tract,
- Olivospinal tract and
- Medial longitudinal fasciculus.

1. Rubrospinal tract

Origin. This tract arises from the large cells (nucleus magnocellularis) or red nucleus in the mid brain.

Course. After arising from the red nucleus, the fibres of this tract cross to opposite side in the lower part of the segmental of mid brain (ventral segmental decussation). Then, the tract descends through the pons and medulla and follows a course similar to that of lateral corticospinal tract in the lateral funiculus of the spinal cord (Figs 10.1-9 and 10.9-3).

Termination. The fibres terminate mainly on interneurons along with the corticospinal fibres.

Functions. This tract exhibits facilitatory influence on the flexor muscles and inhibitory influence on the extensor muscles of the body.

- The red nucleus also receives the corticorubral fibres from the ipsilateral motor cortex. The corticorubro-spinal tract thus formed may act as an alternate route of pyramidal system to exert influence on the lower motor neurons.
- The rubrospinal tract is most important and much better developed in some animals than in human. In human beings, the red nucleus is relatively small and the rubrospinal tract reaches only the upper three cervical segments of the spinal cord.

2. *Vestibulospinal tracts*

There are two vestibulospinal tracts: lateral and medial.

(i) *Lateral vestibulospinal tract*

Origin. Fibres of this tract arise from the lateral vestibular (Deiters') nucleus. These fibres are somatotopically arranged. Fibres to cervical segments arise from the cranioventral part, those to thoracic segments from the central part and those to lumbosacral segments from the dorsocaudal part of lateral vestibular nucleus.

Location and course. This tract is uncrossed and lies in the anterior funiculus of the spinal cord (Fig. 10.1-9); shifting medially as it descends.

Termination. The fibres extend up to caudal segments of the cord and terminate into the neurons of ventral grey column (laminae VII and VIII). Through the interneurons these are projected to the alpha and gamma neurons of lamina IX; some fibres directly reach the alpha neurons.

Functions. Vestibular nucleus receives afferents from the vestibular apparatus mainly from utricles. This pathway is principally concerned with adjustment of postural muscles to linear acceleratory displacements of the body. Lateral vestibulospinal tract mainly facilitates activity of extensor muscles and inhibits the activity of flexor muscles in association with the maintenance of balance.

(ii) *Medial vestibulospinal tract*

Origin. The fibres of this tract arise from the medial vestibular nucleus.

Location and course. This tract descends through the anterior funiculus (within the sulcomarginal fasciculus). The fibres are mostly uncrossed but some fibres are crossed.

Termination. The fibres end in the anterior motor neurons directly or through internuncial neurons (laminae VII and VIII) of the cervical segments of spinal cord.

Functions. This part of the vestibular nucleus receives signals from the vestibular apparatus mainly from the semicircular canals. Functionally, medial vestibulospinal tract is the donor connection of medial longitudinal fasciculus. This tract provides a reflex pathway for movements of head, neck and eyes in response to the visual and auditory stimuli.

3. *Reticulospinal tracts*

There are two reticulospinal tracts: the medial (pontine) reticulospinal tract and lateral (medullary) reticulospinal tract.

(i) *Medial (pontine) reticulospinal tract*

Origin. It arises in the medial pontine reticular formation.

Course. The tract descends, mostly uncrossed, in the anterior funiculus of spinal cord.

Termination. The fibres terminate in the laminae VII and VIII of spinal grey matter and through internuncial neurons influence alpha and gamma neurons of lamina IX.

(ii) *Lateral (medullary) reticulospinal tract*

Origin. The fibres of this tract originate from the gigantocellular component of medullary reticular formation.

Course. These fibres are mostly uncrossed and a few crossed. This tract descends in the lateral funiculus medial to the lateral corticospinal and rubrospinal tracts (Fig. 10.1-9).

Termination. The fibres terminate in the internuncial neurons of laminae VII, VIII and IX of the spinal cord.

Functions of reticulospinal tracts. The reticular formation of the brain stem receives input mostly from the motor cortex through the corticoreticular fibres which accompany the corticospinal tracts. Thus the corticoreticulospinal tracts form additional polysynaptic pathways from the motor cortex to the spinal cord. These tracts are concerned with control of movements and maintenance of muscle tone. The reticulospinal tracts, probably, also convey autonomic information from higher centres to the intermediate region of spinal grey matter and regulate respiration, circulation and sweating.

The pontine and medullary reticular nuclei mostly function antagonistic to each other.

4. Tectospinal tract

Origin. Fibres of this tract arise from the superior colliculi.

Course. The fibres cross the midline in the lower part of segmental of the mid brain forming dorsal segmental decussation. Then the tract descends through the pons and medulla into the anterior white funiculus of the spinal cord (Fig. 10.1-9).

Termination. The fibres terminate in upper cervical levels by synapsing on the anterior horn cells through internuncial neurons located in laminae V and VII of the spinal grey matter.

Function. This tract forms the motor limb of the reflex pathway for turning the head and moving the arms in response to visual, hearing or other exteroceptive stimuli.

5. Olivospinal tract

Origin. This tract originates from the inferior olivary nucleus.

Course and termination. The tract fibres descend and terminate ipsilaterally in the anterior horn cells of the spinal cord.

Functions. Inferior olivary nucleus receives afferent fibres from the cerebral cortex, corpus striatum, red nucleus and spinal cord. It influences muscle activity. Probably, it is involved in the reflex movements arising from the proprioceptors.

6. Medial longitudinal fasciculus

Origin. The medial longitudinal fasciculus (MLF) extends from the mid brain downwards. The fibres of this tract take origin from different area of the brain stem namely:

- Vestibular nuclei,
- Reticular formation,
- Superior colliculus,
- Interstitial nucleus of Cajal and
- Nucleus of posterior commissure.

Course. The MLF in the brain stem is closely related to the nuclei of third, fourth, sixth and twelfth cranial nerves. It is also related to the fibres of seventh nerve (as they wind round the abducent nucleus), and to some fibres arising from the cochlear nuclei. Below, the MLF becomes continuous with the anterior intersegmental tract of spinal cord (Fig. 10.1-9), which descends through the posterior part of anterior white funiculus. This tract is well defined only in the upper cervical segments. Below this level, the fibres run along with the fibres of medial vestibulospinal tract.

Termination. Along with the fibres of the medial vestibulospinal tract, the fibres of this tract make connections with ventral horn cells that innervate the muscles of neck.

Functions. MLF plays an important role in the pathway of ocular movements. Its function can be summarized as:

- It ensures harmonious movements of the eyes and neck (head) in response to vestibular stimulation and auditory stimuli.
- It facilitates simultaneous movements of the lips and tongue as in speech.

DESCENDING TRACTS ENDING IN THE BRAIN STEM

Corticonuclear tracts

Origin. These arise from the cerebral cortex along with the corticospinal tracts (see page 702).

Course and termination. These fibres descend along with corticospinal tract fibres as part of corona radiata and then pass through posterior limb of the internal capsule. In the brain stem, they cross to the opposite side at various levels and end by synapsing with cells of the cranial nerve nuclei, either direct or through interneurons.

Functions. The nuclei of cranial nerves that supply skeletal muscles are functionally equivalent to ventral horn cells of the spinal cord. These are controlled by corticonuclear fibres.

Cortico-ponto-cerebellar pathway

Origin. This pathway consists of the fibres arising in the cerebral cortex of the frontal, temporal, parietal and occipital lobes.

Course. After origin from the cerebral cortex, the fibres descend through the corona radiata and internal capsule to reach the crus cerebri (Fig. 10.1-9). These fibres synapse with the pontine nuclei of the same side. Axons of the neurons in the pontine nuclei form the transverse fibres of the pons. These fibres cross the mid line and pass into the middle cerebellar peduncle of the opposite side and reach the cerebellar cortex.

Functions. This pathway forms the anatomical basis for control of cerebellar activity of cerebral cortex.

Other fibres ending in the brain stem

Other fibres arising from the cerebral cortex end in the following masses of grey matter of brain stem:

- Red nucleus (corticorubral fibres),
- Tectum (corticotectal fibres),
- Substantia nigra,
- Inferior olivary nucleus (cortico-olivary fibres) and
- Reticular formation (corticoreticular fibres).

The above fibres ultimately form part of extrapyramidal system.

Table 10.1-2 Major descending tracts of the spinal cord

Tract	Location	Origin*	Termination	Functions
Pyramidal tracts				
• Lateral corticospinal (crossed pyramidal) tract	Lateral white column of spinal cord.	Primary motor cortex (area 4), pre-motor cortex (area 6) of the opposite cerebral hemisphere (*upper motor neurons*).	Anterior horn cells of the spinal cord (*lower motor neurons*).	Controls conscious skilled movements especially of hands (contraction of individual or small group of muscles particularly those which move hands, fingers, feet and toes).
• Anterior corticospinal (uncrossed pyramidal) tract	Anterior white column.	Primary motor cortex (area 4), pre-motor cortex (area 6) of the opposite cerebral hemisphere (*upper motor neurons*).	Anterior horn cells of the spinal cord (*lower motor neurons*).	Same as that of lateral corticospinal tracts.
Extrapyramidal tracts				
• Rubrospinal tract	Lateral white column.	Red nucleus of the opposite side located in mid brain.	Anterior horn cells of the spinal cord.	Unconscious co-ordination of movements (controls muscle tone and synergy).
• Vestibulospinal tract	Anterior white column.	Vestibular nucleus.	Anterior horn cells of the spinal cord.	Unconscious maintenance of posture and balance.
Reticulospinal tracts				
• Medial reticulospinal tract	Anterior white column.	Reticular formation in medulla.	Anterior horn cells of the spinal cord.	Mainly responsible for inhibitory influence on the motor neurons to the skeletal muscles.
• Lateral reticulospinal tract	Lateral white column.	Reticular formation in mid brain, pons and medulla.	Anterior horn cells of the spinal cord.	Mainly responsible for facilitatory influence on the motor neurons to the skeletal muscles.
Tectospinal tract	Anterior white column.	Superior colliculus of the opposite side.	Cranial nerve nuclei in medulla and anterior horn cells of the upper spinal segments.	Controls movements of head, neck and arms in response to the visual stimuli.

*Location of cell bodies of neurons from which the axons of tract arise.

The major descending tracts in the spinal cord are summarized in Table 10.1-2.

LESIONS OF SPINAL CORD

TRANSECTION OF THE SPINAL CORD

Transection of the spinal cord can be divided into three types:

- Complete transection,
- Incomplete transection and
- Hemisection.

COMPLETE TRANSECTION OF SPINAL CORD

Common causes of complete transection are:

- Gunshot injuries,
- Dislocation of spine and
- Occlusion of the blood vessels.

Common site of involvement is at the mid thoracic level.

Clinical stages

The effects (symptoms and signs) produced by complete transection of the spinal cord occur in following three stages:

- Stage of spinal shock,

- Stage of reflex activity and
- Stage of reflex failure.

A. Stage of spinal shock

Spinal shock refers to the cessation of all the functions and activity below the level of the section immediately after injury.

Effects depend on the site of injury, complete transection in cervical region (above C_5) is usually fatal, because of cutting of connections between respiratory centre and respiratory muscles leading to paralysis of respiratory muscles.

In quick transection of spinal cord, the patient feels as it has been cut into two portions, the upper portion (higher centres and mind) is unaffected, but the whole body below the level of injury is deprived of all the sensations and motor activity.

Cause of stage of spinal shock (also called stage of flaccidity) is not known, but it is related to cessation of tonic neuronal discharge from upper brain stem or supraspinal pathway.

Duration and severity of spinal shock depends upon the evolution of animal. Higher the animal, more profound and longer lasting is the spinal shock. This is probably due to encephalization, i.e. greater dependence of spinal cord on higher centres. Therefore, spinal shock lasts for few minutes in frogs, for few hours in cats and dogs, for days in monkeys and in human beings it lasts for about 3 weeks.

In higher animals, the entire nervous system is integrated as a functional unit, therefore damage to any part of the nervous system disturbs its smoothness of working and the functional failure is more severe. This is called *diaschisis*.

Characteristic effects during spinal shock can be summarized:

1. Motor effects include:

- *Paralysis of the muscles* occurs below the level of section. Depending upon the site of lesion, when both lower limbs are paralysed (transection between cervical and lumbosacral enlargements), it is called *paraplegia* and when all the four limbs are affected (transection below C_5) it is called *quadriplegia.*
- *Loss of tone* occurs in the paralysed muscles. So the muscles become atonic or flaccid. This is called state of flaccid paralysis.
- *Areflexia*, i.e. all the superficial and deep reflexes are markedly decreased or lost.

2. Sensory effects. All the sensations are lost below the level of transections.

3. Vasomotor effects. The sympathetic vasoconstrictor fibres leave the spinal cord between T_1 and L_2. Therefore, depending upon the site of lesion, the vasomotor effects produced are:

- Transection of cord below L_2 segment will produce no effect or very little fall in the blood pressure.
- Transection at the level of T_1 segment cut off all the thoracolumbar sympathetic neurons from the medullary cardiovascular centre. As a result, there occurs loss of sympathetic tonic discharge causing arteriolar dilatation leading to a sharp fall in blood pressure (MBP may fall from a normal resting value of 100 mm Hg to about 40 mm Hg). Fall in blood pressure is less marked as the section shifts more distally towards L_2 segment.

Absence of movements due to paralysis of muscles further retards the circulation and also the venous return producing cold and blue (cyanotic) extremities. Skin becomes dry and scaly and bed sores may develop.

Note. It is important to note that after paralysis of the muscles the body temperature becomes subnormal (as muscular contraction is a major source of heat production). When hot bottles are given to raise the body temperature, under such circumstances bed sores develop.

4. Visceral effects produced are:

- *Urinary bladder* is paralysed, however, the sphincter vesicae regains tone early leading to retention of urine.
- *Rectum* is also paralysed. Since the bowels become hypotonic there occurs constipation.
- Penis becomes flaccid and erection becomes impossible.
- When lesion is at T_6 level, all impulses coming in from the abdominal viscera are cut off from the brain; therefore, gripping sensations or distension of viscera are not appreciated.

B. Stage of reflex activity

If the patient survives the stage of spinal shock, gradually he/she gains few functions. That is why this is also called stage of recovery. After about 3 weeks period, depending largely upon the general health of the patient, the reflex activity begins to return to the isolated segments of spinal cord below the level of lesion. Various developments which take place, in a chronological order, in this stage are:

1. Smooth muscles regain functional activity first of all and urinary bladder becomes automatic, i.e. reflex evacuation is gradually established in a perfectly normal manner. Similarly, reflex defaecation is also established.

2. Sympathetic tone of the blood vessels is regained, next to smooth muscles, when connector cells in spinal cord begin to act independently of the vasomotor centre (VMC). As a result:

- *Blood pressure* is restored to normal,

- *Skin*, which has become dry and scaly, now shows sweating again and becomes more healthy. Bed sores, if any, heal up rapidly.

3. Skeletal muscle tone then recovers slowly after 3–4 weeks. Recovery of muscle tone is reflex in character and is produced by impulses entering the spinal cord from the muscles.

- Tone of flexor muscles returns first, therefore, flexors become less hypotonic than extensors leading to '*paraplegia in flexion*' (both lower limbs are in state of flexion).
- In 'spinal man' the limbs cannot support the weight of the body.
- No wasting of muscles is seen because though the muscles are paralysed for voluntary movements they are in constant reflex activity.

4. Reflex activity begins to return after few weeks of recovery of muscle tone. Recovery of reflex excitability is due to the development of *denervation hypersensitivity* to the mediators released by the remaining spinal excitatory endings and the growing of collaterals from existing neurons with the formation of additional excitatory ending on interneurons and motor neurons.

- *Flexor reflexes* return first, and to elicit flexor reflex a painful stimuli is required. The first reflex which usually appears is Babinski's reflex (i.e. *Babinski's* sign is positive).
- *Extensor reflexes* return after a variable period of 1–5 weeks of appearance of flexor reflexes. Initially, the *knee jerk* appears, then the *ankle jerk* may return still later. Generally, about 6 months after the occurrence of transection marked activity appears in the extensor arcs. This results in an exaggerated extensor reflexes with the appearance of extensor spasms.
- *Mass reflex* can be elicited in some cases by scratching the skin over the lower limbs or the anterior abdominal wall, depending upon the level of lesion. It is characterized by spasm of flexor muscles of both the limbs, evacuation of bladder and profuse sweating below the level of the lesion.

Note. The mass reflex can be utilized to provide paraplegic patients a degree of bladder and bowel control. Patients can be trained to initiate urination and defaecation by intentionally producing mass reflex with the help of a stroke or a pinch on their thighs.

C. Stage of reflex failure

The failure of reflex activity may occur when general condition of the patient starts deteriorating due to malnutrition, infections or toxaemia, under such circumstances:

- Reflexes become more difficult to elicit,
- The threshold for stimulus increases,
- Mass reflex is abolished and
- The muscles become extremely flaccid and undergo wasting.

INCOMPLETE TRANSECTION OF SPINAL CORD

In incomplete transection, the spinal cord is gravely injured but does not suffer from complete transection (i.e. a few tracts are intact).

Effects

Effects of incomplete transection can be divided into three clinical stages:

- Stage of spinal shock,
- Stage of reflex activity and
- Stage of reflex failure.

A. Stage of spinal shock

Features of this stage are similar to those described in the stage of spinal shock of complete transection of spinal cord (see page 707).

B. Stage of reflex activity

Features of this stage differ remarkably from that of the stage of reflex activity of complete transection of spinal cord:

1. Tone appears in extensor muscles first (cf complete transection in which tone appears in flexor muscles first). This is because of the fact that in incomplete transection, some of the descending fibres in the lateral column of the cord, especially the vestibulospinal and reticulospinal tracts may escape injury; and both these tracts mainly reinforce activity of extensor motor neurons. Because of comparatively higher tone in the extensor muscles, a condition called 'paraplegia in extension' results (cf complete transection in which paraplegia in flexion is seen).

2. Extensor reflexes (stretch reflexes) return first and flexor reflexes reappear later (cf complete transection in which flexor reflexes return first). Extensor reflexes which can be elicited in this stage in incomplete transection and not in complete transection) are:

- *Phillipson reflex.* It refers to the extension of the opposite limb produced by gentle flexion of one limb. The flexed limb then becomes extended and the opposite one flexed, i.e. the response alternates in each limb producing a steppage movements.
- *Extensor thrust reflex.* It refers to a physiological extensor response (i.e. active contraction of quadriceps, and posterior calf muscles with straightening of limb) obtained by pressing the foot upward with the palm of the hand in a patient in whom the lower limb has been passively flexed and allowed to rest on the bed.

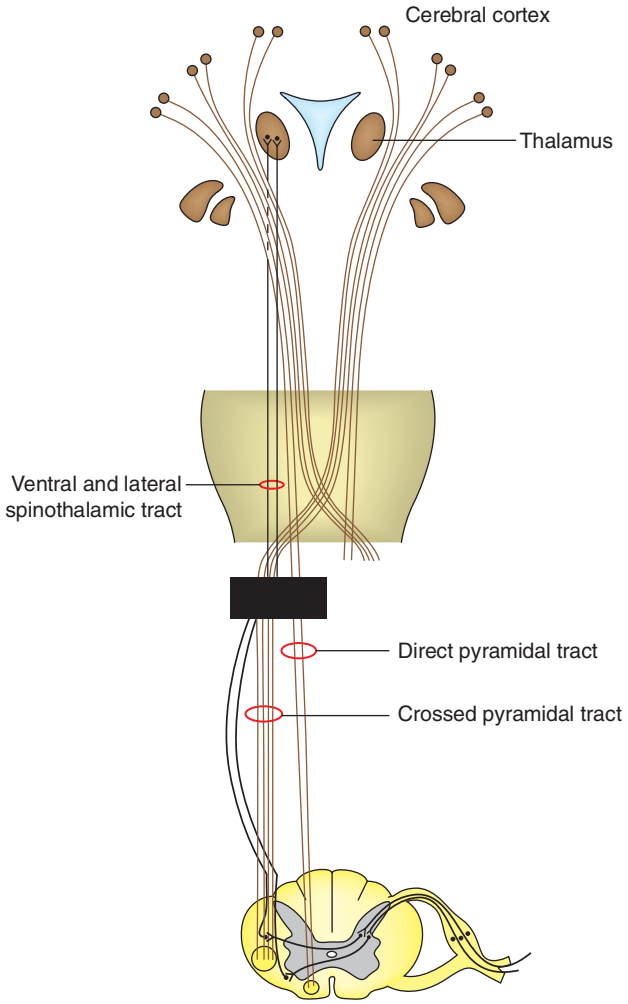


Fig. 10.1-15 Hemisection of the spinal cord.

- *Crossed extensor reflex.* It refers to the occurrence of forcible extension of the opposite limb associated with withdrawal (flexor) reflex produced by noxious stimulus to the sole of foot of one limb.

3. Mass reflex is not elicited in incomplete transection (cf complete transection in which mass reflex is elicited). This is because the controlling effect of brain stem persists through motor fibres (vestibulospinal and reticulospinal), which have escaped injury.

C. Stage of reflex failure

Features of this stage are similar to that of stage of reflex failure with complete transection of spinal cord (see page 708).

HEMISECTION OF THE SPINAL CORD (BROWN-SEQUARD SYNDROME)

Hemisection of the spinal cord refers to a lesion involving one lateral half of the spinal cord (Fig. 10.1-15). It can occur in following accidental injuries. It can also be produced for experimental studies in the animals.

The effects of hemisection of the spinal cord can be described in two stages:

- Immediate effects and
- Late effects.

Immediate effects

Immediate effects following hemisection of the spinal cord are those of 'spinal shock' (see page 707).

Late effects

If the patient survives, typical motor and sensory changes develop after recovery from the spinal shock. These changes constitute the *Brown-Sequard syndrome* and can be described as:

- Changes at the level of section,
- Changes below the level of section and
- Changes above the level of section.

A. Changes at the level of hemisection

I. *Changes on the same side*

1. Sensory changes. All the sensations are lost (complete anaesthesia) at the level of hemisection on the same side. This occurs because of complete damage to posterior nerve root, posterior horn cells and spinothalamic fibres (which cross to the opposite side).

2. Motor changes at the level of hemisection on the same side include:

(i) Complete lower motor neuron (LMN) type paralysis is seen due to damage to the anterior horn cells. That is:

- Flaccid paralysis of muscles (paralysis with loss of muscle tone),
- All the reflexes are lost,
- Muscle power is lost and ultimately
- Muscles degenerate and undergo wasting due to loss of tone. For detailed features of LMN paralysis (see page 711).

(ii) Complete and permanent vasomotor paralysis occurs due to damage of the lateral horn cells.

II. *Changes on the opposite side*

1. Sensory changes. There occurs some loss of pain, temperature and crude touch sensations due to injury to the fibres of spinothalamic tract, which cross horizontally in the same segment and may be caught up in the lesion. But tracts of *Gall and Burdach* (fasciculus gracilis and fasciculus cuneatus) are not affected, so the sensations carried by these two tracts are not affected.

2. Motor changes. Usually no motor change occurs. If it occurs, it is very mild and is similar to the effects of lower motor neuron lesion.

B. Changes below the level of section

I. *Changes on the same side*

1. Sensory changes. There is dissociated sensory loss:

- *Injury to uncrossed fibres of tracts of Gall and Burdach* causes loss of fine touch, tactile localization, tactile discrimination, sensation of vibration, conscious kinaesthetic sensation and stereognosis.
- *No injury to spinothalamic tracts,* which cross to the opposite side so crude touch, pain and temperature sensations are not lost.

2. Motor changes. There occurs upper motor neuron (UMN) type of paralysis due to injury to the pyramidal tracts. Features of UMN paralysis include:

- Increased muscle tone, leading to spastic paralysis,
- Loss of superficial reflexes,
- Exaggeration of deep reflexes,
- Positive Babinski's sign,
- Rigidity of limbs and
- No degeneration and wasting of muscles.

3. Vasomotor changes. There occurs temporary loss of vasomotor tone due to damage to the descending fibres from the VMC in the medulla to the lateral horn cells. This leads to:

- Dilatation of blood vessels and
- Fall in blood pressure.

However, soon the intact lateral horn cells start acting as supplementary VMC and vasomotor tone returns leading to normalization of blood pressure.

II. *Changes on the opposite side*

1. Sensory changes. Dissociated sensory loss occurs as:

- Injury to crossed spinothalamic tracts causes loss of following sensations on the opposite side below the level of lesion: crude touch, pain and temperature.
- No injury to uncrossed tracts of Gall and Burdach, so following sensations on the opposite side below the level of lesion are not lost: fine touch, tactile localization, tactile discrimination, vibratory sense, conscious kinaesthetic sensation and stereognosis.

2. Motor changes. Usually, there occurs no motor change on the opposite side below the level of lesion. Upper motor neuron lesion-type paralysis of a few muscles, however, may occur sometimes due to possible damage to the some fibres of direct pyramidal tracts of the same side when these fibres cross.

In nutshell, below the level of lesion there occurs:

- Extensive motor loss but little sensory loss on the same side and
- Extensive sensory loss but little motor loss on the opposite side.

C. Changes above the level of lesion

I. *Changes on the same side*

1. Sensory changes. A band of hyperaesthesia, i.e. increased cutaneous sensations are present in one or two segments above the level of section on the same side. This occurs due to irritation of the neighbouring posterior nerve roots above the level of section.

2. Motor changes. Twitching of muscle in upper one or two segments on the same side may occur due to irritation of the neighbouring anterior nerve roots above the level of section.

LESIONS OF SENSORY SYSTEM IN SPINAL CORD

DEAFFERENTATION (DORSAL NERVE ROOT LESION)

Injury to the dorsal nerve root (afferent nerve) produces following effects:

1. Loss of all sensations, i.e.

- Loss of exteroceptive senses with anaesthesia and analgesia.
- Loss of conscious muscle sense producing ataxia,
- Loss of unconscious muscle sense from stretch receptors of muscle spindle, with hypotonia or atonia and
- Loss of visceral senses.

2. Loss of all reflexes. All the reflexes, superficial as well as deep, are lost.

3. Muscle tone is lost.

4. Marked weakness in the movements of parts occurs because the higher centres concerned with reflex control of posture are deprived of afferent impulses from joints and muscles.

SYRINGOMYELIA

Syringomyelia is a rare disease, in which there occurs excessive overgrowth of neuroglial tissue accompanied by cavitation in the grey matter around the central canal of the spinal cord. This disease involves the cervical enlargement of the cord more frequently.

Characteristic features

I. Sensory features are predominant and occur in the form of dissociated anaesthesia, i.e. loss of pain and temperature with retention of touch sensation.

II. Motor features may also occur due to further spread of gliosis and cavitations:

1. Flaccid paralysis of the upper limb muscle (LMN-type paralysis) may occur initially due to involvement of anterior horn cells.

2. Progressive spastic paralysis of the legs (UMN-type paralysis) may occur later on due to involvement of pyramidal and extra pyramidal tracts.

TABES DORSALIS

Tabes dorsalis is a disease, usually caused by syphilis, in which there occurs bilateral degeneration of posterior nerve roots and posterior funiculi, especially fasciculus gracilis. It is characterized by following features:

1. Lightening pains occur in intermittent attacks due to stimulation of pain fibres in dorsal nerve roots in the initial stages.

2. Loss or decrease of pain sensibility occurs after sometime producing following features:

- *Trophic disturbances* in the form of perforating ulcers of the skin at pressure points. *Charcot joint refers* to the deformed joints produced by repeated trauma due to loss of pain sensations. There is no proper support and movements of the joints of the body become uncontrolled.
- Anaesthesia of central part of face occurs due to involvement of fifth cranial nerve.
- Anaesthesia at the upper chest, inner border of hands, around the anus and over the legs occurs due to the involvement of dorsal nerve roots in cervicothoracic and lumbosacral regions.

Table 10.1-3 Lower versus upper motor neuron lesions

Lower motor neuron lesion (LMNL)	Upper motor neuron lesion (UMNL)
1. LMNL refers to the involvement of neurons (α and γ) of anterior horn of spinal cord and neurons of cranial nerve nuclei.	1. UMNL refers to the involvement of motor neurons that influence the activity of LMN of spinal cord or cranial nerve nuclei located in brain stem. Thus in UMNL the pyramidal and extrapyramidal descending tracts are involved.
2. LMN paralysis is typically observed in poliomyelitis, when the polio virus selectively affects the lower motor neurons of spinal cord and brain stem.	2. UMN paralysis occurs commonly in vascular accidents or space occupying lesions.
3. Usually a single or individual muscle is affected.	3. Usually a group of muscles are affected.
4. Flaccid paralysis of the involved muscle as muscle tone is lost due to involvement of stretch reflex arc.	4. Spastic paralysis of the involved muscles as the inhibitory higher control is lost and stretch reflex arc is intact.
5. Muscle power is lost and ultimately muscles degenerate and undergo wasting due to disuse (disuse atrophy).	5. No degeneration and wasting of muscles as they are constantly involved in reflex activity (though the voluntary movements are lost).
6. Areflexia, i.e. all the superficial as well as deep reflexes are lost.	6. Superficial reflexes {abdominal, cremasteric, anal are lost but deep reflexes are exaggerated (because of increased gamma-motor discharge)}.
7. Babinski's sign is negative, i.e. on stroking the outer edge of sole of the foot with firm tactile stimulus there occurs plantar flexion (downward movement). It is called flexor response (withdrawal reflex) and is considered a normal response.	7. Babinski's sign is positive, i.e. on stroking the outer edge of the sole of the foot with firm stimulus, there occurs dorsiflexion of the great toe and fanning out (abduction) of small toes. It is also called extensor response. Positive Babinski's sign indicates involvement of corticospinal tract. In normal infants, this sign is positive prior to myelination of the corticospinal tract (i.e. below 1 year of age).
8. Clonus is absent.	8. Clonus is present. It refers to a sustained series of rhythmic muscle jerks when a quick stretch is applied to a tendon. Ankle clonus is usually observed in UMNL by a sudden dorsiflexion of the foot.
9. Clasp knife reflex is absent.	9. Clasp knife reflex is present, i.e. muscular resistance to passive movement is exaggerated, this resistance is strong at the beginning of movement, but yields suddenly in a clasp knife fashion as more force against resistance is applied. The initial resistance is offered because of the stretch reflex developed in extensor muscles, e.g. triceps of the elbow. The sudden relax of resistance is due to the activation of inverse stretch reflex.

3. Loss of deep sensations. Following sensations are lost on the same side at and below the level of lesion: position sense, vibratory sense, sense of stereognosis and discriminative touch.

4. Loss of reflexes. Both superficial and deep reflexes are lost in tabes dorsalis mostly because of loss of sensations.

5. Sensory ataxia occurs due to lack of co-ordination of voluntary involvement. In it the patient walks on a broad base with the legs apart and eyes fixed to the ground for correcting the steps. Typically, the patient raises the legs excessively high and slopes the feet on the ground. *Romberg's sign* is positive.

MULTIPLE SCLEROSIS

Multiple sclerosis (MS) is a demyelinating disorder having widespread disseminated involvement of white matter of the CNS. Because of this it was also called disseminated sclerosis. It is currently considered to be an autoimmune disease, pathologically characterized by focal inflammation, demyelination and gliosis or scarring. A remitting and relapsing course is the most common, with either complete recovery or residual damage with each attack.

Manifestations of MS depend upon the area of CNS involved. Commonest symptoms reported are:

- Limb weakness (75%),
- Sensory loss (37%),
- Paraesthesia (24%) and
- Optic neuritis (37%).
- Diplopia, vertigo and ataxia are comparatively less common.

SUBACUTE COMBINED DEGENERATION OF THE SPINAL CORD

- Subacute combined degeneration of the spinal cord is usually associated with pernicious anaemia, due to lack of intrinsic factor which is essential for the absorption of vitamin B_{12}.

In this condition there occur bilateral degeneration of white fibres of the dorsal column and lateral column of the spinal cord, especially involving the lumbosacral segments. Its manifestations include:

- Loss of position and vibrating senses of the lower extremities and
- Signs of upper motor neuron lesions, such as bilateral spasticity, exaggerated tendon reflexes and positive Babinski's sign.

LESIONS OF MOTOR SYSTEM

Lower versus upper motor neuron lesion

Difference between lower motor neuron lesions and upper motor neuron lesions are shown in Table 10.1-3.

Chapter 10.2

Physiological Anatomy, Functions and Lesions of Cerebellum and Basal Ganglia

CEREBELLUM

PHYSIOLOGICAL ANATOMY

Cerebellum, the largest part of hind brain, consists of two lateral parts called the *cerebellar hemispheres* connected in the midline by a narrow central region called the *vermis*.

EXTERNAL FEATURES (FIG. 10.2-1)

Surfaces. The cerebellum has two surfaces:

- *Superior surface* is related to the tentorium cerebelli and
- *Inferior surface* is related to the *hollow* of occipital bone.

Folia. The surfaces of the cerebellum are thrown into numerous transverse folds called folia.

Fissures. The surface of the cerebellum presents three main fissures (Fig. 10.2-1):

1. Primary fissure lies on the anterosuperior aspect of the cerebellum. It is V-shaped with open being forwards. It forms the posterior limit of the anterior lobe.

2. Horizontal fissure separates the superior surface of the cerebellum from its inferior surface and thus follows the convex posterior and anterolateral border of the cerebellar hemisphere. It is of no functional significance.

3. Posterolateral fissure is situated anteriorly on the inferior surface of the cerebellar hemisphere and separates the posterior lobe from the flocculonodular lobe.

ANATOMICAL PARTS

The cerebellum consists of two cerebellar hemispheres and a median vermis has been divided into many parts which have functional and morphological significance. To show the various parts of cerebellum in a single illustration, it is usual to represent the organ as if it has been opened out (flattened) so that the superior and inferior aspects both can be seen (Fig 10.2-1).

Parts of vermis

Vermis is so named because it resembles a worm, which is bent on itself to form a complete circle. Superior and inferior

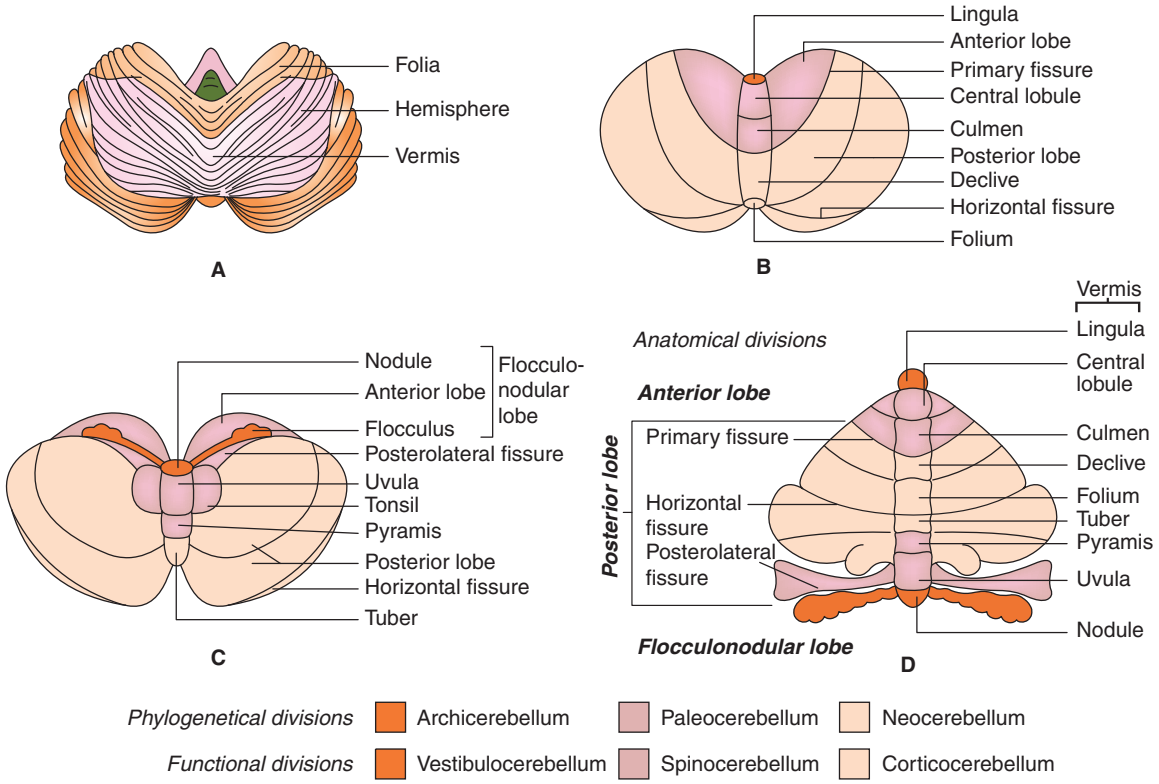


Fig. 10.2-1 Gross anatomy of cerebellum: A, superior surface showing folia; B and C, superior and inferior surface showing fissures, lobes and parts of vermis and D, schematic diagram to show the parts of vermis and hemisphere in unrolled cerebellum.

surfaces of the vermis are termed superior vermis and inferior vermis. Proceeding from above downwards the opened up vermis (as seen in Fig. 10.2-1) consists of following parts: lingula, central lobule, culmen, declive, folium, tuber, pyramis, uvula and nodule.

Parts of hemisphere

With the exception of the lingula, each part of the vermis is related laterally to a part of hemisphere as shown in Fig. 10.2-1 and Table 10.2-1.

ANATOMICAL DIVISIONS

Anatomically, the cerebellum has been divided into three lobes:

1. Anterior lobe is that part of cerebellum which lies in front of the primary fissure on the superior surface (Fig. 10.2-1). The parts of vermis and hemisphere forming the anterior lobe are shown in Fig. 10.2-1 and Table 10.2-1.

2. Posterior lobe is that part of cerebellum which lies between the primary fissure and posterolateral fissure. Parts of the vermis and hemisphere forming the posterior lobe are shown in Fig. 10.2-1 and Table 10.2-1.

Table 10.2-1 Lobes of cerebellum and parts of vermis and hemisphere forming them

Lobes of cerebellum	Part of vermis	Part of hemisphere
Anterior lobe	Lingula	No lateral projection
	Central lobule	Alae
	Culmen	Anterior quadrangular lobule
Primary fissure		
Posterior lobe	Declive	Posterior quadrangular lobule
	Folium	Superior semilunar lobule
Horizontal fissure		
	Tuber	Inferior semilunar lobule
	Pyramis	Biventral lobule
	Uvula	Tonsil
Posterolateral fissure		
Flocculo-nodular lobe	Nodule	Flocculus

3. Flocculonodular lobe is that part of the cerebellum which lies anterior to the posterolateral fissure on the inferior surface. It consists of (Fig. 10.2-1 and Table 10.2-1.):

- *Nodule,* which is rostral part of the vermis and

- *Flocculli,* which are irregular shaped masses attached to nodule on each side. They are almost completely separated from the rest of cerebellum.

PHYLOGENETICAL DIVISIONS

Phylogenetically, i.e. according to evolutionary stages, the cerebellum consists of three subdivisions:

1. Archicerebellum. It is the oldest part to develop. It consists of (Fig. 10.2-1):

- Flocculonodular lobe and
- Lingula.

2. Paleocerebellum. Phylogenetically, it is the next part to appear. It consists of (Fig. 10.2-1):

- Entire anterior lobe except lingula, and following parts of posterior lobe—pyramis, uvula, and paraflocculus.

3. Neocerebellum. It is the latest part to develop. It consists of whole of the posterior lobe except pyramis and uvula (Fig. 10.2-1).

FUNCTIONAL DIVISIONS

Functionally, cerebellum is divided into three divisions:

1. Vestibulocerebellum. It includes the flocculonodular lobe, which is its principal component and has vestibular connections only.

- Nucleus fastigial is its effector nucleus.
- It is concerned with control of body posture and equilibrium.

2. Spinocerebellum. It includes the parts forming paleocerebellum, i.e. entire anterior lobe except lingula and some parts of posterior lobe (pyramis, uvula and paraflocculus).

- *Nucleus interpositus,* i.e. nucleus globossus and nucleus emboliformis are its effector nuclei.
- It receives proprioceptive inputs from the spinal cord and is concerned with control of axial (trunk) and limb muscles postural reflexes.

3. Corticocerebellum. Corticocerebellum also called as central cerebellum includes whole of the posterior lobe except pyramis and uvula.

- Nucleus dentatus is its effector nucleus.
- It occupies the more lateral regions of the cerebellar cortex and receives information from the cerebral cortex and pons.
- It is concerned with smooth performance of highly skilled voluntary movements.

HISTOLOGICAL STRUCTURE

Histologically, cerebellum consists of (Fig. 10.2-2):

- Cerebellar cortex (outer grey matter layer),
- White matter (formed by afferent and efferent nerve fibres of cerebellum) forming medullary core and
- Deep cerebellar nuclei (masses of grey matter embedded in the medullary core).

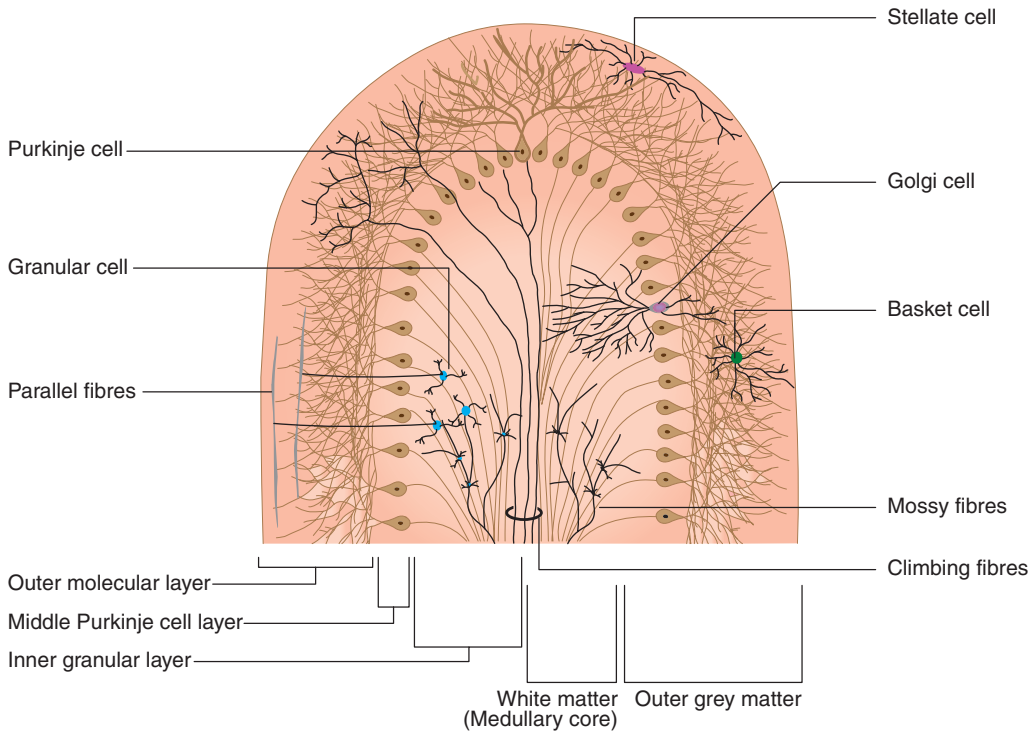


Fig. 10.2-2 Histology of cerebellar cortex.

I. Cerebellar cortex

Grossly in cut section, the cerebellar cortex is seen as extensively folded on itself constituting the *folia,* i.e. leaf-like parts which are marked off from one another by fissures. In striking contrast to the cortex of the cerebral hemisphere, the cerebellar cortex has a uniform structure in all parts of the cerebellum.

Microscopically, the grey matter of cerebellar cortex consists of five main types of neurons (stellate cells, basket cells, Purkinje cells, granule cells and Golgi cells) which are arranged in three layers:

- Molecular layer (most superficial),
- Purkinje cell layer (middle layer) and
- Granule cell layer (inner layer).

1. Molecular layer is composed of two types of neurons (stellate and basket cells) and unmyelinated nerve fibres.

(i) Stellate cells are star-shaped and more superficially located. Their *dendrites* synapse with parallel fibres of granule cells and their *axons* synapse with dendrites of Purkinje cells.

(ii) Basket cells are located deep in the molecular layer.

- They receive inputs from the parallel fibres.
- Their axons branch and form basket around the cell bodies of Purkinje cells (hence the name).
- Each basket cell may synapse with about 70 Purkinje cells.

Nerve fibres present in the molecular layer are parallel fibres (axon of granule cells), dendrites of Purkinje cells and climbing fibres from inferior olivary nucleus.

2. Purkinje cell layer. It is composed of a single layer of large flask-shaped Purkinje cells (biggest neurons in the body).

- Dendrites of Purkinje cells extend into the molecular layer and provide a huge surface area for axodendritic synapses.
- Axons of these cells make synaptic connection with the deep cerebellar nuclei in the medullary core. They act as the sole output neurons from the cerebellar cortex and exert inhibitory influence to the deep cerebellar and lateral vestibular nuclei.

3. Granule cell layer consists of granule and Golgi cells, with their processes and sensory mossy fibres with their synaptic glomeruli.

(i) Granule cells are very small, numerous (about 10 billion) spherical neurons. Axons of the granule cells ascend into the molecular layer and form the parallel fibres which make excitatory synapses with dendrites of Purkinje cells, stellate cells, basket cells and Golgi cells.

(ii) Golgi cells are large cells and less numerous than the granule cells.

- Their dendrites project into the molecular layer and receive inputs from the parallel fibres.
- Their cell bodies receive inputs via collaterals from the incoming climbing fibres and Purkinje cells.
- Their axons branch extensively and make inhibitory synaptic connection with the dendrites of granule cells.

Sensory inputs to cerebellar cortex

The cerebellar cortex receives sensory inputs (afferent fibres) from other parts of the brain by two types of sensory fibres: the climbing fibres and mossy fibres. Both sets of fibres reach the cerebellum through the peduncles and are excitatory in nature.

1. Climbing fibres arise from the neurons of inferior olivary nucleus situated in medulla and reach the cerebellum via olivocerebellar tract.

- Climbing fibres establish 'one-to-one' connection with the Purkinje cell dendrites and excite them to discharge.
- Collaterals of climbing fibres synapse with all other types of neurons in the cerebellar cortex and also the deep cerebellar nuclei.

2. Mossy fibres. Unlike climbing fibres, the mossy fibres have many sources of origin. These are axons of spinocerebellar, vestibulocerebellar, reticulocerebellar, cuneocerebellar and cortico-ponto-cerebellar tracts. Mossy fibres are named so because they resemble moss plant.

- Each mossy fibre makes synaptic connections with the dendrites of many granule cells forming synaptic glomeruli. The glomeruli also contain the inhibitory ending of Golgi cells.
- Each mossy fibre activates about 450 Purkinje cells through the granule cells and their parallel fibres. Thus a climbing fibre excites a single Purkinje cell, whereas a mossy fibre through granule cells and parallel fibres fires several thousand Purkinje cells.

  Note. Both the climbing and mossy fibres of sensory inputs exert excitatory influence.

- Out of five types of neurons in the cerebellar cortex, only the granule cell is excitatory; it releases the excitatory neurotransmitter *glutamate*. The other four types of neurons are inhibitory and release the inhibitory neurotransmitter *gamma aminobutyric acid (GABA).*

II. White matter of cerebellum

The cerebellar cortex, i.e. outer grey matter surrounds inner medullary core of white matter (an arrangement opposite to what is seen in spinal cord). White matter is formed by both

afferent and efferent fibres. These fibres can be classified in three groups:

1. Projection fibres of the cerebellum leave or enter the cerebellum and connect it with other parts of the central nervous system. These are arranged in three bundles:

- *Inferior cerebellar peduncle* consists of fibres connecting cerebellum with medulla.
- *Middle cerebellar peduncle* contains the fibres which connect cerebellum with pons and
- *Superior cerebellar peduncle* which connects the cerebellum with mid brain.

2. Association fibres connect different regions of the same cerebellar hemisphere.

3. Commissural fibres connect the areas of two halves of cerebellar cortex with each other.

III. Deep cerebellar nuclei

Within the white matter of medullary core of cerebellum are embedded four pairs of masses of grey matter called deep cerebellar nuclei. They lie in close relationship with the roof of the fourth ventricle, and therefore are known as roof nuclei (Fig. 10.2-3).

1. Dentate nucleus or nucleus dentatus is the largest cerebellar nucleus. Shaped like a purse made up of a crenated lamina of grey matter it has an open anteriomedian hilum to receive the fibres of superior cerebellar peduncle.

2. Emboliform nucleus or nucleus emboliformis is an oval mass of grey matter located just anteromedian to the hilum of dentate nucleus.

3. Globossus nucleus lies medial to the nucleus emboliformis and therefore, together are referred to as nucleus interpositus.

4. Fastigial nucleus or nucleus fastigii is nearly spherical and lies close to the midline just over the roof of fourth ventricle.

NEURAL CIRCUITS AND NEURONAL ACTIVITY IN CEREBELLUM

Cerebellum executes its functions through excitatory output of the deep cerebellar nuclei to the brain stem and thalamus. Neural connections within the cerebellar cortex, i.e. intrinsic cerebellar circuit (Fig. 10.2-4) is basically concerned with modulating or timing the excitatory output of deep cerebellar nuclei via the fibres of Purkinje cells. This is done in accordance with the signals received by the cerebellar cortex from different parts of the brain and body. The entire process can be discussed, for the purpose of understanding only, under following headings:

- Afferents to cerebellar cortex,
- Neuronal activity of intrinsic cerebellar circuitry and
- Neuronal activity of deep cerebellar nuclei.

AFFERENTS TO CEREBELLAR CORTEX

Afferents to cerebellar cortex reach via two types of fibres:

1. Climbing fibres. These fibres represent terminations of axons reaching the cerebellum from the inferior olivary nucleus. The climbing fibres excite the Purkinje cells directly and the deep cerebellar nuclei via the collaterals by releasing the excitatory neurotransmitter aspartate.

2. Mossy fibres. All the afferent fibres entering the cerebellum, other than the olivocerebellar, are called mossy fibres. These fibres after reaching the granular layer of cerebellar cortex branch profusely and then each branch terminate within a glomerulus (Fig. 10.2-4). The glomerulus also receives axon terminals of Golgi cells (Fig. 10.2-4).

NEURONAL ACTIVITY OF INTRINSIC CEREBELLAR CIRCUITRY

As a result of excitatory input from the climbing fibres or mossy fibres following activity is set up in the intrinsic cerebellar circuitry.

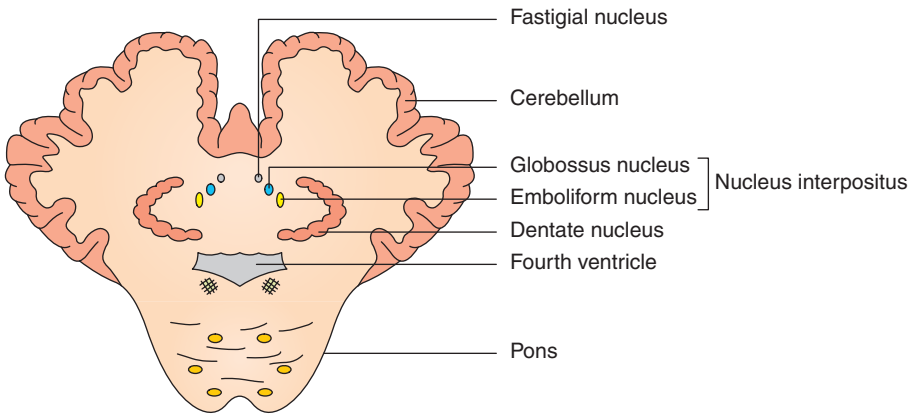


Fig. 10.2-3 Diagrammatic view of nuclei of cerebellum in relation to roof of fourth ventricle.

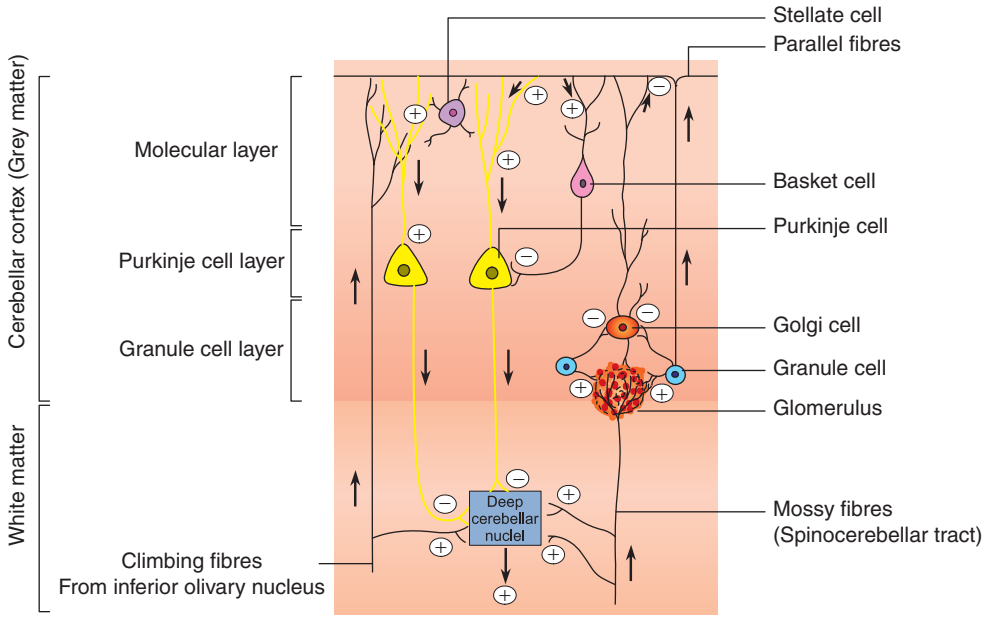


Fig. 10.2-4 Intrinsic cerebellar circuitry.

Feed forward inhibition of Purkinje cells

Granule cells, which are activated by mossy fibres, in turn excite the Purkinje cells. However, this excitation is extremely short lasting. This is because the granule cell also excites basket cell which in turn produces inhibitory post-synaptic potential (IPSP) in the Purkinje cells shortly after stimulation (Fig. 10.2-4). Since Purkinje cell and basket cell are excited by the same excitatory input, this arrangement is called feed forward inhibition. This mechanism helps to limit the duration of excitation produced by given afferent impulses.

Feed forward inhibition of granule cells

As shown in Fig. 10.2-4, the mossy fibres stimulate the granule cells. However, this excitation is short lasting. This is because the mossy fibre also excites Golgi cell which in turn inhibits the granule cell. Since the granule cell and Golgi cells are excited by the same excitatory input (from mossy fibres), this arrangement is said to produce feed forward inhibition of the granule cells.

Feedback inhibition of granule cells

As shown in Fig. 10.2-4, the granule cell is excited by the mossy fibres. The axon of granule cell excites the Golgi cell dendrites, whose axon inhibits the granule cell. Thus excitation of the granule cell is rapidly extinguished by a negative feedback loop. This arrangement is called feedback inhibition of granule cells.

Note. From the above description, it is clear that at least three of the identifiable neural circuits inside the cerebellum are meant for ensuring that discharge of the granule cells and Purkinje cells are extremely precise and short lasting (neural sharpening). This helps in accurate timing of action potentials.

The reverberating circuit

The granule cells and Purkinje cells form a reverberating (echoing) circuit. The main function of the reverberating circuit is to revive and strengthen the non-functional synapses, when two neurons discharge by repeatedly and synchronously. Hebb enunciated this principle, first of all it can be understood by following example.

Suppose a person is making alternate supinations and pronations of his hand rhythmically. This is made possible by the reverberating circuit of cerebellum. Therefore, this capability is impaired in cerebellar disorders and is called adiadochokinesis. This can be explained as,

As shown in Fig. 10.2-5, when the person makes supination movement the signal of descending command from the motor cortex is relayed through the inferior olive to Purkinje cell (PC-1) through the climbing fibres (CF-1). The stimulated Purkinje cell (PC-1) sends a signal to the muscles of supination.

When the muscles of supination contract, a set of proprioceptive receptors are stimulated, from where the sensory information is conveyed through the mossy fibre to the granule cell (GrC-1) of the cerebellum. The stimulation of the granule cell (GrC-1) sets up excitation in the parallel fibres. Although parallel fibres make synaptic connection with Purkinje cells, but most of these are not functional. Therefore, excitation of parallel fibres initially does not cause excitation of Purkinje cells. It is important to note that excitation of parallel fibres occurs a little while after the excitation of Purkinje cell (PC-1), which has already explained

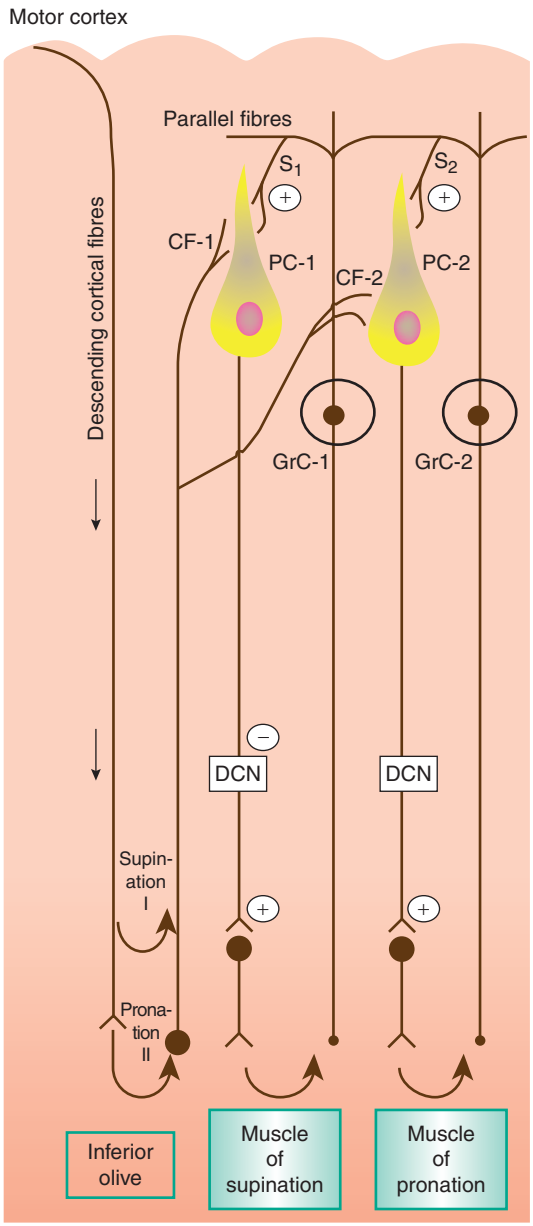


Fig. 10.2-5 Reverberating circuits are shown to illustrate their functioning in supination and pronation movements. (PC = Purkinje cell; CF = climbing fibre; GrC = granule cell; DCN = deep cerebellar nuclei.)

above is extremely short lived. Excitation of parallel fibres a little while after the excitation of Purkinje cell can be compared to an echo which is heard moments after the original sound stops. Because of this similarity, this circuit is called reverberating (echoing) circuit.

Pronation movement (which follows supination) will lead to stimulation of another Purkinje cell (PC-2) through another climbing fibre (CF-2). If the pronation is appropriately timed, the stimulation of PC-2 coincides with the echo of PC-1, i.e. stimulation of parallel fibres of GrC-1.

When supination occurs again, the stimulation of PC-1 coincides with the echo of PC-2, i.e. stimulation of parallel fibres of GrC-2.

When the rhythmic supination and pronation are practised repeatedly, the synaptic connections between PC-1 and parallel fibres of GrC-2; and between PC-2 and parallel fibres of GrC-1 get strengthened. Once this happens, the Purkinje cells, PC-1 and PC-2, get automatically stimulated rhythmically through each other's echo. Under such circumstances, stimulation of Purkinje cell through the climbing fibres is needed only to trigger off the rhythmic sequence.

NEURONAL ACTIVITY OF DEEP CEREBELLAR NUCLEI

- The deep cerebellar nuclei are made of excitatory neurons. These nuclei receive excitatory inputs via collaterals from the mossy fibres, climbing fibres and also other excitatory inputs.
- These sensory inputs maintain the deep cerebellar nuclei in a continuously excitatory state.
- In turn, these nuclei send excitatory impulses to thalamus and brain stem nuclei.
- However, Purkinje cells output is inhibitory to deep cerebellar nuclei via inhibitory neurotransmitter GABA. Purkinje cells inhibit the activities of vestibular nuclei also.

Thus, from the above discussion, it is clear that neuronal activity in the cerebellar cortex plays an important role in modulating the excitatory signals of following pathways:

- From deep cerebellar nuclei to thalamus and then to cerebral cortex, and motor pathway to spinal cord, and
- From deep cerebellar nuclei to brain stem nuclei.

Because of this modulating or timing effect, the cerebellar cortex is able to well organize and co-ordinate the different movements of the body.

CONNECTIONS OF THE CEREBELLUM

CEREBELLAR PEDUNCLES

The cerebellum receives afferents from various sources and sends efferents to different targets. The afferents enter and the efferents leave the cerebellum through three pairs of cerebellar peduncles:

1. Inferior cerebellar peduncle

Inferior cerebellar peduncles, also called as restiform bodies, connect the cerebellum to the dorsolateral aspect of medulla. These constitute the main entrance gates of cerebellum as they contain predominantly afferent fibres.

Afferent fibres passing through inferior cerebellar peduncle are:

- Dorsospinocerebellar tract,
- External arcuate fibres,
- Reticulocerebellar tract,

- Olivocerebellar tract and
- Vestibulocerebellar tract.

Efferent fibres leaving through inferior cerebellar peduncle are:

- Cerebellovestibular pathway,
- Cerebelloreticular pathway and
- Cerebello-olivary pathway.

2. Middle cerebellar peduncle

The middle cerebellar peduncle, also known as brachium pontis, connects the cerebellum to the dorsum of pons. Most of the fibres of the middle cerebellar peduncle are commissural fibres, which connect the area of both the halves of cerebellar cortex.

Afferent fibres, which enter through this peduncle, are: cerebro-pontine-cerebellar fibres.

3. Superior cerebellar peduncle

Superior cerebellar peduncles, also called brachium conjunctiva, connect the cerebellum to the back of mid brain. Since most of the efferent fibres travel through this peduncle, so it is also called exit gate of the cerebellum.

Afferent fibres passing through superior cerebellar peduncle include (Fig. 10.2-6):

- Ventral spinocerebellar tract and
- Tectocerebellar tract.

Efferent fibres leaving through superior cerebellar peduncle are:

- Dentorubral fibres,
- Dentothalamic fibres and
- Cerebellar reticular fibres to mid brain.

AFFERENT AND EFFERENT CONNECTIONS

Salient points of afferent and efferent connections of cerebellum are summarized.

Afferent connections

1. Dorsal spinocerebellar tract. It is an uncrossed tract, which arises from the ipsilateral Clarke's column of cells and carries proprioceptive information from the limbs of the same side. It enters through the ipsilateral inferior cerebellar peduncle and is distributed to the spinocerebellum (for details see page 701).

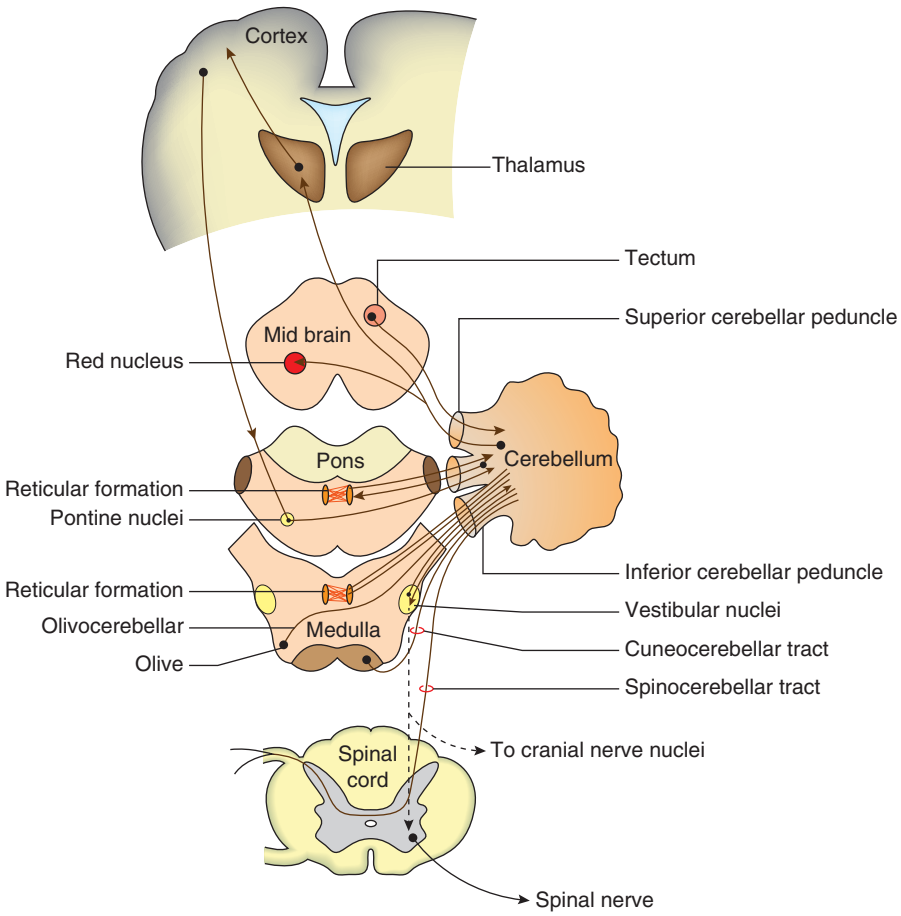


Fig. 10.2-6 Diagrammatic depiction of main connections of cerebellum.

2. Ventral spinocerebellar tract. The fibres of this tract arise from the marginal cells in the dorsal grey horn of spinal cord. After taking origin the fibres cross the midline and ascend in the opposite side. It carries the proprioceptive information from the limbs of the opposite side. It enters the cerebellum through superior cerebellar peduncle and is distributed to the lower limb area of the cortex of spinocerebellum (details on page 701).

3. Cuneocerebellar tract. It arises from the accessory cuneate nucleus and carries proprioceptive impulses from the upper limb, upper trunk and neck. It enters through ipsilateral inferior cerebellar peduncle and is distributed to the spinocerebellum.

4. Olivocerebellar tract. This tract arises from the olivary nucleus and crosses to the opposite side. The olivary nucleus receives afferents from three sources:

- Brain stem nuclei of the same side,
- Spinal cord through spino-olivary tract of same side and
- Cerebral cortex of opposite side.

It carries proprioceptive impulses from the whole body and output signals from the cerebral cortex. It enters the cerebellum through the inferior cerebellar peduncle of opposite side and is distributed to all parts of the cerebellar cortex and the deep cerebellar nuclei.

5. Cortico-ponto-cerebellar tract. The cortico-pontine fibres arise from the cerebral cortex and end in pontine nuclei from where the ponto-cerebellar fibres cross to enter the opposite side through the middle cerebellar peduncle. Cerebral cortex and the cerebellum work in close co-operation in order to affect the proper co-ordination of muscular action in voluntary movements (see page 722).

6. Tectocerebellar tract. It arises from the superior and inferior colliculi of tectum of mid brain and enters the cerebellum via superior cerebellar peduncle. It carries visual impulses from the superior colliculi and auditory impulses from inferior colliculi (see page 697).

7. Vestibulocerebellar tract. It arises from the ipsilateral vestibular nuclei and carries information concerning the position and the movements of head. It enters through the inferior cerebellar peduncle and conveys impulses to vestibulocerebellum (flocculonodular lobe) via deep cerebellar nuclei (nucleus globossus, nucleus emboliformis and nucleus fastigii).

8. Rubrocerebellar tract. It arises from the red nucleus and is both crossed and uncrossed. It enters through the superior cerebellar peduncle and is distributed mainly to the dentate nucleus. It transmits impulses which have originated from the motor cortex and relayed in red nucleus.

9. Reticulocerebellar tract. It arises from the lateral reticular nucleus, enters through the ipsilateral inferior cerebellar peduncle and is distributed to the whole of cerebellar cortex.

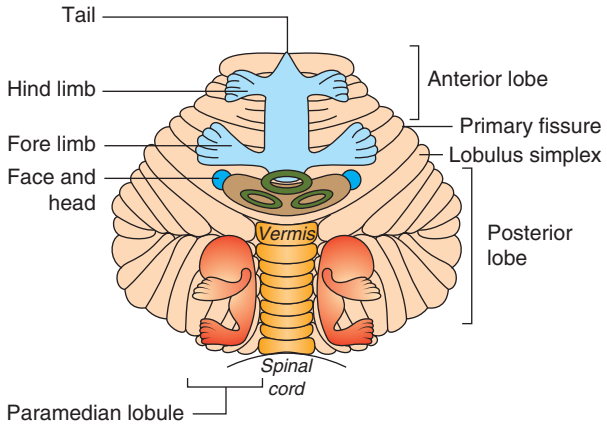


Fig. 10.2-7 Localization of sensory projection areas on the cerebellar cortex.

Localization of the sensory impulses to the cerebellum

Cerebellar cortex, like that of sensory and motor cerebral cortex, exhibits point-to-point representation of sensory impulses (tactile, proprioceptive, visual and auditory) from the whole body. For the purpose of representation, the cerebellar cortex has been divided into three zones:

- *Vermal zone,* i.e. cortex of vermis area (a narrow band in the centre of cerebellum),
- *Paravermal* or intermediate zone, i.e. cortex of medial halves of the cerebellar hemisphere and
- *Lateral zone,* i.e. cortex of lateral halves of the cerebellar hemisphere (Fig. 10.2-7).

Areas of representation. There is a double representation on the superior surface (anterior area) and on the inferior surface (posterior area).

- *Anterior area.* In the anterior area, the body representation is an inverted ipsilateral projection. Axial parts of the body lie in the vermis whereas limbs and facial region lie in the intermediate zone of cerebellar cortex.
- *Posterior area*. In this area, the body representation is a bilateral projection, less defined and is erect.
- There is same topographical representation of motor areas in the cerebellum as is for the sensory areas. Stimulation of these areas produces movements in parts of the body that correspond roughly to those from which sensory impulses are received.
- In addition to the proprioceptive impulses, the cerebellum also receives auditory and visual impulses. The auditory and visual areas lie primarily in the lobulus simplex, folium and the tuber vermis (Fig. 10.2-7).
- The large lateral zones of cerebellar cortex do not have topographical representations of the body. These areas

receive signals entirely and exclusively from the cerebral cortex, and premotor areas of frontal cortex, somatosensory and sensory association areas of parietal cortex. These connections play important role in planning and coordinating rapid sequential muscular activities of the body.

Efferent connections (Fig. 10.2-6)

1. Dento-rubro-thalamo-cortical path. It is a multisynaptic path, synapsing at least in the red nucleus and the thalamus. The fibres originate from the dentate nucleus, which is recent in origin and most well developed in man and so the red nucleus, thalamus and cerebral cortex of opposite side.

2. Cerebello-thalamic-cortical path. It has also the same function, i.e. controlling or influence over the opposite motor cortex.

3. Cerebelloreticular path. Anterior horn cells of spinal cord are controlled by the cerebellum through reticulospinal tracts (direct fibres from cerebellum to anterior horn cells do not exist).

4. Cerebellovestibular fibres pass through the inferior cerebellar peduncle to the vestibular nuclei. These efferents control anterior horn cells of the spinal cord through the vestibulospinal tract.

5. Cerebello-olivary fibres reach the inferior olivary nucleus.

6. Some fibres from the cerebellum also reach the nucleus of oculomotor nerve and the tectum.

FUNCTIONS OF CEREBELLUM

CONTROL OF BODY POSTURE AND EQUILIBRIUM

Vestibulocerebellum. It is concerned with control of body posture and equilibrium.

Afferents to cerebellum are concerned with control of body posture and equilibrium include:

- *Vestibulocerebellar tracts,* which carry input from the vestibular nuclei, which convey afferents from the macula of saccule and utricle for static equilibrium and from the ampullary crests of semicircular ducts for kinetic equilibrium.
- *Spinocerebellar and cuneocerebellar tracts* carry feedback about tone of muscles or position of the limbs in space.
- *Reticulocerebellar tracts* bring feedback about activities of extrapyramidal tracts.

Efferents. The flocculonodular lobe and fastigial nuclei project output fibres through inferior peduncle to vestibular and reticular nuclei of brain stem. The vermal cerebellum sends back the information to spinal cord indirectly through fastigial nuclei.

Mechanism of action. The efferents from the cerebellum influence the spinal motor neurons to keep the body posture upright through the vestibulospinal and reticulospinal tracts, and regulate the position of eyes in relation to movements of the head by connecting motor nuclei of extraocular muscles (3rd, 4th and 6th cranial nerves) via medial longitudinal fasciculus.

It is important to note that the cerebellum does necessary corrections for maintaining posture and equilibrium without participation of conscious will and that the corrections made are highly smooth and precise.

CONTROL OF MUSCLE TONE AND STRETCH REFLEXES

Spinocerebellum is mainly concerned with control of muscle tone and anticipatory adjustment of muscle contraction during movement.

Afferents. *Spinocerebellar, cuneocerebellar* and *olivocerebellar* tracts carry proprioceptive and tactile inputs from the limbs, trunk, neck and other parts of the body. These give feedback about tone of muscles or position of limbs and body. Spinocerebellum also receives auditory and visual impulses through tectocerebellar tract. It also receives the cortical impulses via pontine nuclei.

Efferents. The spinocerebellum is projected into the cerebellar nuclei—fastigii, emboliformis and globossus. Fibres from these nuclei pass through fastigiobulbar, cerebelloreticular and cerebello-olivary tracts and ultimately, to relay to the α and γ motor neurons through the reticulospinal and olivospinal tracts.

Mechanism of action. Spinocerebellum regulates the postural reflexes by modifying muscle tone. It facilitates the gamma motor neurons in the spinal cord via cerebello-vestibulo-spinal and cerebello-reticulo-spinal tracts. The γ motor neurons reflexly modify the activity of α motor neurons and thus regulate the muscle tone. Thus cerebellum forms an important site of linkage of α-γ systems responsible for muscle tone.

Proofs. Temporary suppression of anterior lobe activity by surface cooling abolishes discharge from the γ motor neurons resulting in hypotonia and disturbance in posture. This discharge reappears by warming.

CONTROL OF VOLUNTARY MOVEMENTS

Cerebellum is not able to initiate any motor activity, but coordinates movements initiated by the motor cortex. Therefore, lesions of cerebellum are associated not with paralysis but with disturbances in the smoothness of movements.

Control of movements by cerebellum includes regulation of time, rate, range (extent), force and direction of muscular activity.

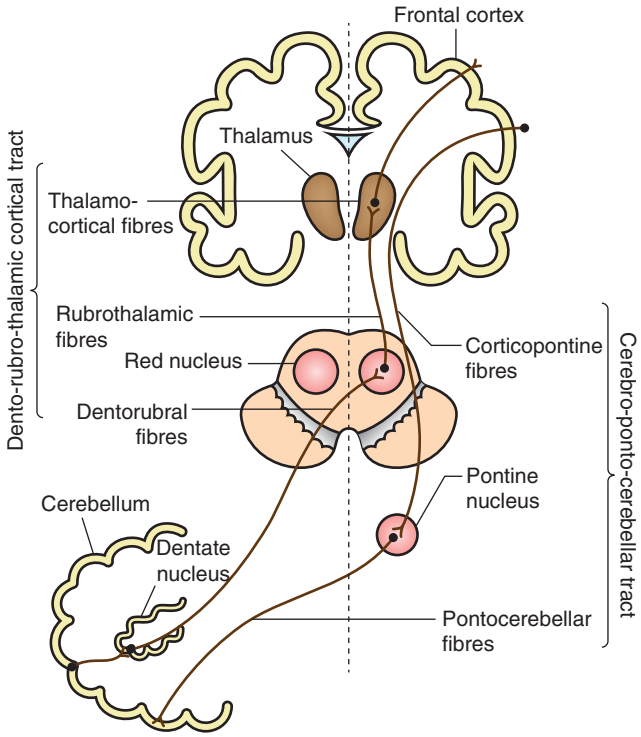


Fig. 10.2-8 Cerebro-cerebello-cerebral circuit (open feedback loop).

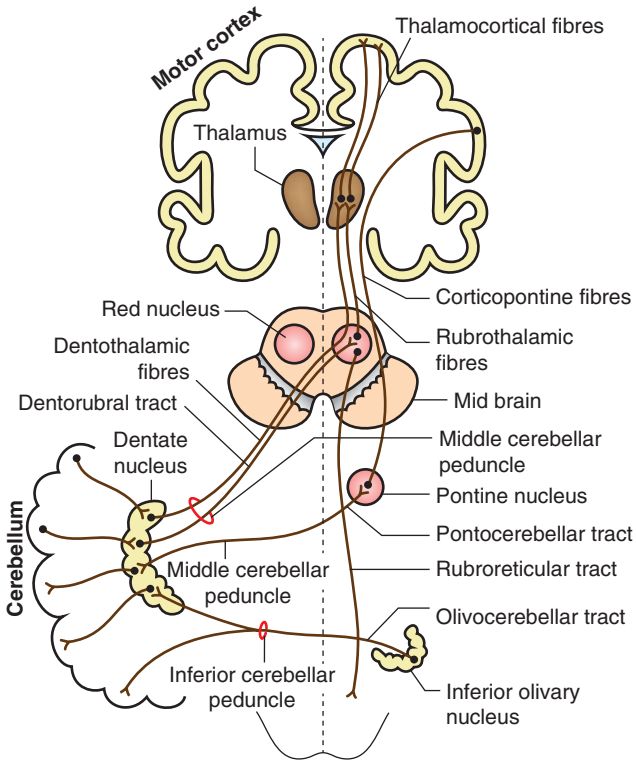


Fig. 10.2-9 Connections of corticocerebellum (closed circuit).

Pathway of control of voluntary movements

Corticocerebellum takes part in smooth performance of highly skilled voluntary movements because of its afferent and efferent connections, which form two feedback loops open and close.

A. Open feedback loop. Also known as cerebro-cerebello-cerebral connection or *afferent efferent circuit* consists of following fibres (Fig. 10.2-8):

1. Cerebro-ponto-cerebellar tract

2. Dento-rubro-thalamic cortical tract, which includes following fibres:

- *Dentorubral fibres* start from the dentate nucleus and pass via superior cerebellar peduncle to end in red nucleus of opposite side.
- *Rubrothalamic fibres* start from the red nucleus and go to thalamus.
- *Thalamocortical fibres* connect the thalamus to area 4 and 6 of motor cortex of cerebrum.

Functions. The cerebro-cerebello-cerebral circuit modulates the motor command of pyramidal tract with a programming of movement.

B. Closed feedback loop. It is formed by the fibres from the cerebral motor cortex to the paravermal cerebellum to the cerebral motor cortex as (Fig. 10.2-9):

Afferent limb is formed by:
Collaterals of corticospinal tracts (which influence the contralateral lower motor neuron of the spinal cord), while descending through brain stem synapse with the ipsilateral pontine nuclei, inferior olivary nucleus and contralateral reticular nucleus of the medulla.

(i) Pontocerebellar fibres from the pontine nucleus and *olivocerebellar* fibres from the inferior olivary nucleus reach the contralateral cerebellar cortex; *reticulocerebellar fibres* from the lateral reticular nucleus are projected to the ipsilateral cortex. All the aforesaid fibres are connected to the paravermal (intermediate) region of cerebellum and provide on their way collaterals to the deep cerebellar nuclei.

(ii) Paravermal cerebellar cortex is in turn connected to the nucleus interpositus and partly to the dentate nucleus.

Efferents from dentate nucleus pass through superior cerebellar peduncle, cross the midline and form decussation with the fibres of opposite side. After forming the decussation, these fibres divide into two groups:

(i) Dentothalamic fibres. Some fibres arising from the dentate nucleus pass through the red nucleus without having any synapse and terminate in the thalamus. Thalamus in turn projects into the motor cortex via thalamocortical fibres.

(ii) Dentorubral fibres. The remaining fibres terminate in the red nucleus of opposite side. From red nucleus, the following tracts arise:

- *Rubrothalamic tract* terminates in thalamus, from where thalamocortical fibres arise and reach the cerebral cortex.
- *Rubroreticular tract* terminates into reticular formation, which projects into spinal cord via reticulospinal tract.
- *Rubrospinal tract.* Red nucleus also projects directly into spinal cord through rubrospinal tract.

Mechanism of action

The cerebellum controls the voluntary movements by following actions:

1. Comparator function. The cerebellum integrates and co-ordinates the patterns of movement involving mostly the distal parts of limbs, especially the hands, fingers and feet by its comparator function.

- When the motor cortex sends impulses through the corticospinal tracts to the lower motor neurons for commanding movements of exploratory nature, it sends messages on the way to the paravermal cerebellum about the sequential intended plan of movements for the next fraction of second.
- The cerebellum also gets feedback from the proprioceptive endings of muscles, tendons and joints about what actual movements result.

The paravermal cerebellum (intermediate zone of cerebellar cortex) then compares the intended movement with the actual movement and through nucleus interpositus sends corrective signals to motor cortex through thalamus and red nucleus. This system of cerebellar *comment* on every command of motor cortex is completed within 10–20 ms.

Thus, the exploratory movement initiated in the motor cortex is corrected by the paracerebellum via the closed loop circuit (Fig. 10.2-9).

2. Damping action. By its comparator action, the cerebellum provides smooth co-ordinate movements of agonist and antagonist muscles of the limbs for the performance of acute purposeful patterned movements. However, all the movements are pendular and have a tendency to overshoot. The corticocerebellum sends impulses to the cerebral cortex to discharge appropriate signals to the muscles so that, any extra or exaggeration of muscular activity does not occur and thus prevents the overshooting. In this way, the movements become smooth and accurate. This action of corticocerebellum is called damping action.

3. Timing and programming the movements. When the perfection of movements is fully assured, the planning of sequence of movements and the timing of the learned movements is then maintained from the association cortex to the motor cortex through lateral zone of cerebellar cortex along with the associated dentate nucleus forming open loop circuit (Fig. 10.2-8).

- *Planning of sequence of movements.* Lateral parts of cerebellar hemisphere communicate with the pre-motor and sensory portion of the cerebral cortex and there is two-way communication between these areas. Plan is transmitted from the cerebral cortex to cerebellum and two-way traffic between two areas provides appropriate transition from one movement to the next. All this plan is stored in the cerebral cortex in the form of memory. So, after the learning process is over these activities are executed easily and smoothly in sequence.
- *Timing function.* Lateral cerebellar hemisphere also provides appropriate timing for each movement, without which succeeding movements may begin far early or too late.
- *Predicting events.* Cerebellum also plays role in predicting events, e.g. rate of progression of auditory and visual phenomena. From changing visual scene person can predict how rapidly he can approach an object.

It is important to note that all fast-skilled movements, such as typing, writing, playing of music instruments etc. is possible because of timing and programming of movements by the cerebellum.

4. Control of ballistic movements. The rapid alternate movements which take place in different parts of the body while doing any skilled work like dancing are called ballistic movements. The cerebellum co-ordinates the action of agonist and antagonist muscles, especially when they occur rhythmically. This is explained under 'reverberating circuit' (see page 718).

5. Servomechanism. From the above discussion it is clear that cerebellum plays an important role in learning of motor skills. Once the skilled works are learnt, the sequential movements could be executed without any interruption. Cerebellum lets the cerebral cortex to discharge the signals, which are already programmed and stored at sensory motor cortex and does not interfere much. However, if there is any disturbance or interference, the cortico-cerebellum immediately influences the cerebral cortex and corrects the movements. This action of corticocerebellum is known as servomechanism.

OTHER FUNCTIONS OF CEREBELLUM

Recent studies have shown that the importance of the cerebellum may extend beyond control of motor activity as:

1. Influence on autonomic system. It has been postulated that the cerebellum may influence autonomic functions.

Cerebellar influence on the autonomic system is probably mediated through the hypothalamus and reticular formation.

2. Influence on conduction in ascending sensory pathway may be exerted by the cerebellum through the reticular formation and thalamus.

3. Control of eyeball movements. The oculomotor, trochlear and abducent nuclei, which supply extraocular muscles of eye movements are brought under the cerebellar control through vestibular nuclei. Medial longitudinal fasciculus is involved in these connections.

CEREBELLAR LESIONS

SIGNS OF CEREBELLAR DYSFUNCTIONS

Common signs observed in patients with cerebellar dysfunction due to lesions of cerebellum are:

Disturbances in tone and posture

1. Atonia or hypotonia. Hypotonia refers to the reduction and atonia to loss of tone in muscles. It occurs due to reduction of the facilitatory neocerebellar output to the descending inhibitory reticular formation.

2. Attitude changes in the unilateral lesions of the cerebellum are:

- *Rotation of the face towards opposite side* (pulled by the healthy muscles).
- Lowering of the shoulder on the affected side.
- Outward rotation and abduction of the leg on involved side.
- Trunk is bent with concavity towards the affected side; this is because the weight of the body is thrown on the unaffected leg.

3. Deviation movement. The arm held straight out in front of the body, deviates laterally when the eyes are closed. In bilateral lesions both arms deviate.

4. Effect on deep reflexes. The deep or tendon reflexes become weak and pendular. For example, *pendular knee jerk* in which after the initial reflex response the leg shakes several to and fro movements before it comes to rest. It occurs due to hypotonia of the quadriceps muscle.

Disturbances in equilibrium

The patient suffering from disturbance of equilibrium, walks on a wide base, sways from side to side (drunken like gait), and is unable to maintain the upright posture and falls on closing the eyes (Romberg's sign) due to involvement of vestibular system.

Disturbances in movements

1. Ataxia, i.e. lack of co-ordination of movements, is the hallmark of cerebellar disorder. It is characterized by:

- *Decomposition of movements,* i.e. the movements seem to occur in stages at different joints.
- *Asynergia,* i.e. lack of co-ordination between the protagonist, synergist and antagonist muscles.
- *Dysmetria,* i.e. movements are incorrect in range, direction and force. The movements may overshoot their intended mark (hypermetria) or fall short of it (hypometria).

2. Intention tremors become evident during purposeful movements, and diminish or disappear with rest. These tremors become more marked as the hand approaches the object (i.e. are observed at the end of movement) and are coarse, oscillating, to and fro and rhythmic. These are significantly observed when the efferent pathways of superior cerebellar peduncles are involved.

3. Nystagmus refers to the regular and rhythmic to and fro involuntary oscillatory movements of the eyes, occurring due to inco-ordination of extraocular muscles. Cerebellar nystagmus occurs during damage to flocculonodular lobes and occurs at rest (when neither the person nor the visual scene is moving).

4. Dysarthria or scanning speech occurs due to in co-ordination of various muscles and structures involved in speech. The speech is slurred, prolonged and explosive with pauses at wrong places.

5. Astasia refers to unsteady voluntary movements.

Charcot's triad. It is a syndrome characterized by nystagmus, intention tremors and scanning speech seen in disseminated sclerosis causing disturbance of connection of cerebellum with brain stem.

CLINICAL TESTS FOR CEREBELLAR DYSFUNCTIONS

Clinically, cerebellar dysfunctions can be demonstrated by following tests:

1. Finger–nose test. The patient has great difficulty in promptly bringing the finger of outstretched arm to touch the tip of his nose. This is because the intention tremors become more severe as the hand approaches the face.

2. Adiadochokinesia, i.e. the patient is unable to rapidly perform alternating movements, e.g. supination and pronation, of the forearm.

3. Rebound phenomenon. When the patient attempts to do a movement against a resistance, and if the resistance is suddenly removed, the limb moves forcibly in the direction

towards which the effect was made. This is called rebound phenomenon. It is due to the absence of breaking action of antagonistic muscles.

4. Gait test. When the patient is asked to walk on a straight line, he is unable to do so (even with eyes open); he follows a zigzag path.

5. Past pointing, i.e. the movement goes beyond the intended point. This is called overshooting and is a manifestation of dysmetria.

BASAL GANGLIA

PHYSIOLOGICAL ANATOMY

COMPONENTS OF BASAL GANGLIA

According to anatomic definition, basal ganglia are subcortical nuclear masses which include corpus striatum (amygdaloid body, and claustrum). They are so named, as they develop in the basal part of cerebral hemisphere. However, from the physiological viewpoint, the term basal ganglia include:

- corpus striatum,
- subthalamic nucleus (body of Luys) and
- substantia nigra.

Corpus striatum

Corpus striatum (Fig. 10.2-10) comprises subcortical masses of grey matter which are situated in the white core of each cerebral hemisphere. It is divided almost completely by the fibres of internal capsule into two parts:

(i) Caudate nucleus (medial part) and
(ii) Lenticular nucleus (lateral part), which is further subdivided into two parts:
- Putamen (an outer part) and
- Globus pallidus (an inner part).

Phylogenetically and functionally, the corpus striatum can be divided into two parts:

- Neostriatum or striatum and
- Paleostriatum or pallidum.

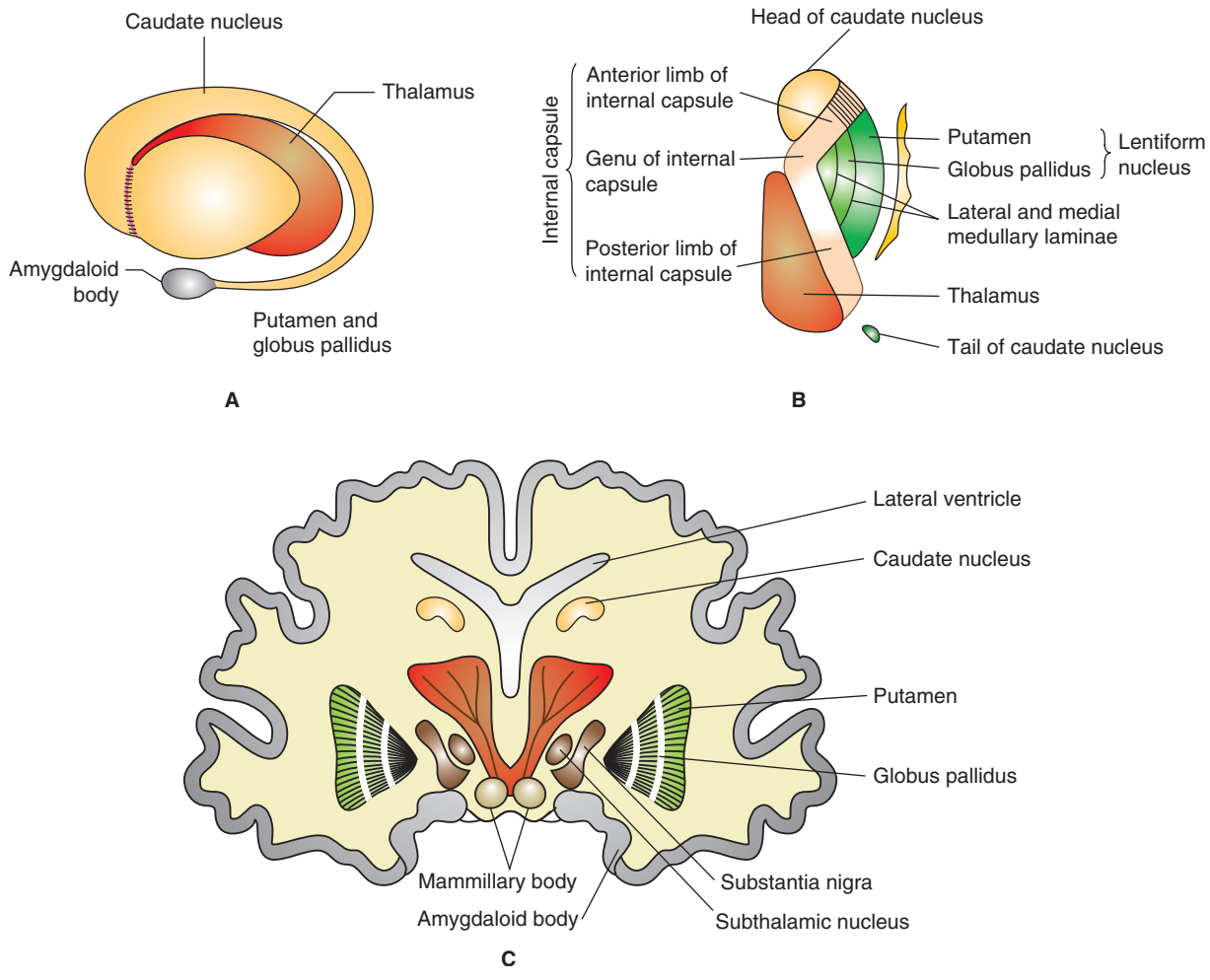


Fig. 10.2-10 Anatomy of basal ganglia: A, lateral view; B, horizontal section and C, frontal section.

1. Neostriatum or striatum. Phylogenetically, the caudate nucleus and putamen are of more recent origin and hence called neostriatum or for short striatum. Functionally and structurally also the caudate nucleus and putamen are similar. The striatum is divided into:

- Dorsal striatum and
- ventral striatum.

2. Paleostriatum refers to globus pallidus, which is older and primitive part. It is also called pallidum, as it is pale (pallid). Pallidum is subdivided into:

- Dorsal pallidum and
- Ventral pallidum.

Salient features of nuclei of corpus striatum (Fig. 10.2-10)

Caudate nucleus. It is a highly curved comma-shaped band of grey matter. It consists of head, body and tail. Caudate nucleus is separated from the lentiform nucleus almost completely by the fibres of internal capsule, except lower part of its head where it is continuous with putamen nucleus (part of lentiform nucleus). This area of continuity is known as *fundus striati*. The tail of caudate nucleus ends by becoming continuous with putamen and lies in close relation to amygdaloid body.

Lenticular nucleus. It is shaped like a biconvex lens and is triangular in both coronal and horizontal sections. It is divided into two parts by an external lamina of white matter:

Putamen is the outer part of lentiform nucleus. It is dark in colour and is roughly quadrilateral.

Globus pallidus is the inner small part, which is paler in appearance. It is further divided by an internal lamina of white matter into:

- External segment (GPe) and
- Internal segment (GPi).

Subthalamic nucleus

Subthalamic nucleus (body of Luys) is a biconvex mass of grey matter, which is situated lateral to red nucleus and dorsal to substantia nigra in the mesencephalon.

Subthalamic nucleus is separated from the ventral nuclei of thalamus by a thin sheet of grey matter known as *zona inserta*.

Substantia nigra

Substantia nigra is a sheet made up of small unpigmented and large pigmented nerve cells. It appears dark in unstained sections as neurons within it contain the pigment neuromelanin. It extends along the entire length of mid brain. Its cranial end reaches close to the subthalamic nucleus. Substantia nigra is divisible into two parts:

1. Pars compacta is the dorsal part of substantia nigra. Pars compacta of the two sides are continuous with each other across the ventral tegmentum. It contains two types of neurons:

- Dopaminergic neurons constitute about 75% and
- Cholinergic neurons are about 25%.

2. Pars reticularis is the ventral part of substantia nigra. Superiorly, it becomes continuous with the globus pallidus. Most of the neurons in the pars reticularis are GABAergic.

CONNECTIONS OF BASAL GANGLIA

- Striatum (caudate nucleus and putamen) forms the main input side of the basal ganglia.
- Striatum in turn projects mainly to globus pallidus and substantia nigra.
- Pallidum (globus pallidus) is the main output side of the basal ganglia.

Therefore, connections of the basal ganglia (Fig. 10.2-11) can be considered under following three headings:

(i) Afferents or input to striatum,
(ii) Projections from striatum and
(iii) Efferents or output from globus pallidus.

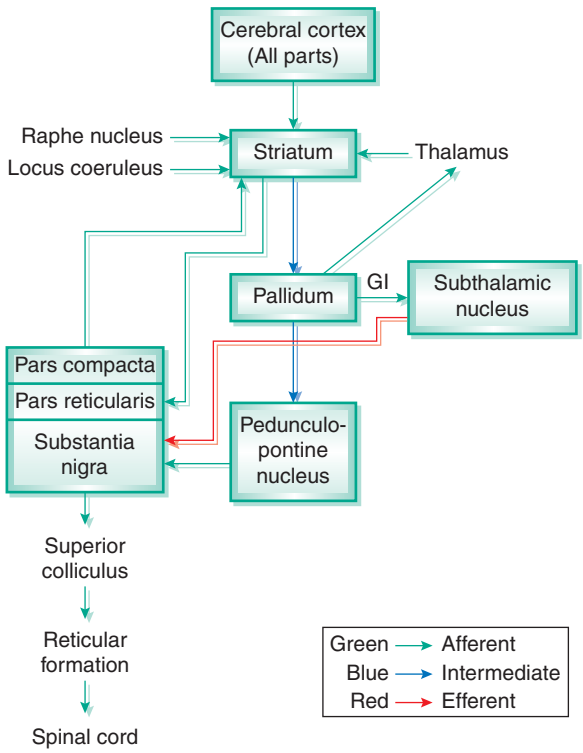


Fig. 10.2-11 Connections of basal ganglia.

Afferents or input to striatum

The striatum (caudate nucleus and putamen) is regarded as the input side of the basal ganglia receiving following afferents.

1. Corticostriate projections. These originate from all parts of the cerebral cortex (pre-motor, supplementary motor cortex and primary somatosensory) and terminate in striatum. These fibres are glutamatergic.

2. Thalamostriate fibres. These originate from the centromedian nucleus of thalamus and terminate in striatum.

3. Nigrostriate fibres. These originate from the pars compacta part of substantia nigra and terminate in the striatum. These are dopaminergic fibres. They are distributed in a typically ordered manner.

4. Raphe striate fibres are serotoninergic fibres received by the striatum from raphe nuclei in the reticular formation of brain stem.

5. Locus coeruleus striate fibres are noradrenergic fibres received by the striatum from the locus coeruleus.

Projections from striatum

1. Striatum to globus pallidus. Striatum (caudate nucleus and putamen), which receive most of the afferents gives robust projection to both segments of globus pallidus. These are GABAergic inhibitory projections.

2. Striatum to substantia nigra. Striatum also gives GABAegric inhibitory impulses to pars reticulata of the substantia nigra.

Efferents or output from globus pallidus

The pallidum (globus pallidus) is the output side of basal ganglia. The efferents of pallidum are as follows:

1. Efferents to thalamus. These fibres are called thalamic fasciculus or *ansa fascicularis*. They arise from the internal segment of globus pallidus (GP_1) and go to ventroanterior, ventrolateral and centromedian nuclei of thalamus. From thalamus fibres project on to prefrontal and premotor cortex.

2. Efferents to subthalamic nucleus, which in turn project to substantia nigra.

3. Efferents to substantia nigra. The pallidum projects to the substantia nigra. These fibres take three routes:

- Some reach the substantia nigra directly,
- Others go via subthalamic nucleus and
- Still others via pedunculopontine nucleus.

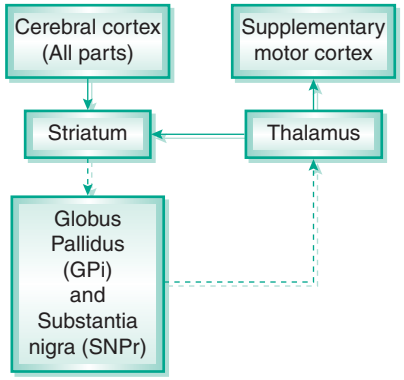


Fig. 10.2-12 Pathway of primary feedback loop (cortex-basal ganglia–cortex neuronal circuit).

Substantia nigra, in turn sends following descending projections:

(i) Substantia nigra brain stem reticular formation-reticulo-spinal tract pathway.
(ii) Substantia nigra superior colliculus-tectospinal tract pathway.
(iii) Substantia nigra-habenula.

4. Efferents to red nucleus. This pathway includes fibres from globus pallidus-red nucleus-rubrospinal tract pathway.

FUNCTIONAL NEURONAL CIRCUITS OR LOOPS

The functional neuronal loops can be grouped as:

1. Primary feedback loop and
2. Additional feedback loop.

1. Primary feedback loop or cortex-basal ganglia-motor cortex circuit

The primary functional neuronal circuit or loop is formed by (Fig. 10.2-12):

- Afferents from all parts of cerebral cortex to striatum (excitatory glutamatergic).
- Projection of striatum to globus pallidus and substantia nigra (GPi and SNpr). (GABAergic inhibitory).
- Efferents from GPi and SNpr to thalamus (GABAergic inhibitory).
- Projections from thalamus to motor cortex and striatum.

Functions. Cortex-basal ganglia-cortex neuronal circuit provides a negative feedback loop to control the activity of motor cortex.

Parts. The primary feedback loop (cortex-basal ganglia-cortex neuronal circuit) consists of two parts, i.e. two distinct loops built into it:

- Caudate loop and
- Putamen loop.

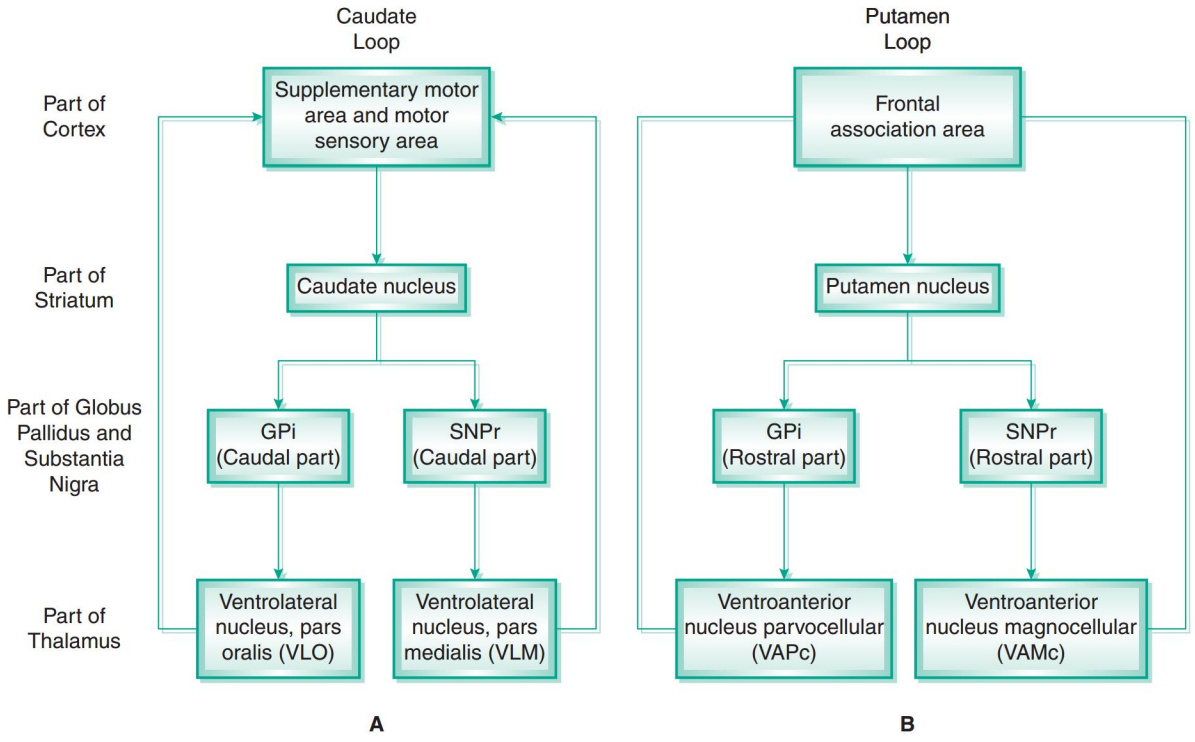


Fig. 10.2-13 A, Pathway of caudate loop and B, pathway of putamen loop.

(i) Caudate loop or circuit is shown in Fig. 10.2-13A. This loop passes through caudate part of pallidum (GPi) and caudal part of substantia nigra (SNpr) and ventrolateral parts of thalamus.

Functions. Caudate loop plays role in cognitive control of motor activity (thinking process of brain). It also plays role in the control of eye movements.

(ii) Putamen loop or circuit. As shown in Fig. 10.2-13B, the putamen loop passes through rostral part of GPi and SNpr, and ventroanterior part of thalamus.

Function. Putamen loop is mainly responsible for the motor control of body movements.

FUNCTIONS OF BASAL GANGLIA

Control of voluntary motor activity

Basal ganglia control the voluntary movements which are initiated by the motor cortex. Role of basal ganglia in control of voluntary motor activity includes:

- Cognitive control of motor activity,
- Timing and scaling of intensity of movements and
- Subconscious execution of some movements.

1. Cognitive control of motor activity. Physiological studies have shown that neural discharge in basal ganglia, like cerebellum, begins well before the movements begin. Therefore, it is believed that basal ganglia, like the cerebellum, are involved in the planning and programming of the movement (Fig. 10.2-14).

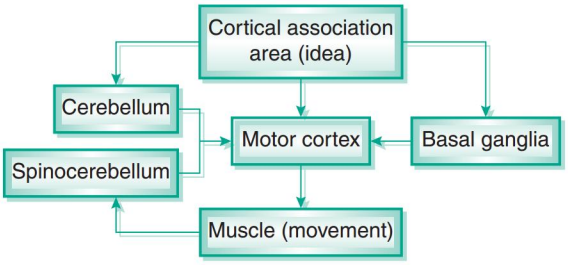


Fig. 10.2-14 Planning and programming of movement.

Most of the motor actions occur as a consequence of thoughts generated in mind. This process is known as cognitive control of motor activity.

Pathway. The cognitive control of motor activity is executed by the basal ganglia through the feedback loops (functional neuronal circuit). As described above, the caudate loop is primarily involved in the cognitive control of motor activity.

2. Timing and scaling of the intensity of movements. Two important capabilities of brain in controlling the movements are:

- Timing of the movements, i.e. how rapidly the movements should be performed and

- Scaling of the intensity of movements, i.e. how large the movement should be.

In higher animals, the basal ganglia act as an important co-ordinating centre of extrapyramidal system. In the absence of basal ganglia, the timing and scaling function becomes very poor.

3. Subconscious execution of some movements. Basal ganglia subconsciously execute some movements during the performance of trained motor activities, i.e. skilled activities. Examples of movements executed sub-consciously at the level of basal ganglia are:

- Swinging of arm while walking,
- Crude movement of facial expression that accompany emotions,
- Movements of limbs while swimming.

Control of clutch and brake while driving (constant attention is required during initial stages; however, they are carried out subconsciously by basal ganglia as they become routine).

Importance. By subconscious control of activities, the basal ganglia relieve cortex from routine acts so that cortex can be free to plan its actions.

Pathway. As described on page 729, the putamen feedback circuit is concerned with control of subconscious execution of some movements during the performance of trained motor activities as listed above.

Control of reflex muscular activity

The basal ganglia exert inhibitory effect on spinal reflexes and regulate activity of muscles which maintain posture. Visual and labyrinthine reflexes are important in the maintenance of posture. The co-ordination and integration of impulses for these activities depend upon basal ganglia.

Control of muscle tone

Muscle spindles and the gamma motor neurons of spinal cord (which are responsible for maintaining the tone of the muscles) are controlled by basal ganglia, especially substantia nigra.

Pathway includes projection from cortical inhibitory area-striatum-pallidum-substantia nigra-reticular formation-spinal cord.

Proof. In lesion of basal ganglia muscle tone increases. Rigidity (lead-pipe type) is a characteristic feature of Parkinson's disease.

Role in arousal mechanism

Globus pallidus and red nucleus are involved in arousal mechanism because of their connections with reticular formation. Extensive lesions in globus pallidus are associated with drowsiness, leading to sleep.

DISORDERS OF BASAL GANGLIA

PARKINSON'S DISEASE

Parkinson's disease, also called *paralysis agitans* or shaking palsy, was first described by James Parkinson in 1817.

Aetiopathogenesis

Primary idiopathic condition. Parkinson's disease occurs in elderly people due to idiopathic degeneration of nigrostriatal system of dopaminergic neurons. There is a steady loss of dopamine and dopamine receptors with age in the basal ganglia in normal individuals; however, it is markedly precipitated in individuals developing Parkinson's disease.

Secondary causes. In addition to the primary idiopathic degeneration of substantia nigra, features similar to Parkinson's disease can occur in some other conditions. The term *Parkinsonism nigra* is used to denote such a condition, which may occur due to following causes:

- Viral encephalitis,
- Cerebral arteriosclerosis,
- Complication of certain drugs (e.g. phenothiazine) which block dopamine (D_2) receptors
- Experimentally, parkinsonism can be produced acutely by injection of the drug MPTP (methyl-phenyl-tetrahydro-pyridine).

Pathogenesis. A current view of the pathogenesis of Parkinson's disease is that there is an imbalance between excitation and inhibition in the basal ganglia created by the loss of the dopaminergic inhibition of the putamen (Fig. 10.2-15). The resulting increase in inhibitory output to the external segment of the globus pallidus decreases inhibitory output from the subthalamic nucleus, and this increases the excitatory output from this nucleus to the internal segment of globus pallidus. This in turn increases the inhibitory output from this segment to the thalamus, causing a reduction in excitatory drive to the cerebral cortex.

Clinical features

Parkinson's disease has both hypokinetic and hyperkinetic features. Its cardinal features are a triad of akinesia, rigidity and tremor; of which akinesia is a hypokinetic feature while rigidity and tremors are hyperkinetic features.

1. Akinesia or hypokinesia

The patient is unable to initiate the voluntary movements (akinesia) or the voluntary movements are decreased (hypokinesia).

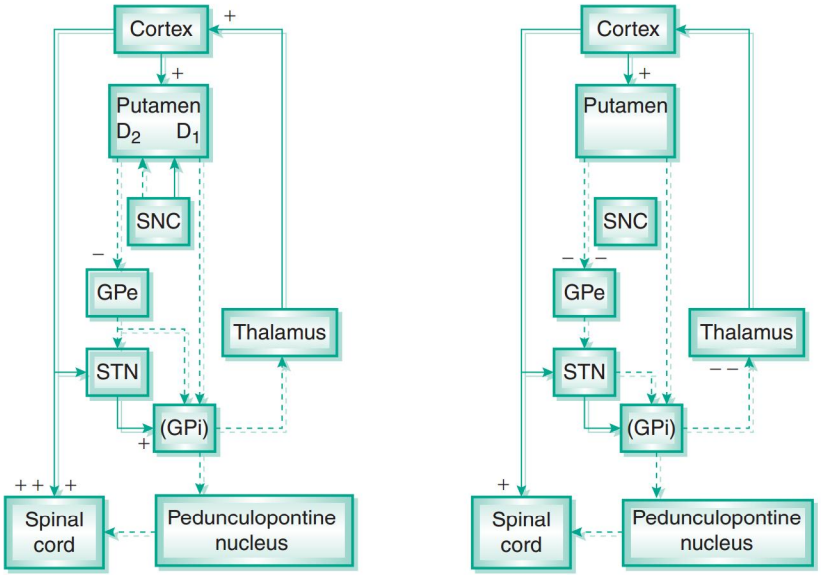


Fig. 10.2-15 Basal ganglia-thalamo-cortical circuitry: A, in normal; and B, in Parkinson's disease. Solid arrows indicate excitatory inputs and dashed arrows indicate inhibitory outputs. Number of plus (+) and minus (–) signs indicate relative increase and decrease in excitation and inhibition of outputs.

Causes. Akinesia is not due to any paralysis or decrease in muscle power, sensory system is also normal. Difficulty in initiating voluntary movements is because of hypertonicity of the muscles.

Manifestations of akinesia or hypokinesia include:

- Delayed motor initiative, as evidenced by prolonged reaction time.
- Slow performance of voluntary movements (bradykinesia).
- Mask-like facial expression due to decrease in movements of facial muscles.
- Absence of normal associated movements, e.g. swinging of arms during walking.
- Shuffling or festinant-type gait, in which patient is bent forward and walks quickly with short steps as if trying to catch up centre of gravity or preventing himself from falling.
- Retropulsion, i.e. when a walking patient is suddenly pulled backwards, he begins to walk backwards and is unable to stop.

2. Rigidity

It refers to an increase in tone of the muscles.

Characteristic features of rigidity occurring in Parkinson's disease are:

- It occurs due to an increased tone in both the protagonists and antagonist muscles.
- Mainly large proximal group of muscles of limbs, e.g. biceps and knee flexors are affected.
- Usually there occurs uniform resistance to flexion giving a feeling as if lead pipe is being bent (lead-pipe rigidity).
- Sometimes, there is a series of catches during passive motion of the limbs (cogwheel rigidity).
- Due to rigidity, posture becomes that of flexion attitude in which: back is flexed, arms are abducted and flexed and the knees are bent.
- In advanced cases, the rigidity may increase to such an extent that a statue-like appearance is produced with complete absence of movements.
- Rigidity differs from spasticity seen in lesions of pyramidal tracts (see Table 10.2-2).

Cause of rigidity. An increased discharge of γ-efferents supplying the muscle spindle causes rigidity. This fact is confirmed by the observation that local injection of 1% procaine solution into the affected muscles decreases rigidity by abolishing the γ-discharge.

Neural mechanism. In patients with Parkinson's disease, lack of dopaminergic activity due to degeneration of neurons in the substantia nigra shifts the balance towards excitatory cholinergic fibres. As a result, hyperkinetic features of Parkinson's disease appear.

3. Tremors

Tremors, (i.e. involuntary rhythmic oscillatory movements of the distal parts of limbs and head) seen in Parkinson's disease have following characteristics:

- The tremors are present at rest, but disappear during activity. It is hallmark of Parkinson's disease and so popularly known as resting (static) tremors.

Table 10.2-2 Differences between spasticity and rigidity

Feature	Spasticity	Rigidity
1. Lesion	Occurs in pyramidal tract lesions, commonest site being internal capsule.	Occurs in basal ganglia lesion, therefore, called the extrapyramidal rigidity.
2. Muscles involved	One group of muscles either agonist or antagonist (usually antigravity muscles) are involved.	Both agonist and antagonist muscles are involved producing a uniform hypertonia often resulting in general attitude of flexion of the limbs and trunk.
3. Characteristics of Hypertonia	Clasp-knife type of hypertonia is seen in muscles involved, i.e. on passive flexion initially there is marked resistance but then there is sudden completion of movement without much resistance (similar to closure of a pocket knife).	Usually there occurs a uniform resistance to flexion giving a feeling as if lead pipe is being bent (lead-pipe rigidity). Sometimes, there is a series of catches during passive motion of the limb (cogwheel rigidity).
4. Relation of hypertonia to stretch	Spasticity is stretch sensitive, i.e. degree of hypertonia developed during any passive stretch is proportional to the speed of stretch applied.	Rigidity is not stretch sensitive.

- Frequency of tremors ranges from 4–6 times/s.
- It is frequently seen as frill-rolling movements of the hand, i.e. rhythmic contraction of thumb over first two fingers.
- Tremors are suppressed during sleep and exaggerated by stress anxiety and excitement.

The tremors are observed as rhythmic movements of pronation and supination in fingers, hands, lips or tongue.

Neural mechanism. The tremors seem to occur due to pacemaker activity in the nucleus ventralis intermedius of the thalamus. Thalamic neurons exhibit an intrinsic autorhythmicity and probably it gets unmasked due to an increase in the inhibitory input from the pallidum. The thalamic pacemaker activity induces oscillation in the long-loop reflex pathways, which originate from muscle spindle. The reflex path runs through the thalamus up to the cortex and then loops back to extrafusal muscle fibres along the corticospinal tract.

Treatment

1. L-dopa is used in the treatment of Parkinson's disease. It can cross the blood–brain barrier and reaches the brain tissue where it is concentrated into dopamine and thus compensates its deficiency.

- Drug dopamine is not used as it cannot cross the blood–brain barrier.
- Along with L-dopa, carbidopa is also used. It prevents the conversion of L-dopa into dopamine in the liver and thus prevents side effects, which can occur due to excessive dopamine content in liver.
- Carbidopa cannot cross blood–brain barrier and thus in the brain L-dopa is converted into dopamine.
- L-dopa in low doses diminishes rigidity and in high doses reduces tremors.

2. Surgical destruction of the globus pallidus or ventrolateral nucleus of thalamus can also ameliorate the symptoms of Parkinson's disease by restoring the output balance towards normal.

CHOREA AND ATHETOSIS

Chorea is characterized by rapid, jerky, involuntary movements (dancing movements). It occurs due to damage to caudate nucleus. Chorea is seen frequently in children as a complication of rheumatic fever.

Athetosis is characterized by slow, rhythmic, twisting, worm-like, confluent writhing movements of the extremities, affecting chiefly the fingers and the wrists. It occurs due to damage to putamen. Athetosis may occur in children following birth injuries.

HUNTINGTON'S DISEASE

Cause. It is a genetic disease of nervous system inherited as an autosomal dominant disorder usually occurring between 30 and 50 years of age.

Site of lesion. There occurs damage to GABAergic and cholinergic neurons of striatum (caudate and putamen) that project to pallidum. The loss of GABAergic pathway to the external pallidum releases inhibition, permitting the hyperkinetic features of the disease to develop.

Characteristic features of Huntington's disease are:

- An early sign is jerky trajectory of the hand when reaching to touch a spot.
- Later, hyperkinetic choreiform movements appear and gradually increase until they incapacitate the patient.
- Speech becomes slurred and then incomprehensible.
- There occurs progressive loss of memory (dementia).

It is a gradually progressive disease, with no effective treatment, which ultimately leads to death.

HEMIBALLISM

Cause. It is a rare disease caused by damage of subthalamic nucleus. Common cause of damage is haemorrhage in the nucleus. Damage to the subthalamic nucleus reduces inhibitory output from GPiSNpc to thalamus. This leads to disinhibition of thalamic output, resulting in hyperkinetic movements mediated by the corticospinal tracts.

Characteristic features. The most important feature of hemiballism is spontaneous attacks of flail-like, intense and violent movements affecting whole of the opposite half of body.

WILSON'S DISEASE

Wilson's disease, also known as hepatolenticular degeneration, is caused by copper toxicity resulting from impaired biliary excretion of dietary copper. Toxic effects are most pronounced in the liver and brain.

- *Liver* involvement is in the form of cirrhosis.
- *In brain,* the lesions are widespread. However, the changes are more marked in the lenticular nucleus, particularly putamen resulting in symptoms of Parkinsonism, i.e. muscular rigidity, tremors and akinesia.

In this condition, the copper content of substantia nigra is high and plasma level of ceruloplasmin (copper binding protein) is low.

KERNICTERUS

Kernicterus refers to the damage of globus pallidus caused by indirect bilirubin, which crosses the blood–brain barrier. It occurs in haemolytic disease of newborn, which results due to Rh antibodies. In this condition, death is very common. However, if the child survives, it may show rigidity, chorea, athetosis and mental deficiency (also see page 169).

Chapter 10.3

Physiological Anatomy, Functions and Lesions of Thalamus and Hypothalamus

THALAMUS
- **Physiological anatomy**
  - External features
  - Internal structure
  - Classification of thalamic nuclei
  - Connections of thalamus
- **Functions of thalamus**
- **Applied aspects**

HYPOTHALAMUS
- **Physiological anatomy**
  - External features
  - Subdivisions and nuclei of hypothalamus
  - Connections of hypothalamus
- **Functions of hypothalamus**
- **Applied aspects**

THALAMUS

PHYSIOLOGICAL ANATOMY

The thalamus proper (i.e. dorsal thalamus) along with the ventral thalamus (old name subthalamus), epithalamus and hypothalamus constitutes the diencephalon. The diencephalon along with the cerebral hemispheres forms the so-called forebrain. It is important to note that the thalamus proper is now called *dorsal thalamus*.

EXTERNAL FEATURES

The dorsal thalamus is a large ovoid structure placed immediately lateral to the third ventricle. It has an anterior and a posterior end and four surfaces viz. dorsal, ventral, medial and lateral.

- *Anterior end* (or pole) lies just behind the interventricular foramen.
- *Posterior end* (or pole) is expanded and is called pulvinar. It lies just above and lateral to the superior colliculus.
- *Dorsal or superior surface* of the thalamus is convex and triangular in outline. It forms the part of floor of the central part of lateral ventricle (Figs 10.3-1 and 10.3-2).
- *Ventral or inferior surface* of the thalamus is related to the hypothalamus anteriorly (Fig. 10.3-1) and to the ventral thalamus posteriorly (Fig. 10.3-2).
- *Medial surface* forms the greater part of the lateral wall of the third ventricle and is lined by ependyma. The medial surfaces of the two thalami are connected by a short bar of grey matter called the *interthalamic adhesion*. Inferiorly, the medial surface is separated from the hypothalamus by *hypothalamic sulcus* (Fig. 10.3-1).
- *Lateral surface* of thalamus is related to the posterior limbs of the internal capsule.

INTERNAL STRUCTURE

Like other parts of brain, thalamus consists of grey matter (mainly) and white matter (Fig. 10.3-3).

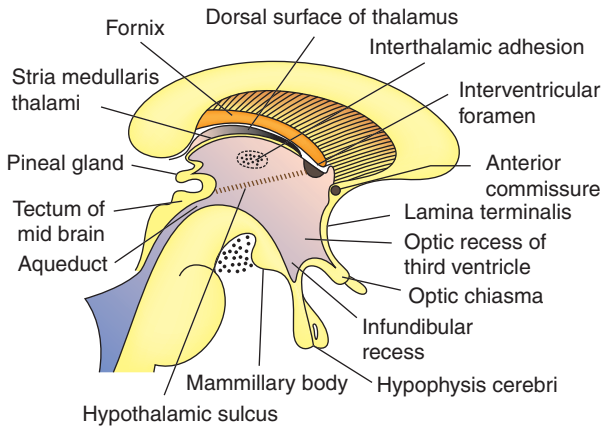


Fig. 10.3-1 Median section through the brain showing medial surface of thalamus and hypothalamus.

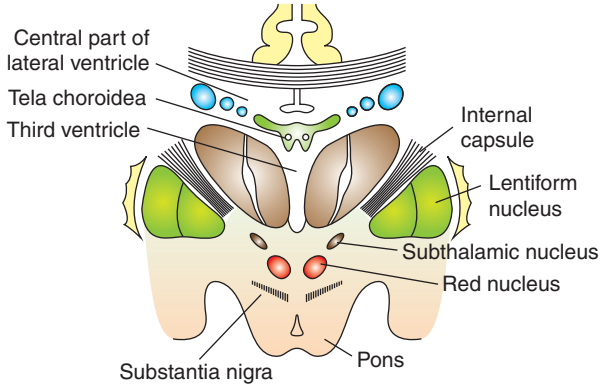


Fig. 10.3-2 Coronal section through the brain passing through the basilar part of pons showing the relations of ventral surface of thalamus and subthalamic structures.

White matter

White matter is scanty in thalamus and includes:

- *Stratum zonale,* a thin layer of white matter covering the superior surface of thalamus.
- *External medullary lamina* is a thin layer of white matter covering the lateral surface of thalamus. It consists of thalamocortical and corticothalamic fibres.
- *Internal medullary lamina* is a Y-shaped sheet of white matter placed vertically in the grey matter of thalamus. It consists mainly of internuclear thalamic connections.

Grey matter

Grey matter of thalamus in divided into three masses of nuclei by the Y-shaped internal medullary lamina (Fig. 10.3-3):

- Anterior part,
- Lateral part and
- Medial part.

CLASSIFICATION OF THALAMIC NUCLEI

Anatomical classification of thalamic nuclei

Anatomically, thalamic nuclei can be classified as (Fig. 10.3-3):

1. Anterior group of nuclei. The mass of grey matter enclosed within the bifurcation of the internal medullary lamina is called *anterior nucleus*.

2. Lateral group of nuclei. The mass of grey matter present in the lateral part of thalamus is subdivided into ventral and dorsal group of nuclei each containing three nuclei:

(i) Ventral group of nuclei includes:

- Ventral anterior nucleus,
- Ventral lateral (lateroventral) nucleus and
- Ventral posterior (posteroventral) nucleus which is further divided into two parts:
  - Ventral posterolateral nucleus and
  - Ventral posteromedial nucleus

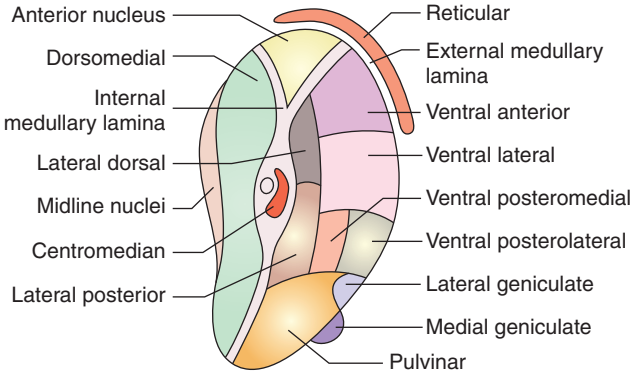


Fig. 10.3-3 Horizontal section of right thalamus (superior aspect) showing nuclear subdivisions.

- Medial and lateral geniculate bodies are present in the posterior zone of ventral groups of nuclei.

(ii) Dorsal group of nuclei are:

- Lateral dorsal nucleus,
- Lateral posterior nucleus and
- Pulvinar.

3. Medial group of nuclei

- Dorsomedial nuclei present in the medial part of thalamus,
- Centromedian nucleus and other interlaminar nuclei present within the internal medullary lamina and
- Midline nuclei that lie between the medial part of thalamus and the ependyma of the third ventricle.

Functional classification of thalamic nuclei

Functionally, the thalamic nuclei can be grouped under two divisions:

- Non-specific projection nuclei and
- Specific projection nuclei.

A. Non-specific projection nuclei

Non-specific projection nuclei are those which receive impulses for diffuse secondary responses from the reticular activating system (RAS) and project diffusely to the whole of neocortex. These include:

- Midline nuclei and
- Centromedian nucleus.

B. Specific projection nuclei

Specific projection nuclei receive specific sensations and project to specific portions of neocortex and limbic system. Depending upon the type of sensation, the specific projection nuclei can be divided into four groups:

I. Specific sensory relay nuclei. These include:

1. Medial geniculate bodies;
2. Lateral geniculate bodies and
3. Posteroventral group of nuclei.

II. Motor control nuclei. These include:
1. Ventrolateral group of nuclei and
2. Ventral anterior nucleus.

III. Visceral efferent control nuclei. These include:
1. Anterior group of nuclei and
2. Dorsomedial nucleus.

IV. Integrative and perceptual function control nuclei.
1. Pulvinar nucleus,
2. Lateral posterior nucleus and
3. Dorsal lateral nucleus.

CONNECTIONS OF THALAMUS

Afferent and efferent connections of the various thalamic nuclei based on their functional classification are as (Fig. 10.3-4):

Connections of non-specific projection nuclei

These are functionally associated with *diffuse thalamic projection,* which produce marked changes in the electrical activity of the cerebral cortex when they are stimulated.

Non-specific projection nuclei include:
- Centromedian nucleus and other intralaminar nuclei, and
- Midline nuclei.

Afferents to these nuclei come from RAS, basal ganglia and other thalamic nuclei.

Efferents from these nuclei project to the stratum and the entire neocortex.

Connections of specific projection nuclei

These nuclei receive specific sensations and project to specific portion of neocortex and limbic system. Depending upon the type of sensation, the specific projection nuclei can be divided into four groups:

I. Specific sensory relay nuclei

1. Medial geniculate bodies

Afferents. Medial geniculate bodies receive a 'topically' organized projection of auditory fibres from the cochlear nerve, lateral lemniscus and also from the inferior colliculi.

Efferents. The medial geniculate bodies (MGB) project on to the auditory area of the cerebral cortex (area 41 and 42). For details see page 929.

APPLIED ASPECT

Destruction of a small part of MGB produces deafness of a particular band of sound frequency.

2. Lateral geniculate bodies

Afferents. Lateral geniculate bodies (LGB) show an orderly organized representation of the retina. They receive projections from the optic tracts from both eyes (temporal fibres of the same side and nasal fibres of the opposite side). They also receive projections from the superior colliculi. In the LGB, the macula is represented in the caudal two-thirds, whereas the remaining retina is represented in the rostral one-third.

Efferents from LGB (optic radiations) project topographically on the visual cortex of the occipital lobe (areas 17, 18 and 19). For details see page 905.

3. Ventral posterior nucleus (Fig. 10.3-4)

Afferents. The ventral posterior nucleus has two divisions: ventral posterior lateral (VPL) and ventral posterior

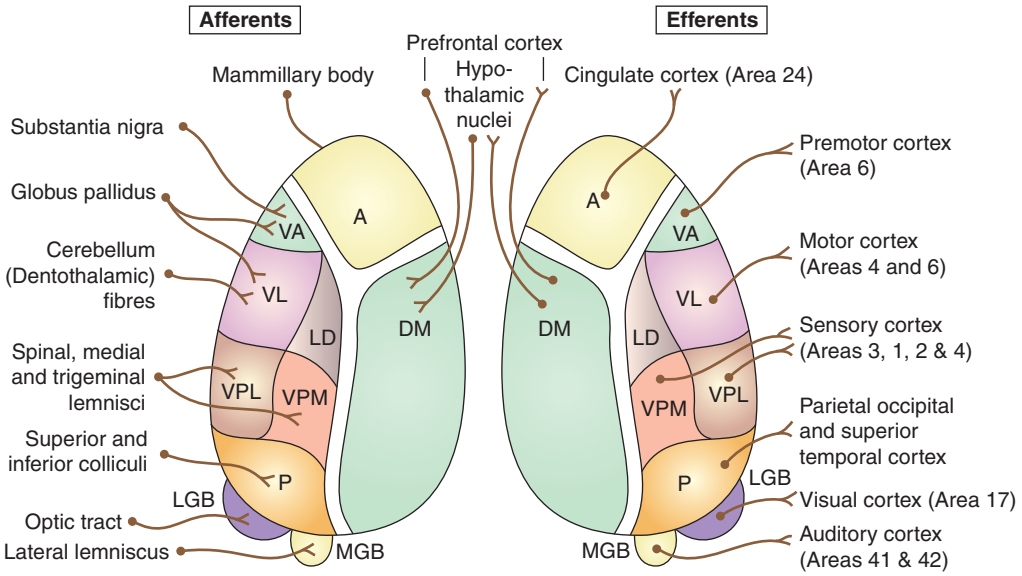


Fig. 10.3-4 Afferent and efferent connections of some of the thalamic nuclei. (A = Anterior nucleus; VA = ventral anterior nucleus; VL = ventral lateral nucleus; VP = ventral posterior nucleus; P = pulvinar; LGB = lateral geniculate body; MGB = medial geniculate body.)

medial (VPM). VPL and VPM are the sites of termination of ascending somatic afferent tracts.

- *The medial lemniscus* carrying afferent fibres from the gracile nucleus, cuneate nucleus and spinothalamic afferents terminate in the VPL. Thus, VPL receives somatosensory impulses (touch-pressure, pain, proprioceptive, temperature and kinaesthetic) from the trunk and limbs, i.e. the whole body except face.
- *The trigeminal lemniscus* carrying afferents from face and taste fibres terminate in the VPM. Thus VPM receives somatosensory impulses from the face along with sensations of taste.
- *In the ventral posterior nucleus,* a topographic representation of the body can be demonstrated.

Efferents. Ventral posterior nucleus is the main sensory nucleus and its efferents go to the sensory cortex, areas 3, 1, 2 (post-central gyrus) via posterior limb of the internal capsule.

II. Motor control nuclei

1. Ventral lateral nucleus

Afferents. Ventrolateral (VL) nucleus is the chief motor nucleus of the thalamus. It acts as a relay station for cerebellar impulses. It receives the dentothalamic fibres from the dentate nucleus of the opposite cerebellar hemisphere. It also receives fibres from the globus pallidus via thalamic fasciculus.

Efferents from VL nucleus project on the primary motor cortex and premotor cortex (areas 4 and 6) via posterior limb of internal capsule.

These relay proprioceptive information and voluntary motor functions.

2. Ventral anterior nucleus (Fig. 10.3-4)

Afferents. The ventral anterior (VA) nucleus is involved in the programming of movements controlled by the basal ganglia. Its afferents come from the globus pallidus, cerebellum and substantia nigra.

Efferents from VA go to the premotor cortex (area 6).

III. Visceral efferent control nuclei

The nuclei concerned with visceral efferent control mechanism are:

1. Anterior nucleus (Fig. 10.3-4)

Afferents. Anterior (A) nucleus belongs to the Papez circuit of limbic system (see page 850). It is concerned with the recent memory and emotions. It receives afferents from the hippocampus directly via fornix and relayed through the mammillary body (mammillothalamic tract).

Efferents from the anterior nucleus go to the cingulate gyrus (area 24) of the cerebral cortex.

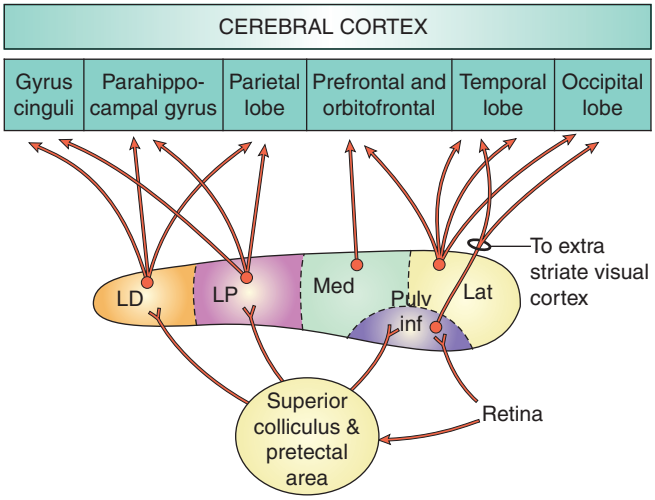


Fig. 10.3-5 Scheme to show the connections of lateral group of thalamic nuclei. (LD = Lateral dorsal; LP = lateral posterior; Med = medial; Pulv = pulvinar; Lat = lateral.)

2. Dorsal medial nucleus

The dorsal medial nucleus has reciprocal connections with the prefrontal cortex and hypothalamus. Its point-to-point interconnections with the prefrontal cortex implying important functions in thinking, memory, judgement and in emotional behaviour.

IV. Integrative and perceptual function control nuclei

1. Pulvinar nucleus (Fig. 10.3-5)

Afferents. Pulvinar nucleus is concerned with integration of visual, auditory and other sensations. It receives afferents from the superior and inferior colliculi.

Efferents go to the parietal, occipital and superior temporal cortex (auditory and visual association areas).

2. Lateral posterior nucleus (Fig. 10.3-5)

Afferents. It receives fibres from the superior colliculus.

Efferents from the lateral posterior nucleus reach the cerebral cortex of the superior parietal lobule. They also reach the cingulate and parahippocampal area.

3. Dorsal lateral nucleus (Fig. 10.3-5)

Afferents. It receives afferents from the superior colliculus.

Efferents. Projections reach the cingulate gyrus, the parahippocampal gyrus and parts of the hippocampal formation. Some fibres reach the cortex of the parietal lobe.

FUNCTIONS OF THALAMUS

1. Sensory relay centre. Almost all the sensory impulses (except olfactory) reach the thalamic nuclei, which relay them to the cerebral cortex by a series of projection fibres collectively termed as the thalamic radiations (ascending

thalamocortical system). Because of this, thalamus is usually considered the head ganglion of all the sensory system.

2. Centre for integration of sensory impulses. The thalamus is not only a great relay station for all sensations, but also forms a major centre for integration and modification of peripheral sensory impulses before the impulses are projected to specific areas of cerebral cortex. This function of thalamus is called processing of sensory information. Because of this, thalamus is usually considered as a functional gateway of cerebral cortex.

3. Crude centre for perception of sensations. In addition to processing and relaying of sensations, thalamus also acts as a crude centre for sense perception. Pain sensation is perceived in the thalamus itself. Usually, the sensations have two qualities: the discriminative nature and the affective nature.

- The *discriminative nature* is the ability to recognize type, location and other details of the sensation. It is the function of cerebral cortex.
- The *affective nature* is the capacity to determine whether a sensation is pleasant or unpleasant and agreeable or disagreeable. It is the function of thalamus.

4. Centre for integration of motor function. Thalamus receives the output from the basal ganglia and the cerebellum before projecting it to the motor cortex, thereby helping in integration of motor functions by unconscious regulation of muscle tone.

5. Role in arousal and alertness reaction. Majority of nonspecific ascending impulses from RAS are relayed to thalamus before proceeding to the cortex. Through these fibres the thalamus is involved in controlling the level of consciousness and maintaining state of alertness and wakefulness.

6. Role in emotional aspect of behaviour. Because of intimate connections between thalamus and frontal cortex and hypothalamus; the thalamus is involved in subjective feeling of various emotions. Thus it acts as a part of limbic system. It also forms a part of the Papez circuit and is concerned with the recent memory and emotions.

7. Role in language. Thalamus is also concerned with language (speech) function. Integration between different cortical parts by subcortical connections in the thalamus helps to achieve speech.

8. Role in synchronization of electroencephalogram. Thalamus also plays an important role in the genesis of synchronization of electroencephalogram.

9. Centre for integration of visceral and somatic function. Thalamus receives somatic as well as autonomic sensations, and is also connected with hypothalamus. Because of this it also acts as a centre for integration of visceral and somatic functions.

10. Centre for sexual sensations. Thalamus also acts as a centre for perception of sexual sensations.

11. Centre for reflex activity. All the sensory fibres relay in thalamus, so it forms the centre for many reflex activities.

APPLIED ASPECTS

THALAMIC SYNDROME

The thalamic syndrome is a disturbance of emotional responses to sensory experience.

Cause. Thalamic syndrome is produced by the damage to posteroventral and posterolateral nuclei as a result of thrombotic blockage of thalamogeniculate branch of posterior cerebral artery.

Symptoms and signs in thalamic syndrome occur on the opposite side of the body. These include:

I. Sensory symptoms due to involvement of posteroventral nucleus are:

1. Astereognosis occurs due to loss of tactile localization, tactile discrimination and stereognosis.
2. Thalamic phantom limb, i.e. patient is unable to locate the position of limbs with closed eyes and searches for the limb in air. This occurs due to loss of kinaesthetic sensations.
3. Thalamic over-reaction, i.e. the threshold for pain, touch and temperature is decreased and the sensations become exaggerated and disagreeable.

II. Motor symptoms due to involvement of posterolateral nucleus are:

1. Ataxia, decreased muscle tone and profound muscular weakness occur due to damage to cerebellar afferents.
2. Involuntary movements, any of the following may be associated:

- Involvement of fibres coming from the globus pallidus leads to chorea (quick jerky movements) or athetosis (slow writhing and twisting movements).
- Intention tremors are usually associated with thalamic syndrome.

3. Thalamic hand or athetoid hand refers to the abnormal posture of hand occurring in patients with thalamic syndrome. It is characterized by moderate flexion of the wrist with hyperextended fingers.

HYPOTHALAMUS

PHYSIOLOGICAL ANATOMY

EXTERNAL FEATURES

The hypothalamus is the most important organ of integration in the homeostatic control of internal environment. It is a bilateral diencephalic structure, diffuse nuclear mass situated below the thalamus.

Boundaries of hypothalamus are (Fig. 10.3-1):

- *Superiorly,* hypothalamic sulcus separates it from the thalamus,
- *Inferiorly,* it is related to the structures in the floor of third ventricle, viz. tuber cinereum, infundibulum and the mammillary bodies, which are considered its parts.
- *Medially,* it forms part of the wall of third ventricle.
- *Laterally,* it is in contact with the internal capsule.
- *Anteriorly,* it extends up to anterior commissure and lamina terminalis.
- *Posteriorly,* the hypothalamus merges with the ventral thalamus at a vertical plane just caudal to the mammillary bodies.

SUBDIVISIONS AND NUCLEI OF HYPOTHALAMUS

For convenience of description, the hypothalamus can be divided as:

From medial to lateral into two zones:

- Medial zone and
- Lateral zone.

From anterior to posterior, the hypothalamic nuclear mass is arranged in four regions:

- Preoptic region,
- Supraoptic region,
- Tuberal region and
- Mammillary region.

Hypothalamic nuclei

Nuclear masses of hypothalamus present in different regions of the hypothalamus are (Fig. 10.3-6):

1. Preoptic region. It is located behind the lamina terminalis. It contains preoptic nucleus.
2. Supraoptic region. It lies above the optic chiasma and rostrally continuous with preoptic area. It forms the anterior nucleus group which includes:

- Suprachiasmatic,
- Supraoptic anterior and
- Paraventricular nucleus.

3. Tuberal region. It is the widest region of the hypothalamus and forms the middle nuclear group, which includes dorsomedial, lateral, tuberal, ventromedial and arcuate (infundibular or tuberal) nucleus.
4. Mammillary region. It forms the *posterior nuclear* group, which includes the posterior and mammillary nucleus.

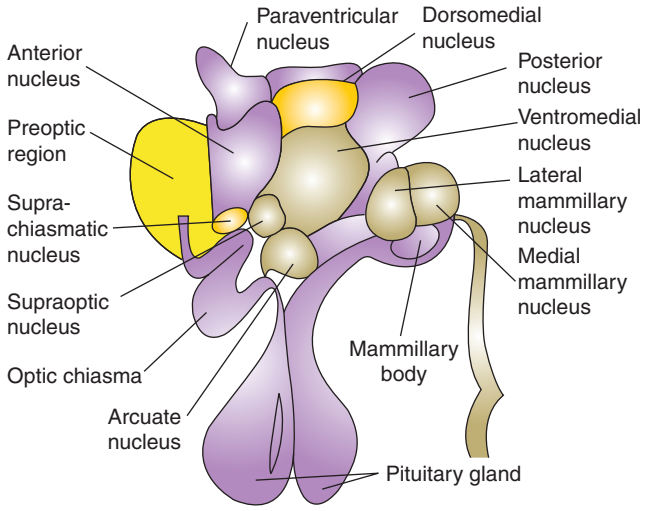


Fig. 10.3-6 Nuclei of hypothalamus.

CONNECTIONS OF HYPOTHALAMUS

The hypothalamus serves as the main integrator of the autonomic nervous system and is concerned with the visceral functions, and is, therefore, connected to other areas having a similar function. These include the various parts of the limbic system, the reticular formation and autonomic centres in the brain stem and spinal cord.

Apart from its neural connections, the hypothalamus also acts by releasing secretions into the blood stream and into the cerebrospinal fluid.

Afferent connections

1. From other parts of limbic system. Hypothalamus receives afferents from other parts of limbic system in the form of following nerve fibre bundles (Figs 10.3-7 and 10.3-8).

- *Medial forebrain bundle* forms the major pathway of the hypothalamus. It consists of both ascending and descending fibres. The descending fibres begin in the anterior olfactory areas (anterior perforated substance, olfactory tubercle and pyriform cortex) and run through the lateral zone of hypothalamus to reach the tegmentum of mid brain. These fibres end in hypothalamic nuclei and raphe nuclei of the reticular formation of the mid brain. These fibres are related to basic emotional drives and to the sense of smell.
- *Fornix* is the main projection for the hippocampal formation and ends in the mammillary body.
- *Stria terminalis* arises from the amygdaloid body, reaches over the thalamus and terminates in the preoptic area and anterior nucleus of hypothalamus.

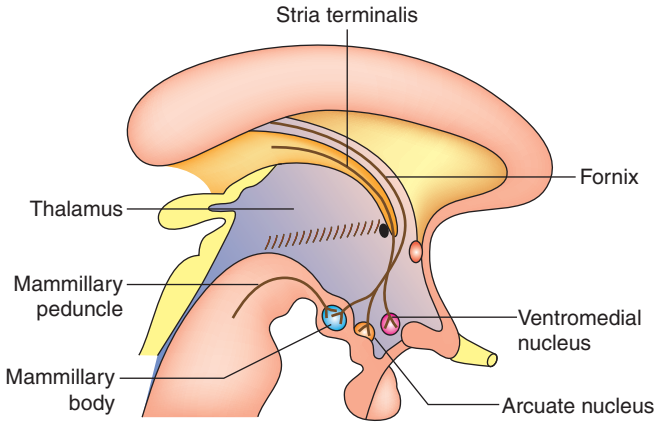


Fig. 10.3-7 Some afferent connections of hypothalamus.

- *Medial hypothalamic tract* runs from the restricted region of hippocampus to the arcuate nucleus. This pathway and stria terminalis are the only two major afferent pathways running directly to medial hypothalamus.

2. From brain stem, the afferents reach to the hypothalamus via following nerve bundles:

- *Mammillary peduncle.* It is a bundle of fibres that connects the tegmentum of the mid brain to the mammillary body. The fibres in it carry gustatory and general visceral impulses from the spinal cord and brain stem centres (nucleus of tractus solitarius and dorsal nucleus of vagus) to the hypothalamus.
- *Dorsal longitudinal fasciculus of Schutz* arises from the periaqueductal grey matter and spreads over dorsal and caudal region of hypothalamus. These fibres also carry visceral impulses to the hypothalamus.
- *Medial forebrain bundle.* The ascending fibres arise from the mid brain and project to the lateral hypothalamic and preoptic nuclei. These fibres also carry visceral impulses to the hypothalamus.
- *Catecholaminergic pathways from the locus coeruleus* ascend monosynaptically to cerebrum and cerebellum. On way to cerebellum they project fibres to thalamic nuclei, hypothalamus, septal area, amygdaloid body and hippocampus. These projections modify the degree of alertness. Ascending catecholaminergic fibres are distributed to supraoptic and paraventricular nuclei, and possibly regulate the output of the releasing hormones of the hypothalamus.
- *Serotoninergic pathways* ascending from the raphe nuclei of the pons and lower mid brain terminate in the hypothalamus, septal nuclei, amygdaloid body and neocortex. Presumably, they regulate the sleep–wake cycle, because total insomnia develops when the serotonin stores are depleted by the use of the drug reserpine.

3. From neocortex. Corticohypothalamic fibres have been described to exist in human beings that interconnect the prefrontal and posterior orbitofrontal regions with preoptic, paraventricular and ventromedial nuclei of hypothalamus.

4. From globus pallidus. The pallidohypothalamic fibres from the globus pallidus go to diffused area of the hypothalamus.

5. From thalamus. The thalamohypothalamic fibres from the dorsomedial and midline nuclei of thalamus go to diffused area of hypothalamus.

6. From retina. The retinohypothalamic fibres are projected from the ganglionic cells of the retina to suprachiasmatic nucleus of hypothalamus through the optic nerve and optic chiasma. This pathway possibly explains the influence of light on the hormonal regulation of reproductive cycle by the hypothalamus.

Efferent connections

Efferents from the hypothalamus go to (Fig. 10.3-8):

1. Autonomic centres. *Posterior longitudinal fasciculus* runs from the autonomic nuclei in the hypothalamus and goes to the autonomic nuclei in the brain stem and spinal cord. Centres in the brain stem receiving such fibres include the nucleus of *solitary tract*, the dorsal nucleus of the vagus, the nucleus ambiguus and the parabrachial nucleus. Fibres descending to the spinal cord end in neurons in the intermediolateral grey column.

2. Other parts of limbic system. The hypothalamic nuclei provide reciprocal connections to other parts of the limbic system mainly through:

- *Stria terminalis,* which connects ventromedial nucleus with amygdaloid nucleus.
- *Medial forebrain bundle,* which connects lateral hypothalamus with septal nuclei, where they relay and then projects to hippocampus.
- *Ventral pathway,* which connects lateral hypothalamus with amygdaloid nucleus.

3. Thalamus. The *mammillothalamic tract* (bundle of Vicq d'Azyr) connects the mammillary body to the anterior nucleus of thalamus, which in turn are connected with cingulated gyrus thus forming a component of Papez circuit of the limbic system. These fibres are responsible for those emotions and aspects of behaviour that are related to preservation of the individual and species.

4. Tegmentum of mid brain. The mammillotegmental tract arises from the mammillary body and terminates in the ventral and dorsal tegmental nuclei of mid brain.

5. Neocortex. Fibres from the hypothalamus project widely to the neocortex. They play a role in maintaining the cortical arousal.

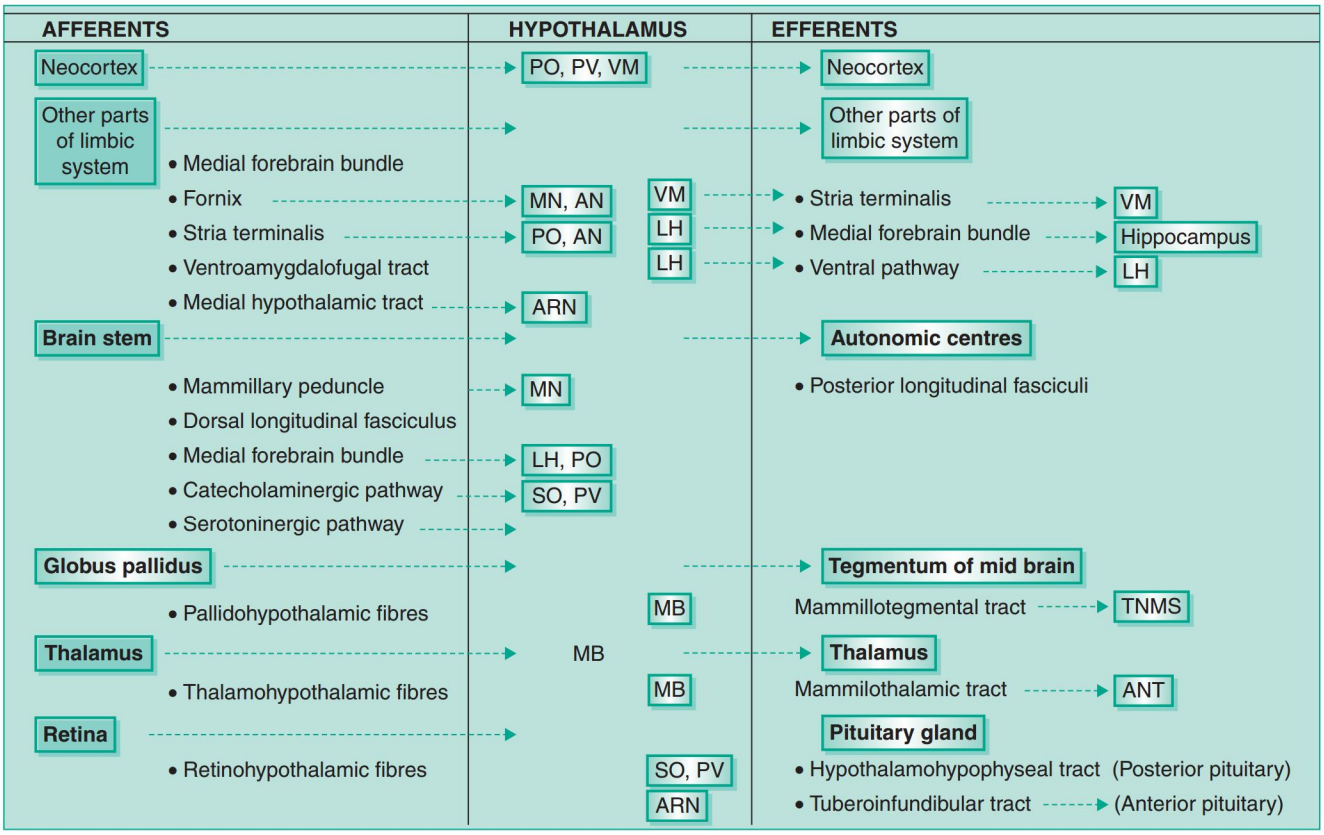


Fig. 10.3-8 Simplified scheme of main afferent and efferent connections of hypothalamus. (PO = Preoptic; PV = paraventricular; VM = ventromedial; MN = medial nucleus; AN = anterior nucleus; LH = lateral hypothalamus; SO = supraoptic; ARN = arcuate nucleus; MB = mammillary body.)

6. Pituitary gland. Influences from the hypothalamus are conveyed to the pituitary gland (hypophysis cerebri) in two different forms:

(i) Hypothalamohypophyseal tract (see page 538). It is composed of axons of the large neurosecretory cells of the supraoptic and paraventricular nuclei of hypothalamus. The fibres of this tract are unmyelinated and pass to posterior lobe of pituitary (neurohypophysis) through the infundibular stem of hypothalamus and form a series of dilated terminals known as *Herring bodies,* which come in contact with the capillary bed of the neurohypophysis. The neurosecretory cells of supraoptic and paraventricular nuclei secrete *vasopressin* (ADH) and *oxytocin*.

(ii) Tuberoinfundibular tract (tuberohypophyseal tract). It consists of fibres arising from the arcuate nuclei of the tuberal region of the hypothalamus, and extends to the median eminence and infundibular stem where the fibres come in close contact with the capillary plexus of the *hypophyseal portal system* (see page 537). The neurons of hypothalamic nuclei synthesize certain *releasing* or *inhibiting factors* (or hormones), which are conveyed as membrane-bound vesicles by the tuberoinfundibular tract to the hypophyseal portal vessels. These hormones are carried by the portal vessels to the anterior pituitary (adenohypophysis) where they regulate the secretion of anterior pituitary hormones.

FUNCTIONS OF HYPOTHALAMUS

Autonomic functions

Sherrington described the hypothalamus as the *head ganglion* of the autonomic nervous system.

- *Anterior hypothalamus* is the parasympathetic area. It has an excitatory effect on this system.
- *Posterior hypothalamus* is the sympathetic area and has an excitatory effect on it.

Autonomic functions subserved by the hypothalamus are:

1. Cardiovascular regulation. Hypothalamus regulates the cardiovascular system through cardiovascular control centres in the reticular regions of medulla and pons.

- Stimulation of posterior and lateral nuclei of hypothalamus increases the heart rate and the arterial blood pressure and produces cutaneous vasoconstriction.

- Stimulation of preoptic area decreases the heart rate and the arterial blood pressure and produces cutaneous vasodilatation.

2. Regulation of pupil size

- Stimulation of posterior and lateral hypothalamus causes dilatation of pupil, while
- Stimulation of anterior and medial parts of preoptic and supraoptic areas produce constriction of pupil.

3. Regulation of peristaltic and secretomotor functions of alimentary tract

- *Stimulation of posterior and lateral hypothalamus* diminishes the secretion and motility of gastrointestinal tract (ergotropic function).
- *Stimulation of anterior and medial hypothalamus* increases peristalsis and secretomotor functions of alimentary tract (trophotropic function).

Endocrinal functions

1. Control of anterior pituitary. The hypothalamus controls the functions of anterior pituitary by secreting certain 'releasing' and 'inhibiting' hormones which reach the anterior pituitary by a neurovascular link through the tuberoinfundibular tract and hypophyseal portal vessels as described on page 537.

Hypothalamus does the following functions through the releasing hormones:

- Controls the metabolism by controlling thyroid gland.
- Through its influence over adrenal cortex, controls the metabolism of different foodstuffs and maintains electrolyte balance.
- Keeps the gonads inhibited till the physical growth is complete. After physical growth is complete this inhibition is removed so that gonads start functioning and gametes are produced (propagation of species). Gonadal hormones acting on the brain bring about physiological changes for mating of male and female.
- Controls the formation of milk by the breasts by controlling prolactin secretion.

2. Regulation of posterior pituitary functions. The hypothalamus regulates the posterior pituitary functions through the hypothalamic–hypophyseal tract (for details see page 538).

Neural control of the posterior pituitary with the secretion of antidiuretic hormone (ADH) by the supraoptic and paraventricular nuclei helps in regulation of water balance by controlling water excretion by kidneys (see page 547).

3. Regulation of uterine contractility and regulation of milk ejection from the breast. Stimulation of paraventricular nucleus of hypothalamus causes its cells to secrete the hormone oxytocin. Oxytocin increases the contractility of uterus. It also contracts the myoepithelial cells that surround the alveoli of breast and cause milk ejection.

Regulation of sleep–wake cycle

The hypothalamus plays an important role in sleep–wake cycle:

- *Anterior hypothalamus* is considered a *sleep facilitatory centre*, as its stimulation leads to sleep.
- *Posterior hypothalamus* acts as *waking centre*, as its stimulation causes wakefulness.
- *Sleep is also considered to occur as a negative phenomenon,* i.e. inhibition of wakefulness centre in the posterior hypothalamus by the anterior hypothalamus also contributes to occurrence of sleep. Lesions in the posterior hypothalamus produce severe coma (for details see page 864).

Control of circadian rhythm

Circadian rhythm refers to the rhythmic fluctuations in certain physiological parameters of the body. These are called circadian rhythms because they often show 24 h cycles (circadian around a day). Many of the rhythms are co-ordinated with each other.

Common rhythmic variations in homeostatic regulatory mechanism are:

- Rhythmic secretion of ACTH (see page 589),
- Rhythmic secretion of growth hormone (see page 540),
- Rhythmic secretion of melatonin (see page 617),
- Sleep–wake cycles (as described above),
- Body temperature rhythm (see page 954) and
- Rhythmic gonadotropin secretion (see page 658).

Basis of circadian rhythm. The circadian rhythms are *internally driven*. The *suprachiasmatic nuclei* of hypothalamus are the main site of most circadian rhythms in the body. These are believed to contain the 'biological clock', which regulates the circadian rhythm according to the 24 h light–dark cycles. The suprachiasmatic nuclei receive important inputs from:

- The eyes via retinohypothalamic fibres (page 617) and
- The lateral geniculate nuclei.

Effect of environmental factors on circadian rhythm. The environmental factors, such as light–dark cycles, temperature, meal timing, etc. only provide *hints* and are required only to set a circadian rhythm cycle of 24 h. Otherwise, the circadian rhythms are internally driven and can occur in the absence of environmental factors as evidenced below:

- Normally, the rats show locomotor activity in the dark (at night) and inactivity in the day time. These cycles of activity and inactivity continue even when the rats are

put permanently in darkened laboratory for a few days with no exposure to light.
- The cycles of activity and inactivity can be disrupted by bilateral lesions of suprachiasmatic nuclei.

Physiological significance of circadian rhythm

- The circadian rhythm enables *homeostatic mechanism* to be utilized immediately and automatically. For example, there is a rhythm in the urinary excretion of ACTH.
- The circadian rhythms have effects on the body's resistance to various drugs. For example, difference in the sensitivity of dose of a potentially lethal drug depends markedly on the time the drug is given.

Disturbances of circadian rhythm can occur during high speed jet travel. One may travel several thousand kilometres within a few hours. As a result, the traveller's external clock (day or night) does not coincide with the internal biological clock. That is, the body may be in rest (night) phase, while it is day time in the country of destination. It results in irritability, mental depression or even physical illness. The symptoms subside in a few days. The condition is called jet lag.

Regulation of food intake

The regulation of food intake is an essential vegetative function of the hypothalamus, which maintains the body weight of an individual relatively constant over a long period. To regulate the food intake, hypothalamus has two centres namely, the *feeding centre* and *satiety centre* located in the tuberal region.

Feeding centre. The lateral hypothalamic nucleus subserves as the feeding centre or hunger centre. When this is stimulated, in animals it creates a sensation of hunger and leads to increased food intake (hyperphagic). This causes obesity. The destruction of feeding centre leads to loss of appetite (anorexia).

Normally, the feeding centre is always active and its activity is inhibited by the satiety centre after food intake.

Satiety centre. Satiety is opposite to hunger, i.e. it is a feeling of fulfilment after food intake. The *ventromedial nucleus* of hypothalamus acts as a satiety centre. Stimulation of this in animals causes sensation of food intake (fulfilment). Destruction of satiety centre leads to hyperphagia. There are following hypothesis regarding regulation of food intake:

- Glucostatic theory
- Lipostatic theory
- Gut peptide theory and
- Thermostatic theory

Glucostatic theory. The cells of satiety centre act as glucoreceptors (also called glucostats), therefore the activity of satiety centre is governed by glucose utilisation of these cells. The satiety centre activity decreases when the glucose supply is inadequate leading to less or no inhibition of feeding centre resulting in its inactivation and individual feels hungry. On the other hand, when there is adequate supply of glucose, satiety centre cells activity increase leading to inhibition of feeding centre and there is feeling of fulfilment.

IMPORTANT NOTE

Polyphagia in diabetes mellitus is explained by the glucostatic theory. There is inadequate glucose utilisation by glucoreceptors of satiety centre (due to deficiency of insulin).

Lipostatic theory. The neurons of feeding centre respond to levels of fatty acids and amino acids. The body fat depots initiate either neural or hormonal signals that are related to the hypothalamus and control the food intake:

- Leptin (Greek word, means thin) is a circulating protein hormone produced by the adipose cells. By its action on hypothalamus it decreases release of Neuropeptide Y resulting in a decrease food intake.

IMPORTANT NOTE

Leptin acts through leptin receptors, mainly present in brown adipose tissue and brain microvasculature. Leptin controls the size of body fat; therefore, obesity occurs due to defective leptin receptor gene (Ob gene).

Gut peptide theory. According to the gut peptide hypothesis, presence of food in gastrointestinal tract (GIT) releases certain polypeptides and GIT hormones (like CCK, Glucagon, GRP, Peptide YY, Somatostatin) that act on the hypothalamus to inhibit food intake. Circulating CCK plays a major role through its receptors (CCK_A and CCK_B) present in the hypothalamus.

Thermostatic theory. Body temperature (core) regulates food intake. Fall in body temperature increases and rise decreases the food intake.

The balanced activity of these two centres is responsible for the normal food intake.

Role of neurotransmitters in food intake. *Food intake is increased* by the stimulation of α_2 adrenergic receptors in medial hypothalamus and centrally acting opioids.

Food intake is decreased by the stimulation of β adrenergic and dopaminergic in lateral hypothalamus and by stimulation of serotonergic pathways.

Role of hypothalamic peptides. Principal hypothalamic polypeptides; (Neuropeptide Y, Orexin-A and Orexin-B,

melanin concentrating hormone (MCH) and Ghrelin) increase the food intake; whereas α MSH, CART (cocaine- and amphetamine-regulated transcript) and CRH decrease food intake.

Regulation of sexual behaviour and reproduction

In animals, hypothalamus plays an important role in maintaining the sexual function, especially in females. A decorticate female animal will have regular oestrous cycle provided the hypothalamus in intact.

A pathway of sex regulation has been identified as amygdala–stria terminalis–preoptic area–tuberal region of hypothalamus. The tuberal region of hypothalamus maintains the basal secretion of gonadotropin releasing hormone (GnRH), and its connection with the preoptic area is essential for the cyclical surge of gonadotropin before ovulation. Electrical stimulation of preoptic area produces ovulation in the experimental animals. Destruction of neural links between the preoptic and tuberal region prevents ovulation.

Role in emotional and instinctual behaviour

The emotional and instinctual behaviour is mainly regulated by the limbic cortex (for details see page 849). The hypothalamus along with the limbic structures is concerned with affective nature of sensory impulses, i.e. whether the sensations are pleasant or unpleasant. These affective qualities are also called a reward and punishment. The two centres in the hypothalamus involved in such a behaviour and emotional changes.

Reward and punishment centres

Reward centre is located along the course of medial forebrain bundle, especially in lateral and ventromedial nucleus of hypothalamus. Electrical stimulation of this area encourages the animal to seek more of such stimulation.

Punishment centre is located in the medial hypothalamus (periventricular zone). The electrical stimulation of this area leads to pain, fear, defence, escape reactions and the other elements of punishment. The experimental animal avoids further stimulation of this area.

Role of reward and punishment centres. Almost anything that we do is related in some way to reward and punishment. If we do something that is rewarding, we continue to do it. If we do something that is punishing we cease to do it.

Therefore, *reward and punishment centres constitute one of the most important of all the controllers of our bodily activities, our drives, our aversions and our motivation.*

Sensory experience that is causing neither reward nor punishment is remembered hardly at all, the animal becomes habituated to such sensory experience and then ignores it. But when the sensory experience causes either reward or punishment, the cortical response becomes progressively more and more intense. Thus *reward and punishment centres help in selecting the information that we learn.*

Rage. Strong stimulation of punishment centres produces a violent and aggressive emotional state called rage. Normally, it is kept in check by counterbalancing activity of ventromedial nuclei of hypothalamus, hippocampus, amygdala and anterior portion of limbic cortex.

Rage reaction is characterized by:

- Development of a defence posture,
- Extension of limbs,
- Lifting of tail,
- Hissing and splitting,
- Piloerection,
- Wide opening of eyes,
- Dilation of pupil and
- Severe savage attack, even on mild provocation.

Sham rage. Normally, the animals and human being maintain a balance between the rage and the opposite state, i.e. calm emotion. This occurs due to the reciprocal connections between the hypothalamus and the cerebral cortex. When the connection between cerebral cortex and hypothalamus is severed by decortication, the experimental animal exhibits outburst of rage on mild peripheral stimulation. This is known as *sham rage*, since the emotions associated with are absent. Thus, sham rage is due to release of hypothalamus from the cortical control, and it can be abolished by lesioning the caudal hypothalamus.

Role in regulation of body temperature

The hypothalamus acts as a principal integrating centre for heat regulation. By adjusting a balance between the heat production and heat loss, it helps to maintain body temperature at 37°C. Hypothalamus accomplishes this function by two centres:

1. Heat loss centre. Anterior hypothalamus, especially preoptic area, acts as a heat loss centre.

- Increase in the temperature of blood flowing through this area increases the activity of temperature-sensitive neurons which results in cutaneous vasodilatation and increased sweating causing more heat loss.
- Lesions of anterior hypothalamus abolish the physiological response to heat exposure.

2. Heat gain centre. The posterior hypothalamus acts as a heat gain centre. Electrical stimulation of posterior hypothalamus results in cutaneous vasoconstriction and shivering.

A lesion of the posterior hypothalamus abolishes not only body response to cold but also to heat as well, because this area is final integration centre for all thermoregulatory signals. Final efferent signals for heat production or heat loss emerge from the posterior hypothalamus.

Regulation of body temperature is discussed in detail in Chapter 12.1 (see page 953).

Role in regulation of water balance

Hypothalamus regulates the water balance of the body by two mechanisms (Fig. 10.3-9):

- *Through thirst centre* by controlling water intake and
- *Through osmoreceptors* in supraoptic nucleus by controlling water loss.

1. Through thirst centre. Thirst centre located in the lateral nucleus of hypothalamus is stimulated by plasma hypertonicity (which occurs when the water content of the body is reduced). This causes intense desire for water and animal drinks large quantities of water. It has been observed that:

- Discrete lesions of thirst centre abolishes fluid intake and the animal dies of dehydration.
- Electrical stimulation of this area in conscious animals (through chronically implanted electrodes) causes the animal to drink water as long as the stimulation continues.
- Injection of hypertonic saline in this area induces the animal to drink large amount of water. But neither isotonic saline, nor distilled water, hypertonic urea caused any drinking. These experiments suggest that the thirst centre monitors the plasma osmolality and is separate from the osmoreceptors involved in ADH release.

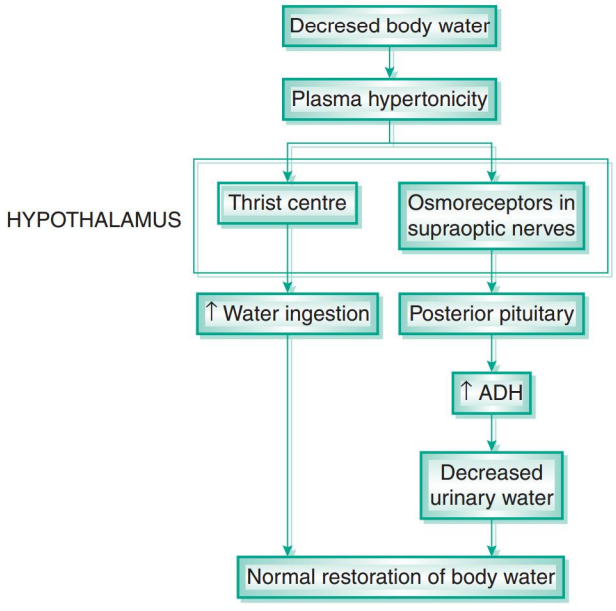


Fig. 10.3-9 Mechanism of regulation of water balance by hypothalamus.

Further, it has been observed that the sensation of thirst is satisfied simply by the act of drinking, even before sufficient water is absorbed from the gastrointestinal tract to correct the plasma osmolality. Oropharyngeal and upper gastrointestinal receptors appear to be involved in this response. However, relief of the thirst sensation via these receptors is short lived. Thirst is completely satisfied only when the plasma osmolality, blood volume and arterial pressure are corrected.

2. Through osmoreceptors in supraoptic nucleus. The increased plasma osmolality also stimulates osmoreceptors in supraoptic nucleus. The stimulated neurons of supraoptic nucleus in turn send impulses to the posterior pituitary gland to secrete hormone ADH. This hormone reaches the kidney tubules through blood and causes increased absorption of water from the collecting ducts of the kidneys. Thus water loss is decreased. When body has excess water exactly opposite events occur.

For further details of water balance control of body fluid osmolality, see page 414.

APPLIED ASPECTS

Lesions of hypothalamus

Lesions of hypothalamus include:

- Tumour,
- Inflammation (or encephalitis),
- Ischaemia due to vascular disorder and
- Damage due to surgical operations in this area.

Disturbances in hypothalamic lesions

Lesions of hypothalamus may result in a variety of disturbances:

- Autonomic disturbances
- Disturbances of body temperature regulation
- Sleep disturbances due to lesions in mammillary body and anterior hypothalamus
- Endocrine abnormalities, e.g. hypogonadism and hypothyroidism
- Disturbance in sexual functions due to involvement of mid hypothalamus
- Disturbance of body water balance due to damage to supraoptic nuclei or infundibular stalk, characterized by excessive thirst and polyuria.
- Emotional disturbances leading to sham rage due to lesions in ventromedial and posterolateral parts.

Clinical conditions in hypothalamic lesions

Lesions of hypothalamus may produce any of the following specific clinical conditions:

1. Diabetes insipidus. It occurs due to deficiency of ADH occurring in tumour or sham lesions of anterior hypothalamus in which supraoptic nuclei are damaged. It is characterized by excessive thirst and polydipsia. For details see page 549.

2. Narcolepsy. It is a hypothalamic disorder with abnormal sleep pattern. Patient gets sudden attacks of unresistable desire of sleep during day time. The duration of sleep is usually short—from few seconds to about 20 min.

3. Cataplexy. Cataplexy refers to a sudden emotional outburst of anger, fear or excitement associated with narcolepsy. The attack lasts for few minutes. In this consciousness is not lost.

Chapter 10.4

Physiological Anatomy and Functions of Cerebral Cortex and White Matter of Cerebrum

CEREBRAL CORTEX

- **External features**
  - Sulci and gyri
  - Lobes of cerebral hemisphere
- **Cortical functional areas**
- **Phylogenetical divisions of cerebral cortex**
  - Allocortex
  - Mesocortex
  - Neocortex
- **Histological structure**
  - Types of cells
  - Laminae of neocortex
- **Areas, connections, functions and applied aspects of different lobes**
  - Frontal lobe
  - Parietal lobe
  - Temporal lobe
  - Occipital lobe

WHITE MATTER OF CEREBRUM

- Association fibres
- Commissural fibres
- Projection fibres
  - Corona radiata
  - Internal capsule

CEREBRAL CORTEX

EXTERNAL FEATURES

Cerebrum. It consists of two cerebral hemispheres which are separated from each other in the upper part by a median longitudinal fissure in which the *falx* cerebri (a fold of dura mater) invaginates. In the lower part, the two cerebral hemispheres are connected by the largest white commissure called *corpus callosum*.

Each cerebral hemisphere has three poles, three surfaces and three borders:

Poles. Three poles of each hemisphere are:

- Frontal pole anteriorly,
- Occipital pole posteriorly and
- Temporal pole that lies between the frontal and occipital poles, and points forwards and somewhat downwards.

Surfaces. Three surfaces of each hemisphere are:

- *Superolateral surface,*
- *Medial surface* and
- *Inferior surface,* which is further subdivided into an anterior orbital part and a posterior tentorial part.

SULCI AND GYRI

The surface of cerebral hemisphere is covered by a thin layer (2–4 mm thick) of grey matter called the cerebral cortex. The entire surface of cerebral hemisphere is folded with intervening grooves of fissures. The folds or convolutions are called gyri and the intervening fissures are called sulci. As a result of the folding of the cerebral surface, the cerebral cortex acquires a much larger surface area (about 2200 cm^2).

LOBES OF CEREBRAL HEMISPHERE

Each cerebral hemisphere is divided into four lobes:

The four lobes of each cerebral hemisphere as seen on superolateral surface (Fig. 10.4-1) are:

- ***Frontal lobe.*** It lies in front of the central sulcus and above the posterior ramus of the lateral sulcus. It is concerned with motor functions.
- ***Parietal lobe.*** It lies between the central sulcus and parieto-occipital sulcus and upper part of first imaginary line.
- ***Temporal lobe.*** It lies below the posterior ramus of the lateral sulcus. It is concerned with hearing.

- ***Occipital lobe.*** It lies behind the parieto-occipital sulcus and its continuation the first imaginary line. It is concerned with vision.

CORTICAL FUNCTIONAL AREAS

On the basis of number and thickness of cortical laminae and cell type (cytoarchitecture), Brodmann divided the cortex into 47 areas.

Classically, cortical functional areas are subdivided into (Fig. 10.4-2):

Motor areas include:
- Primary motor area (Brodmann's area 4),
- Premotor area (area 6),
- Frontal eye field (area 8),
- Supplementary motor area.

Sensory areas include:
- Primary somaesthetic areas (area 3, 1 and 2).
- Secondary (supplementary) somaesthetic area,
- Somaesthetic association areas (area 5, 7 and higher association area 40).

Auditory areas include:
- Primary auditory area (area 41) or auditory area I,
- Auditory association area (area 42) or auditory area II,
- Higher auditory association area (area 22).

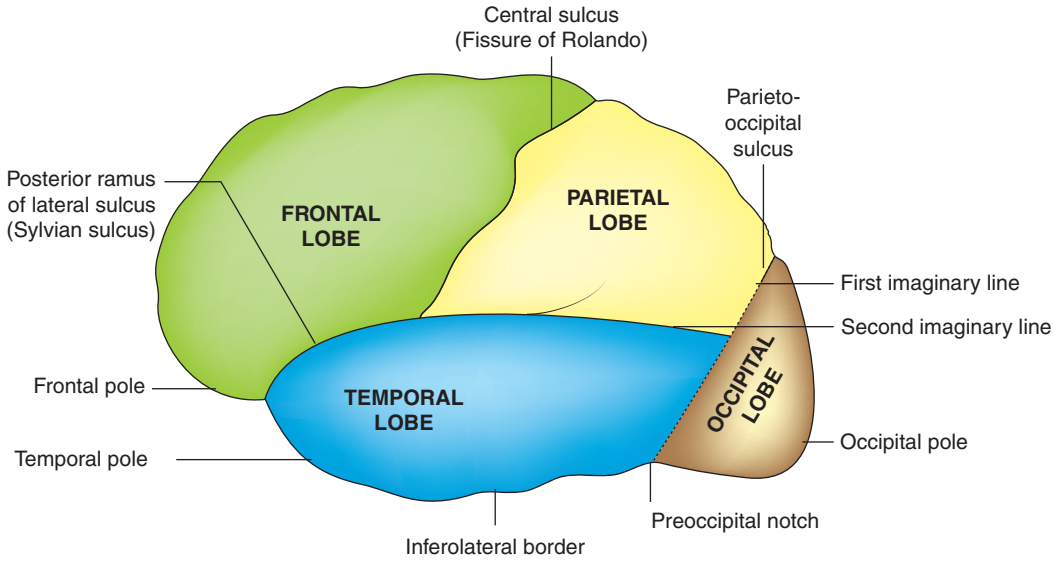


Fig. 10.4-1 Superolateral surface of cerebral hemisphere to show different lobes, poles and borders.

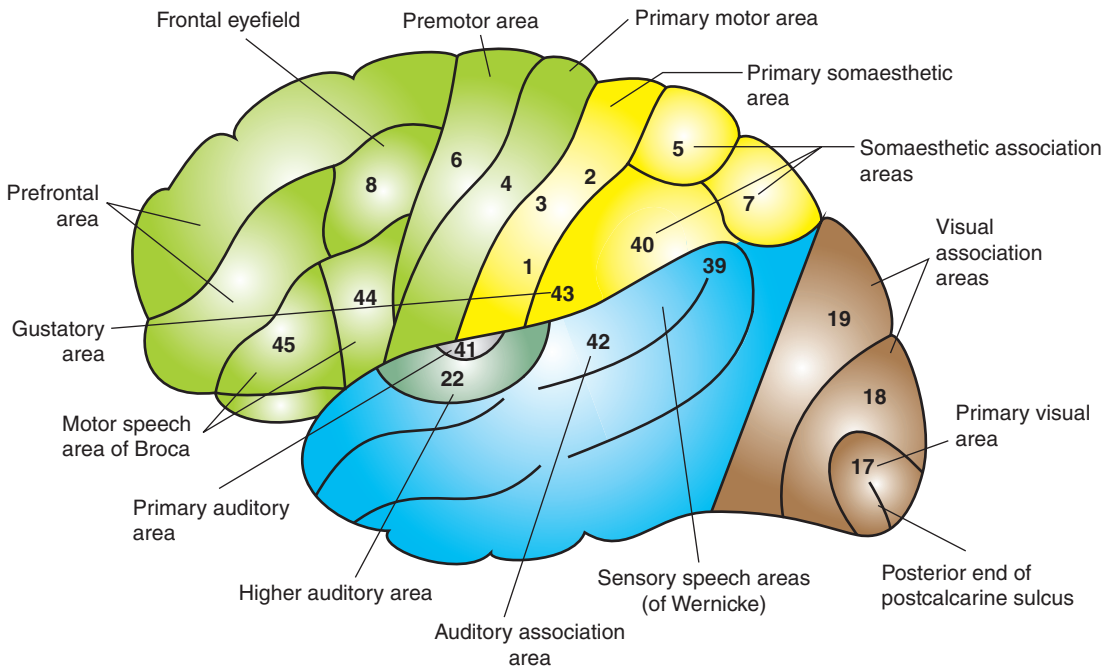


Fig. 10.4-2 Different areas on the lateral surface of the human cerebral cortex.

Visual areas include:

- Primary visual area (area 17) or visuostriate area of visual area I,
- Visual association area 18 (peristriate area) and
- Visual association area 19 (parastriate area).

Speech areas include:
Motor speech area comprises:

- Anterior area (Broca's area) or areas 44, 45 and
- Superior area

Sensory speech areas comprise:

- Area 39 (or reading centre),
- Area 40 and
- Area 22 (Wernicke's area).

Smell area is:

- Area 28.

Gustatory area is:

- Area 43.

Note. The different cortical functional areas are described along with the description of various lobes of cerebral cortex.

PHYLOGENETICAL DIVISIONS OF CEREBRAL CORTEX

The cerebral cortex, also known as pallium, is divided phylogenetically into three: allocortex, mesocortex and neocortex.

1. Allocortex or old cortex forms about 10% of the entire cortex and can be further subdivided into:

- *Archipallium* (ancient cortex), which includes hippocampus and dentate gyrus.
- *Paleopallium* (old cortex) comprises uncus and part of parahippocampal gyrus, which belong to the piriform area of olfactory cortex.

  Since most of the allocortex is located around the peripheral margin of the diencephalon in the form of a ring, it is also called *limbic cortex*. This ring of limbic cortex functions as a two-way communication linkage between neocortex and lower limbic structures. Along with thalamus and hypothalamus, the limbic cortex is concerned with emotional and instinctive behaviour.

2. Mesocortex, which is the transitional zone between allocortex and neocortex, comprises the cingulate gyrus, part of parahippocampal gyrus and *subiculum*.

3. Neocortex, also called an isocortex, comprises rest of 90% of the cerebral cortex in human brain. The actual extent of neocortex has increased with the evolution of mammals. The comparative ratio of allocortex and neocortex in rat, cat, monkey and human being is shown in Fig. 10.4-3.

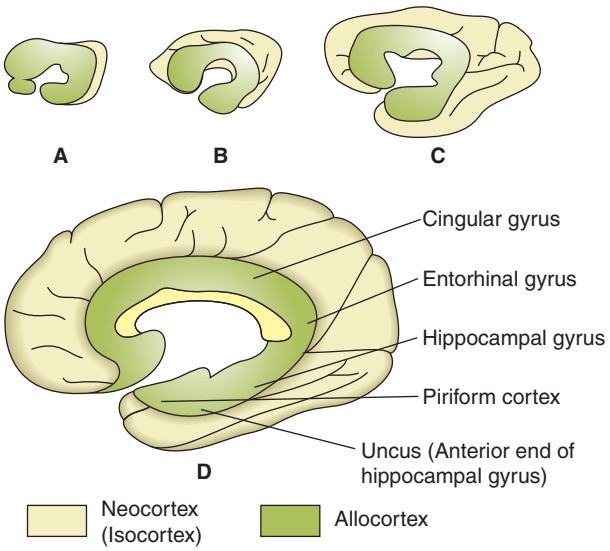


Fig. 10.4-3 Relative extent of allocortex and neocortex in different mammals: A, rat; B, cat; C, monkey and D human.

HISTOLOGICAL STRUCTURE OF CEREBRAL CORTEX

Histologically, the allocortex is composed of three distinctive layers, while the neocortex is composed of six layers named I–VI from outside to inside (Fig. 10.4-4).

Histologically, the cerebral cortex is composed of nerve cells and fibres. Three types of cells may be identified in the cerebral cortex:

1. Pyramidal cells. About two-thirds of all cortical neurons are pyramidal cells. These cells have triangular cell bodies, with the apex generally directed towards the surface of cortex. Axon arises from the base of the cell and a large dendrite arises from the apex and other dendrites arise from the basal angles. The processes of pyramidal cells extend vertically through the entire thickness of cortex. These cells are present in layer II, III and V of neocortex.

2. Stellate or granule cells. These cells form about one-third of the total neurons. These cells have small cell bodies from where the dendrites arise in all directions. Layer IV is packed with such cells and is best developed in primary sensory cortex.

3. Fusiform cells. These are comparatively few in number. Such cells have spindle-shaped cell bodies and are present in layer VI.

Laminae of neocortex

- The six laminae of neocortex numbered I–VI are (Fig. 10.4-4):

I. Molecular or plexiform layer. It mainly consists of transverse nerve fibres dispersed with occasional horizontal cells.

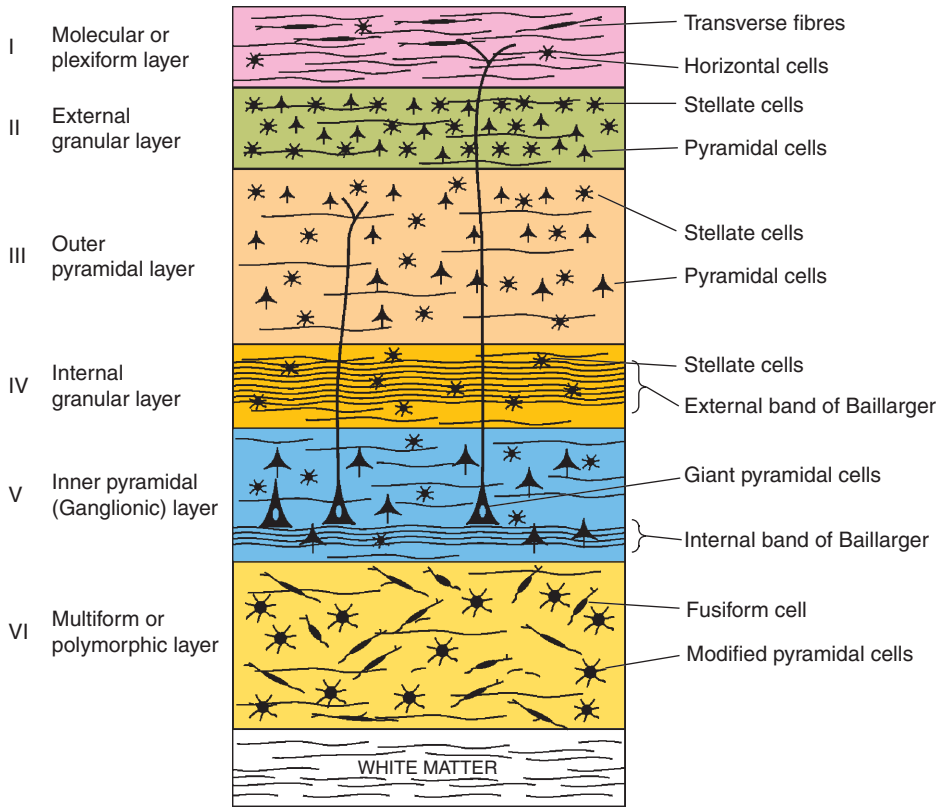


Fig. 10.4-4 Histological structure of neocortex.

- The transverse fibres derived from the apical dendrites of pyramidal cells, axons of stellate and *Martinotti* cells (pyramidal cells with short axon) of deeper layer, which ascend and ramify horizontally in this layer.
- The horizontal cells of Cajal are small pear-shaped or fusiform.

II. External granular layer. It contains numerous stellate or granule cells and a lesser number of small pyramidal cells. It is traversed by afferent and efferent projection fibres. Dendrites of cells of this layer pass into the molecular layer. The axons end in the deeper layer. Some axons enter the white substance of hemisphere.

III. Outer pyramidal layer. It consists mainly of pyramidal neurons and some stellate and basket cells. The pyramidal cells are of two types: the *small cells* lie in the superficial zone and *medium-sized* cells occupy the deeper zone.

IV. Internal granular layer. This layer consists of densely packed stellate cells. The inner zone of this layer is traversed by a prominent aggregation of transversely running fibres called **external band of Baillarger**.

V. Inner pyramidal (ganglionic) layer. This layer consists of large pyramidal cells. It is specially developed in the motor cortex, where these cells are called giant cells or Betz cells. This layer is traversed by a prominent aggregation of transversely running fibres called **internal band of Baillarger**.

VI. Polymorphous or multiform layer. This layer contains neurons of various sizes and shapes, many of which are probably modified pyramidal cells. Many spindle-shaped cells called fusiform cells are present in this layer. This layer also contains cells of Martinotti, whose axons project vertically towards the outer surface of cortex to ramify in the molecular layer. This layer merges with the white matter of cerebral cortex.

IMPORTANT NOTE

- Most of the afferent fibres from the specific nuclei of thalamus make synapses in the laminae I–IV.
- Afferent projections from the non-specific thalamic nuclei and from ascending reticular system terminate in all laminae of cortex.
- Laminae II and IV are concerned with sensorial modalities.
- Laminae III–V are meant for somatomotor or visceromotor activities.
- Laminae I and VI are engaged for integration of association of sensorimotor behaviour.

AREAS, CONNECTIONS, FUNCTIONS AND APPLIED ASPECTS OF DIFFERENT LOBES

A. FRONTAL LOBE

The frontal lobe lies in front of the central sulcus and above the posterior ramus of the lateral sulcus (Fig. 10.4-2). It forms

about one-third of cortical surface. On the basis of function, the frontal lobe is subdivided into two main areas:

- Precentral cortex and
- Prefrontal cortex.

I. PRECENTRAL CORTEX

Precentral cortex refers to the posterior part of the frontal lobe that includes lip of central sulcus, precentral gyrus and posterior part of superior, middle and inferior frontal gyri. Stimulation of different points in this area causes activity of discrete skeletal muscles. Therefore, precentral cortex is also called excitomotor area of cortex. Stimulation of motor area also causes some sensory perception.

Therefore, nowadays the motor cortex and sensory cortex are together known as *sensorimotor cortex*.

Areas in precentral cortex

The precentral cortex includes following important areas:

- Primary motor area (Brodmann's area 4),
- Premotor area (Brodmann's area 6, 8, 44 and 45) and
- Supplementary motor area.

Note. Now, it has been suggested that the sensorimotor cortex (area 4 and 6) is primarily motor and secondarily sensory in function; hence these areas have been designated as M-1.

1. Primary motor area (area 4)

Extent of area 4. It lies in the precentral gyrus extending into the paracentral lobule on the medial surface. The area also includes the anterior wall of the central sulcus and other gyri of the frontal lobe (Fig. 10.4-5).

Structural characteristics. This area contains all the six layers of cortex (see page 749). Special features of this area are the presence of giant pyramidal cells called Betz cells in ganglionic layer and a thin granular layer (that is why also called agranular cortex).

Topographic representation. Different parts of the contralateral half of the body are represented separately in more or less inverted order. Those parts of the body which carry out the most skilled movements, e.g. the fingers and thumb have the largest areas of cortical representation. The areas for tongue, jaw and facial movements lie in the inferior part of the motor cortex; those for the arm, trunk and leg are arranged in sequence in the motor area, which extends to the vertex and on to the medial surface of the cerebral hemisphere. The parts in the paracentral lobule are for the foot and perineum. Thus body is represented upside down (however face is not represented in inverted manner) (Fig. 10.4-6A). Stimulation of the points representing upper parts of the face, pharynx and the vocal cords produces bilateral responses.

Functions. It is concerned with initiation of voluntary movements of the contralateral half of the body and initiation of speech.

2. Premotor area

Premotor area lies anterior to the primary motor area and includes Brodmann's area 6, 8, 44 and 45.

Area 6

Location. It abuts on the primary motor cortex area both above and behind and thereby includes the posterior parts of the superior, middle and inferior frontal gyri. This area is divided into two parts, upper 6a and lower 6b. Cells from this area contribute fibres to pyramidal tracts.

Topographical organization of this area is roughly the same as that of primary motor cortex.

Functions. Area 6 co-ordinates the voluntary action of area 4 and extrapyramidal system and is, therefore, involved in the integration of voluntary movements. Thus, the skilled movements are accurate and smooth.

Electrical stimulation of area 6a in human being causes same effects as that of stimulation of area 4. However, the strength of stimulus must be stronger while stimulating area 6.

- Stimulation of area 6a causes generalized pattern of movements like rotation of head, eyes and trunk towards the opposite side.
- Stimulation of area 6b produces rhythmic, complex co-ordinated movements involving the muscles of face, buccal cavity, larynx and pharynx.

Lesions of area 6 in monkeys lead to loss of skilled movements. The recovery may occur, but the movements become awkward. This also produces grasping reflex. Lesions involving area 6 along with area 4 produce severe symptoms of hemiplegia with spastic paralysis.

Area 8. It is also called frontal eyefield.

Location. It lies anterior to area 6.

Afferents to this area come from the occipital lobe and dorsomedial nucleus of thalamus.

Efferents from area 8 go to nuclei of third, fourth and sixth cranial nerves.

Functions. It is concerned with the control of eye movements.

Electrical stimulation of area 8 causes conjugate movements of eyeballs to the opposite side, opening and closure of eyelids, pupillary dilation and lacrimation.

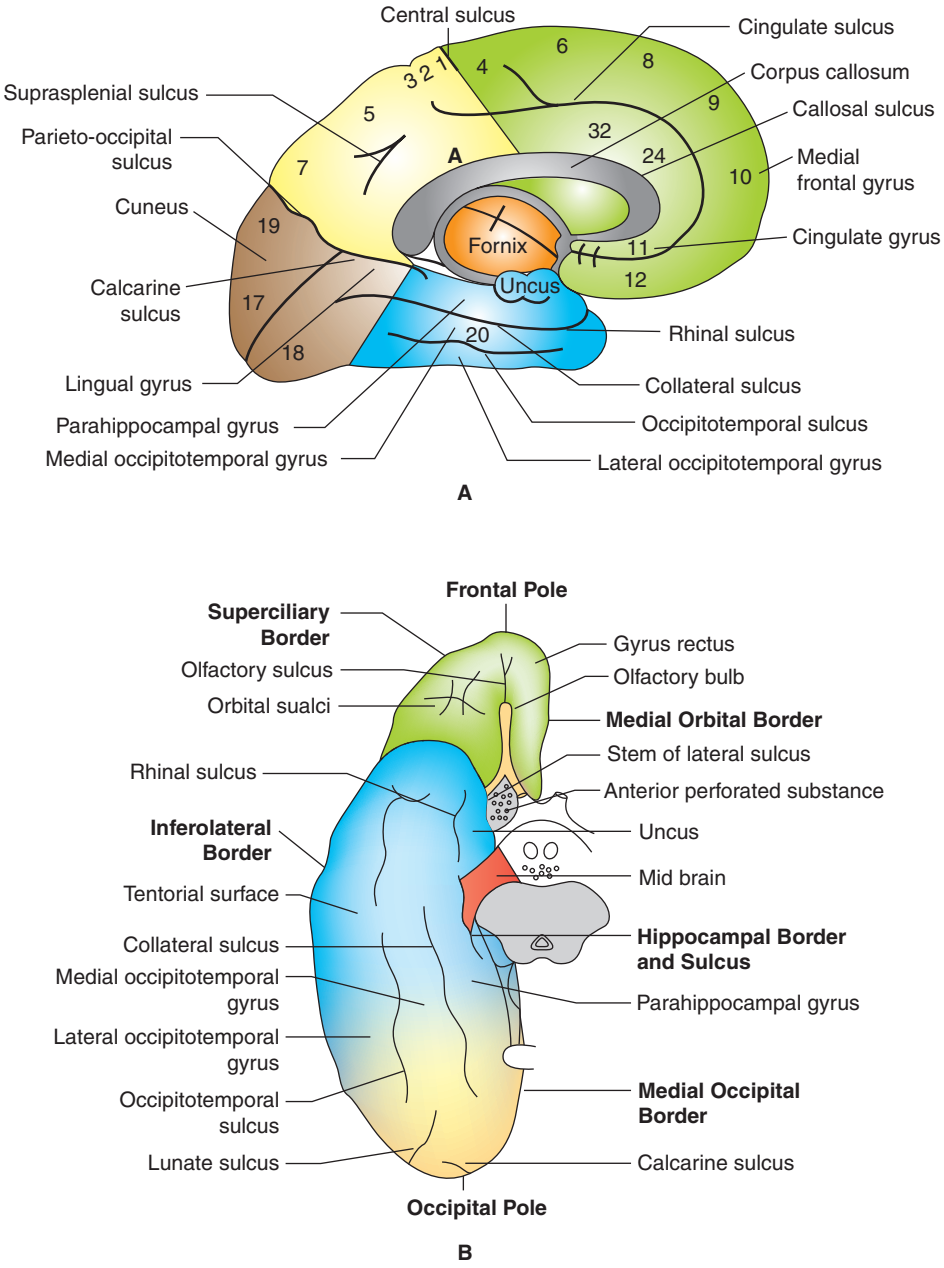


Fig. 10.4-5 Left cerebral hemisphere showing lobes, sulci and gyri: A, medial surface and B, inferior surface.

Lesions of this area turn the eyes towards the affected side. Conjugate movements of the eyes are absent. Pupil and eyelids are not affected.

Area 44 and 45 or Broca's motor speech area.

Location. It is special region of premotor cortex situated in the inferior frontal gyrus.

Functions. This area, specially in dominated hemisphere (*left hemisphere in right handed person*), is concerned with the movements of those structures, which are responsible for the production of voice and articulation of speech, that is, it causes activation of vocal cords, simultaneously with movements of mouth and tongue during speech.

Lesions of this area cause motor aphasia, i.e. inability to speak the word though vocalization is possible.

3. Supplementary motor area

Location. Supplementary motor area is located in the medial surface of frontal lobe rostral to primary motor area (Fig. 10.4-5).

Topographical organization. In this area, components of the upper body are located dorsal to those of the lower body.

Functions. This area in association with the premotor area provides attitudinal movements, fixation movement of different segments of the body and positional movements of head and eyes.

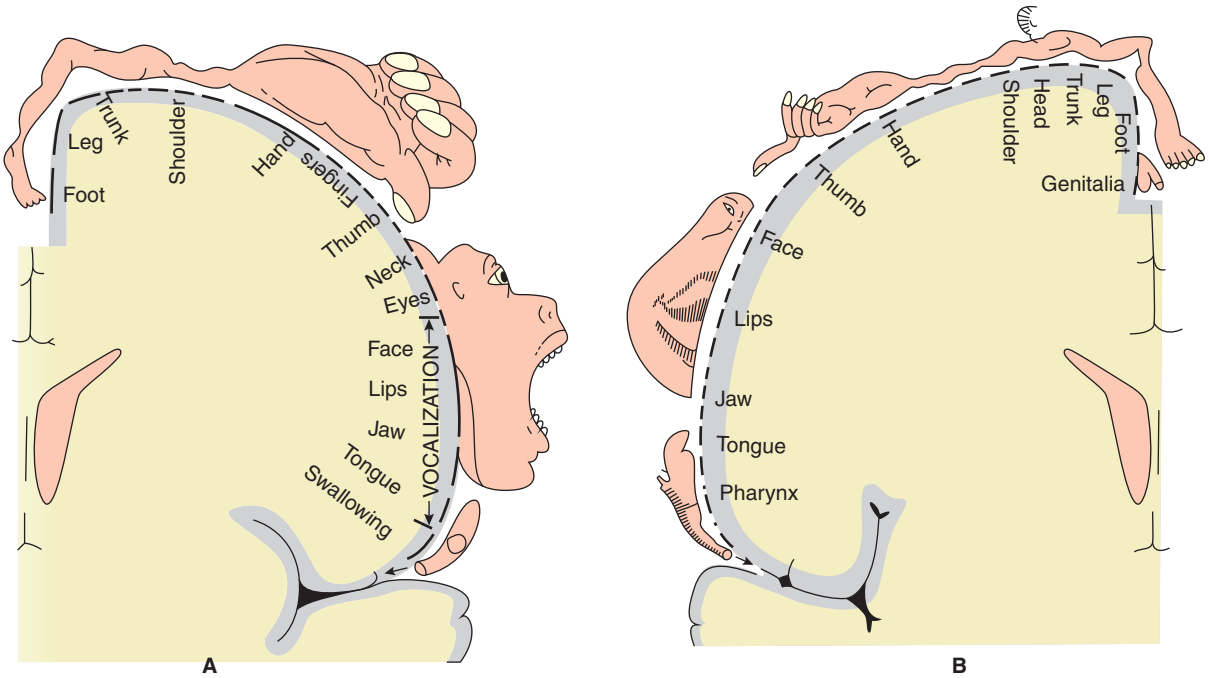


Fig. 10.4-6 Topographical representation (homunculus) of motor, A and sensory, B areas in the cerebral cortex.

Connections of precentral cortex

Afferents to precentral cortex come from the following sources:

1. *Fibres from adjacent regions* include those from:
- Somatic sensory area of parietal cortex.
- Adjacent areas of frontal cortex anterior to motor cortex,
- Subcortical fibres from auditory and visual cortices.

2. *From opposite hemisphere.* Subcortical fibres passing through the corpus callosum connect corresponding areas of cortices in the two sides of brain.

3. *Fibres from thalamus* include:
- Tracts from ventrolateral and ventroanterior nuclei of thalamus which in turn receive from the cerebellum and basal ganglia. They cause co-ordination between functions of motor cortex, basal ganglia and cerebellum.
- Fibres from intralaminar nuclei of thalamus to cause general level of excitability of motor cortex.

Efferents from precentral cortex include:
1. *Corticospinal tract* (pyramidal tract) is the most important tract through which the motor cortex controls the activity of the anterior horn cells in the spinal cord. For details see page 703.

2. *Collaterals from pyramidal tracts* and a large number of fibres from the motor cortex go to deeper regions of cerebrum and brain stem as follows:

- *Adjacent areas of cortex.* Axons of Betz cells send collaterals to the adjacent areas of cortex. These collaterals inhibit adjacent areas (lateral inhibition) and sharpen the boundaries of excitatory signals.
- *Basal ganglia.* A large number of fibres go to caudate nucleus and putamen from where additional pathway goes to brain stem.
- *Red nucleus.* Some fibres go to red nuclei and then to spinal cord through rubrospinal tracts.
- *Reticular substance.* Some fibres go to reticular substance of brain stem. From where fibres pass to spinal cord through reticulospinal tract.
- *Vestibular nuclei.* Some fibres go to vestibular nuclei, from where through the vestibulospinal tract they reach the spinal cord.
- *Pontine nuclei.* A large number of fibres synapse in pontine nuclei and pass to cerebellum (pontocerebellar fibres).
- *Inferior olivary nuclei.* Collaterals also go to inferior olivary nuclei and then to cerebellum through olivocerebellar tract.

II. PREFRONTAL CORTEX

Location. Prefrontal cortex, also called prefrontal lobe or orbitofrontal cortex, is the anterior part of frontal lobe lying anterior to area 8 and 44 (Fig. 10.4-5).

Major areas. Prefrontal cortex has different Brodmann's areas, such as 9 to 14, 23, 24, 29, and 32, 44 to 47.

Connections of prefrontal cortex shown in Fig. 10.4-7 are:

Afferents to prefrontal cortex come from:
(i) Dorsomedial nucleus of thalamus project on to areas 9–12 on the lateral and adjacent medial surface and areas

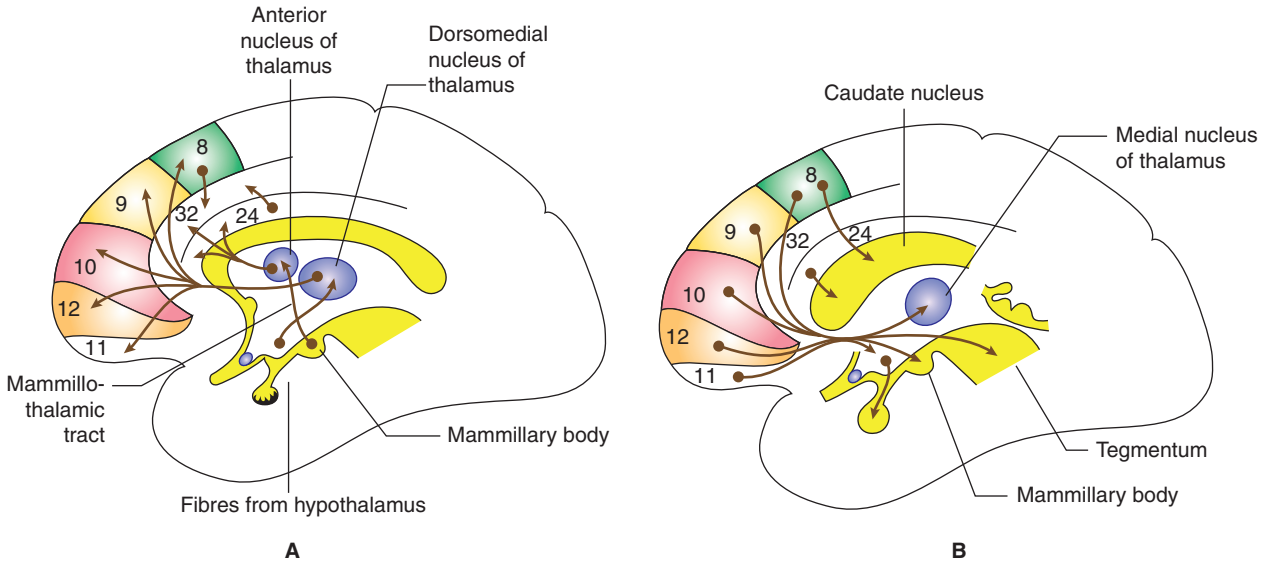


Fig. 10.4-7 Connections of prefrontal lobe: A, afferents and B, efferents.

44–47 in the inferior frontal gyrus. Since the dorsomedial nucleus of thalamus, in turn receives afferents from the posterior hypothalamus; therefore, the impulses which reach the prefrontal lobe via the medial nucleus represent a resultant of hypothalamic and thalamic activity.

(ii) Anterior nuclei of thalamus project on to cingulate gyrus (areas 23, 24, 29 and 32). Since the anterior nucleus of thalamus receives afferents from the mammillary bodies of the hypothalamus, which in turn receives the afferents from the hippocampus via the fornix. The hippocampus is thus ultimately projected to inhibitory area 24.

The prefrontal lobe thus forms a closed-circuit connection with the thalamus called *Papez circuit* (Fig. 10.10-3). This circuit is responsible for resting EEG and plays an important role in the genesis of emotions.

Efferents from prefrontal cortex go to:

(i) *Thalamus.* Fibres from area 9 and 10 go to ventral and medial thalamic nuclei.
(ii) *Tegmental reticular formation.* Fibres from area 9 and 10 also go to reticular formation in the tegmentum.
(iii) *Pontine nuclei.* Fibres from area 10 pass to the pontine nuclei as frontopontine tract and thence to the cerebellum.
(iv) *Caudate nucleus.* The inhibitory area 8 and 2, 4, 5 discharge to the caudate nucleus.
(v) *Mammillary bodies.* Fibres from area 13, the hippocampus, uncus and amygdala project via the fornix to the mammillary bodies of the hypothalamus.

Functions of prefrontal cortex

1. Centre for planned actions. Prefrontal association areas in close association with the motor cortex plan complex patterns and sequence of motor movements.

2. Centre for higher functions. This forms the centre for higher functions like emotions, learning, memory and social behaviour. It is responsible for various autonomic changes during emotional conditions because of its connections to hypothalamus and brain stem.

3. Seat of intelligence. Short-term memories are registered in the prefrontal cortex. It can keep track of many bits of information and also has ability to recall this information bit by bit for subsequent thoughts. It is therefore called seat of intelligence or an organ of mind.

4. Control of intellectual activities. The prefrontal cortex has the following intellectual abilities:

- To prognosticate.
- To plan the future.
- It allows the person to concentrate on the central theme of thought. It helps in depth and abstractness of thought and thereby in elaboration of thought.
- It allows to delay action in response to incoming sensory signals so that sensory information can be weighed until the best response is obtained.
- It allows to consider the consequence of motor actions before their performance.
- It plays role in solution of complicated mathematical, legal and philosophical problems.
- It allows to correct avenues of information in diagnosis of rare diseases.
- It allows to control one's activity according to the moral laws.

APPLIED ASPECTS

Frontal lobe syndrome

Frontal lobe syndrome refers to the symptom complex occurring due to injury or ablation of prefrontal cortex.

Prefrontal leucotomy, i.e. cutting the connection between the thalamus and prefrontal lobe also results in frontal lobe syndrome. Bilateral prefrontal lobectomy (extirpation) also results in a similar condition. In the past, these operations were performed in patients with severe mental illness. However, nowadays due to availability of tranquilizers and other drugs (which can control mental illness) these operations are not conducted because of the associated complications.

Characteristic features of frontal lobe syndrome are:

- *Flight of ideas,* which results in difficulty in planning.
- *Emotional instability,* there occurs lack of restraint leading to hostility, aggressiveness and restlessness.
- *Euphoria,* i.e. a false sense of well-being and failure to realize or indifference to seriousness of other's feelings or emotions.
- *Impairment of memory* occurs for recent memory only. The memory of remote events is not lost.
- *Loss of moral and social sense* is common and there is loss of love for family.
- *Lack of attention and power of concentration* associated with restlessness is a common feature.
- *Lack of initiative following marked depression* of intellectual activity leads to reducing mental drive.
- *Functional abnormalities* may occur in the form of:
  - Hyperphagia, i.e. increased appetite,
  - Loss of control over urinary or rectal sphincters,
  - Disturbances in orientation and
  - Slight tremor.

B. PARIETAL LOBE

Parietal lobe (Fig. 10.4-1) lies between the central sulcus and parieto-occipital sulcus and upper part of first imaginary line. Below it is separated from the temporal lobe by the posterior ramus of lateral sulcus and in continuation of it the second imaginary line.

AREAS OF PARIETAL LOBE

Functionally, parietal lobe can be divided into three parts:

- Primary sensory area (which corresponds to Brodmann's areas 3, 1 and 2),
- Secondary sensory area and
- Sensory association areas (Brodmann's area 5 and 7).

Note. Since stimulation of sensory area also produces some motor response and stimulation of motor area also causes some sensory perception, therefore, nowadays the sensory and motor cortex is combinedly called somatosensory cortex, and:

- Primary sensory area (area 3, 1 and 2) is called primary somatosensory (sensorimotor) area or first somatosensory area (SI), and
- Secondary sensory is called second somatic sensory area (SII).

Primary sensory area (first somatic sensory area)

Location. The first SI occupies the posterior wall of the central sulcus, the post-central gyrus and the post-central part of the paracentral lobule (Fig. 10.4-5).

Major areas. It includes Brodmann's area 3, 1 and 2.

Structurally, the primary sensory cortex is granular cortex which is densely packed with stellate cells, with a few small and medium-sized pyramidal cells.

Topographical organization. The primary sensory cortex receives sensory inputs from the opposite half of the body. The representation of the body within this area is similar to that already noted in the primary motor cortex (page 752, Fig. 10.4-6B).

The sensations derived from the skin are appreciated in the anterior part of the area and proprioceptive sensations in the posterior part of the area.

Electrical stimulation of primary sensory area (SI) produces vague sensations like numbness and tingling.

Lesions. If lesions occur only in the sensory cortex without involvement of thalamus, the sensations are perceived but the discriminative functions are lost. If thalamus is also affected by lesion, there occurs loss of sensations in the opposite side of the body.

Secondary sensory area

Location. Secondary sensory area, also called second somatic sensory area (SII) is situated in post-central gyrus below the area of face of first somatic sensory area. Most of it is buried in the superior wall of the sylvian fissure (lateral cerebral sulcus).

- *Topographical representation.* The SII area receives sensory impulses from SI as well as from the thalamus directly. Like SI, the SII area also manifests a dermatomal (point-to-point) sequence of representation (although there is more overlap). Thus the body is represented twice in the somatic sensory cortex, i.e. in area SI as well as in area SII.
- Neurons in the anterior part of area SII respond to touch whereas neurons in the posterior part can be excited by touch, auditory, visual and nociceptive stimuli.

Lesions of SII produce deficits in discrimination power, whereas sensory processing in SI is not affected.

Sensory association areas

The sensory association areas include area 5 and 7. The area 40 is higher association area.

Area 5. It lies posterior to area SI in the parietal lobe and contains neurons which react to passive or active rotation of a joint or joints. Few neurons respond to tactile stimuli

like the other areas in SI and SII, area 5 also displays a columnar organization (point-to-point representation).

Area 7. It is located in superior parietal lobule deep into the intraparietal sulcus extending close up to the occipital lobe. This area is concerned with more elaborate process of discrimination between the stimuli.

Area 40. This is higher association area, located in supramarginal gyrus, concerned with stereognosis, i.e. recognition of common objects placed in the hand without looking at them. A lesion affecting area 40 produces tactile agnosia (astereognosis and tactile aphasia).

CONNECTIONS OF PARIETAL LOBE

Afferent connections of somatosensory area

First somatic sensory area (SI) receives afferent projections from posteromedial (VPM) and posterolateral (VPL) parts of the ventral posterior nucleus of thalamus, which convey exteroceptive and proprioceptive impulses from the contralateral side but from both the sides of face. The lower part of post-central gyrus acts as taste receptive centre.

Second somatic sensory area (SII) receives afferent projection from area SI as well as directly from the thalamus.

Sensory association area receives impulses from area SI and SII.

Efferents from somatosensory area

- Pyramidal cells of the sensory area contribute fibres to corticospinal, corticobulbar and corticonuclear tracts. These fibres, presumably, modulate the sensory input at the root entry zone of posterior grey column of the spinal cord, and nuclei gracilis and cuneatus of the lower medulla.
- All somatosensory areas (particularly SI) send fibres to the caudate nucleus and putamen.
- Area SI also sends efferent fibres:
  - Back to its own thalamic projection nuclei
  - To the tectum, pons and cerebellum.

Association fibres from the sensory cortex

- Through association fibres the sensory cortex is connected with other cortical areas.
- Association fibres interlinking the areas SI, SII, area 5 and area 4 are involved in somatic sensations (Fig. 10.4-8).

Commissural fibres from the sensory cortex

- Commissural fibres are mostly axons of pyramidal cells of layer III and connect the corresponding somatosensory areas with those of the opposite hemisphere.
- Area SI projects to the contralateral areas SI and SII.
- Area SII projects only to area SII of the opposite hemisphere.

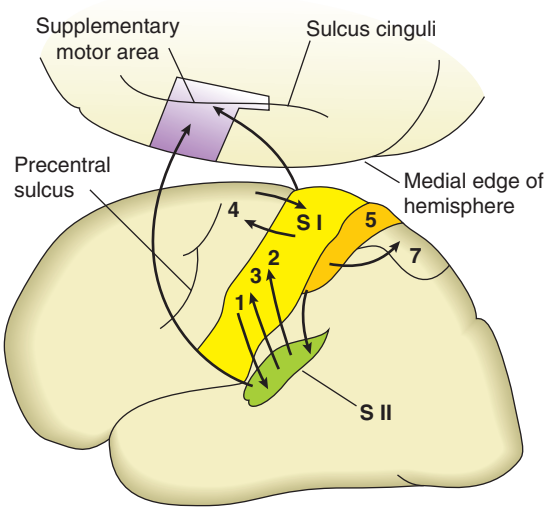


Fig. 10.4-8 Connections of the parietal lobe areas involved in somatic sensations.

FUNCTIONS OF PARIETAL LOBE

First somatic sensory area (SI) (areas 3, 1 and 2) localizes, analyses and discriminates different cutaneous and proprioceptive senses.

Second somatic sensory area (SII) receives sensory impulses from SI and from thalamus directly. Though the exact role of this area is not clear; it is concerned with perception of sensation. Thus, the sensory parts of the body have two representation in area SI and area SII.

Sensory association areas (area 5 and 7) are associated with more elaborate process of discrimination between the stimuli, thus helps in differentiating the relative intensity of different stimuli. Therefore, warm objects are distinguished from warmer, cold from colder and rough from rougher, etc.

Higher association area (area 40) helps in recognition of common familiar objects placed in the hand without looking at them (stereognosis).

Inferior part of post-central gyrus contains centre for taste and general sensations from tongue. Lesion of this part causes loss of taste and general sensations of opposite half of the tongue.

Angular gyrus helps in recognition of spatial relationship by:

- Tactile localization, i.e. the precise point stimulated is accurately localized.
- Tactile (two-point) discrimination, i.e. two points of a compass placed close together are recognized as two and not as one.
- Accurate estimation of the extent and direction of small joint displacements.

APPLIED ASPECTS

A. Unilateral removal of parietal lobe results in defective response to stimuli due to mental imbalance of perception of sensation specially proprioception and fine touch:
- There occurs loss of discrimination and localization ability and temperature sense on opposite side of the body.
- Loss of control of voluntary movements (ataxia) on opposite side.

B. Bilateral removal of post-central gyrus (area 3, 1, 2 and 5) in animals results in:
Complete loss of tactile placing but visual placing in initial stages is retained. Therefore, the animal's limbs are inactive to tactile stimulation with closed eyes.

C. Removal of inferior parietal lobules (specially area 7)
Unilateral lesion causes a marked failure in care of the left half of the body. Since the body images cannot be appreciated.

Body image remains awareness of position of body parts relative to one another.

In severe cases, such individuals shave half of their face, dress half of their bodies and read half of each page.

Bilateral lesion. In bilateral lesion, visual placing is also lost but coarse tactile placing is retained. As optical righting reactions are lost; therefore, individual is unable to make use of visual information that is inability to copy designs etc. (called *constructional apraxia*) and there is *spatial disorientation.*

C. TEMPORAL LOBE

Temporal lobe (Fig. 10.4-1) lies below the posterior ramus of the lateral sulcus and its continuation the second imaginary line. Behind it is separated from the occipital lobe by the lower part of first imaginary line, which connects the upper end of parieto-occipital sulcus to the parieto-occipital notch.

AREAS OF TEMPORAL LOBE

The major areas in the temporal lobe are (Fig. 10.4-5):

- Primary auditory area (area 41 and 42) and
- Auditory association area (area 22, 21 and 20).

Primary auditory area

Primary auditory area, also called audiosensory area, includes Brodmann's area 41 and 42 and forms the centre for hearing.

Location. It is situated in the middle of the superior temporal gyrus on the upper margin and on its deep or insular aspect (Heschl's or transverse temporal gyrus). Heschl's gyrus can be seen only when the lips of the lateral sulcus are widely separated (Fig. 10.4-2).

Connections *of this area* are:
Afferents are received from:
- Medial geniculate body via auditory radiations and
- Pulvinar of thalamus.

Efferents are sent to:
- Medial geniculate body,
- Superior colliculus and
- Pulvinar.

Functions. This area perceives the nerve impulses as sound, i.e. auditory information, such as loudness, pitch, source and direction of sound.

Auditory association area

- Auditory association area corresponds to Brodmann's areas 22, 21 and 20.

Area 22

Location. Area 22, also called Wernicke's area, is a sensory speech centre situated in the posterior part of superior temporal gyrus behind the area 41 and 42 (Fig. 10.4-2) in the categorical hemisphere, i.e. dominant hemisphere.

Functions. It is concerned with:
- Interpretation of the meaning of what is heard and
- Comprehension of spoken language and the formation of ideas that are to be articulated in speech.

Areas 21 and 20

Location. Areas 21 and 20 are located in the middle and inferior temporal gyrus, respectively.

Functions. These areas receive impulses from the primary area and are concerned with interpretation and integration of auditory impulses.

Lesions of these areas impair auditory, short-term memory without impairing visual memory.

APPLIED ASPECTS

Unilateral removal of temporal lobe
Unilateral removal of temporal lobe causes no deafness. This is because each ear is bilaterally represented in the auditory pathway from the medulla upwards and projects about equally to the two cerebral hemispheres. Thus the removal of one auditory cortex has only a slight effect on auditory acuity (sharpness of hearing).

Temporal lobe syndrome
Temporal lobe syndrome, also known as *Kluver–Bucy syndrome,* is produced in animals particularly monkeys after removal of bilateral temporal lobe along with amygdala and uncus.

Characteristic features of this syndrome are:
- *Aphasia,* i.e. disturbances in speech.
- *Visual agnosia,* i.e. inability to recognize the objects in spite of good vision.

- Auditory disturbances in the form of frequent attacks of tinnitus, auditory hallucinations with sounds like buzzing, ringing or humming.
- Hyperphagia and omniphagia, i.e. animals start eating more and eating that diet which it was not eating previously.
- Hypersexuality is noted in male animals due to damage to amygdaloid nuclei and piriform cortex.
- Increased oral activity, i.e. animal starts repeatedly putting up in their mouth to all the moveable objects present in the surrounding.
- Hypermetamorphosis, i.e. animal starts responding to every stimulus, whether it is experienced before or not.
- Dreaming states, i.e. the animals are not aware of their own activities and have the feeling of unreality.

Clinical significance. In human with bilateral temporal lobe diseases or lesions, various above mentioned symptoms are seen.

D. OCCIPITAL LOBE

Occipital lobe lies behind the parieto-occipital sulcus and its continuation down an imaginary line (Fig 10.4-1). It is concerned with vision.

AREAS OF OCCIPITAL LOBE

Occipital lobe is mostly formed of sensory and association areas and has only slight motor function. It contains visual cortex having three areas (Figs 10.4-2 and 10.4-5):

- Primary visual cortex (area 17),
- Visual association area (area 18) and
- Visual association area or occipital eyefield (area 19).

Primary visual cortex is also called striate area (area 17). It lies on the medial surface of the occipital lobe in and near the calcarine sulcus occupying parts of lingual gyrus and cuneus. It also extends to the superolateral surface of the occipital lobe limited in front by the lunate sulcus.

It receives the fibres of the optic radiations which bring impulses from parts of both retinae, and these parts are represented within the area in a specific orderly manner. It constitutes the centre of vision.

Peristriate area, also called visual association area (area 18), lies in the walls of lunate sulcus.

Parastriate area (area 19) is also a visual association area. It lies in the cortex in front of the lunate sulcus.

CONNECTIONS

Afferents to visual cortex come from the lateral geniculate body in the form of optic radiations. The right visual cortex receives impulses arising from the temporal half of right retina and nasal half of the left retina; and the left visual cortex receives those arising from the temporal half of the left retina and nasal half of the right retina.

Thus, there is a point-to-point projection of the retina in the visual cortex in such a way that the right visual cortex is concerned with perception of objects situated to the left of the vertical median line in the visual fields and left visual cortex with the objects situated to the right half.

Efferents from visual cortex go to:

- *Various parts of the cerebral cortex* in both hemispheres, in particular they reach the frontal eyefield, which is concerned with eye movements.
- *Superior colliculus*, the pretectal region, and the nuclei of cranial nerves supplying muscles that move the eyeballs also receive efferents from the visual cortex.
- *Corticogeniculate* projection has also been evidenced physiologically.
- *Thalamus* (pulvinar) also receives efferents from the visual cortex.

FUNCTIONS

- *Primary visual area* (area 17) is concerned with perception of visual impulses.
- *Visual association areas* (area 18 and area 19) are concerned with the interpretation of visual impulses. These are involved in the recognition and identification of objects in the light of past experience.
- *Occipital eyefield area* (area 19) is concerned with the movements of eyeball. Therefore, like other sensory areas, the visual area is also to be regarded as partly motor in function.

WHITE MATTER OF CEREBRUM

Passing through, between and around the subcortical masses of grey matter of cerebrum are tracts of white fibres. The white fibres of cerebrum are of three types (Fig. 10.4-9):

- Association fibres,
- Commissural fibres and
- Projection fibres.

I. ASSOCIATION FIBRES

Association fibres connect the different gyri of the same hemisphere. These are of two types (Fig. 10.4-10):

1. Short association fibres, which connect the adjacent gyri are innumerable.

2. Long association fibres, which connect the widely separated gyri are arranged in five groups:

- *Superior longitudinal fasciculus.* It connects the frontal region to the temporal and occipital region.

- *Inferior longitudinal fasciculus.* It runs from the occipital pole to the temporal lobe.
- *Cingulum.* It runs from below the rostrum of the corpus callosum to the temporal lobe.
- *Fronto-occipital fasciculus.* It also runs from the frontal pole to the temporal and occipital regions. It lies in a deeper plane than the superior longitudinal fasciculus.
- *Uncinate fasciculus.* It connects the anterior speech area (Broca's area) and orbital surface of the frontal lobe with the cortex over the temporal pole.

II. COMMISSURAL FIBRES

Commissural fibres connect the corresponding parts of two cerebral hemispheres with each other. There are five bundles of commissural fibres (Fig. 10.4-11):

- Corpus callosum,
- Anterior commissure,
- Posterior commissure,
- Habenular commissure and
- Hippocampal commissure.

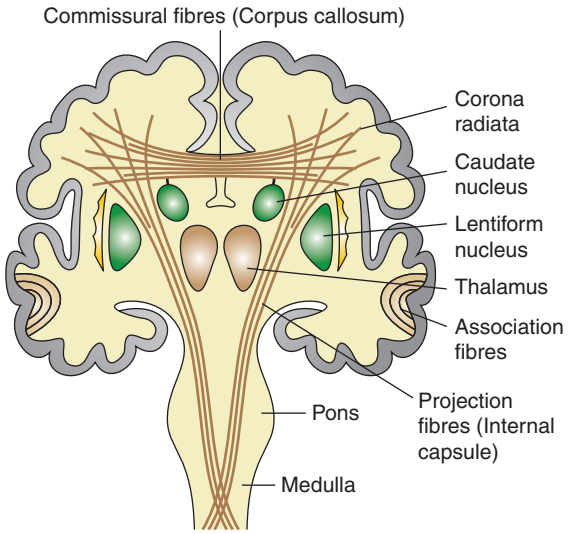


Fig. 10.4-9 Frontal view of coronal section of brain showing the position of association, commissural and projection fibres.

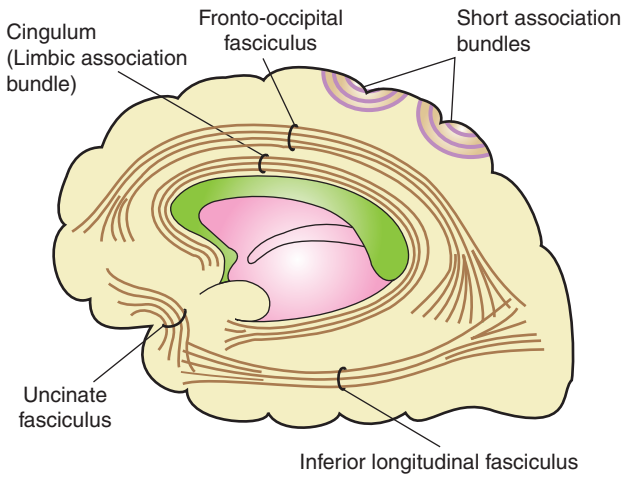


Fig. 10.4-10 Short and long association fibres.

III. PROJECTION FIBRES

Projection fibres connect the cerebral hemispheres with other parts of CNS, e.g. thalamus, brain stem and spinal cord.

Projection fibres include the afferent and efferent tracts contained in the corona radiata and internal capsule.

Afferent projection fibres include thalamic radiations, which according to their disposition have been named as:

- Anterior thalamic radiations,
- Superior thalamic radiations,
- Posterior thalamic radiation (including optic radiations) and
- Inferior thalamic radiation (including auditory radiations).

Efferent or motor projection fibres include:

- Corticobulbar and corticospinal tracts (pyramidal system),
- Corticopontine fibres,
- Corticorubral fibres and
- Corticothalamic fibres.

Corona radiata

Corona radiata (fountain of fibres) refers to that part of projection fibres that radiates from the upper end of internal capsule to the cerebral cortex (Fig. 10.4-9). It contains both the ascending and descending fibres.

Internal capsule

Internal capsule is a thick curved band of projection fibres (ascending and descending) that occupy the space between the thalamus and caudate nucleus medially, and the lentiform nucleus laterally. Superiorly, it fans out as corona radiata and inferiorly, the fibres descend into the crus cerebri.

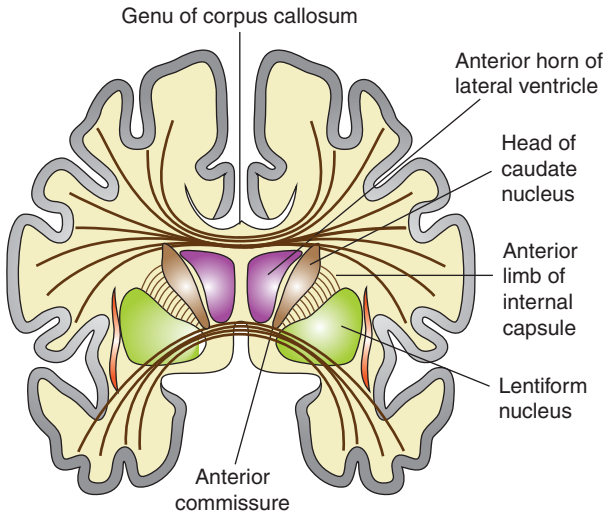


Fig. 10.4-11 Genu of corpus callosum and anterior commissure.

***Subdivisions of internal capsule are** (Fig. 10.4-12):*

1. Anterior limb. It is short and lies between the head of caudate nucleus and lentiform nucleus.

It consists of:

- *Anterior thalamic* radiations, containing reciprocal connections between the dorsomedial and anterior nuclei of thalamus and prefrontal cortex and gyrus cinguli.
- *Corticopontine* (frontopontine) fibres from the frontal cortex to nuclei pontis.

2. Genu. It is the region of the band in the capsule situated medial to the apex of the lentiform nucleus. It connects the anterior and posterior limbs. It contains following fibres:

- *Anterior part* of superior thalamic radiations and
- *Corticonuclear* (corticobulbar) fibres, which extend from the frontal eye field and motor area of cortex to the motor nuclei of the cranial nerves of the opposite side.

3. Posterior limb. It is bounded by the thalamus medially and lentiform nucleus laterally. It is longer than the anterior limb and contains:

- *Corticospinal tract.* The anterior two-thirds of posterior limb contains most of these fibres with upper limb in front, trunk in the middle and lower limb behind.
- *Corticopontine* (parietopontine) fibres.
- *Superior thalamic radiations,* which comprises the fibres having reciprocal connections between the ventral nuclei of thalamus and the parietal lobe.
- *Corticorubral tract,* which arises from the motor and premotor areas of frontal cortex and end in the red nucleus.

4. Retrolenticular or caudal part. It occupies the region behind the lentiform nucleus and consists of:

- *Posterior thalamic radiations* having fibres of reciprocal connections between the lateral geniculate body and the occipital lobe.
- *Optic radiations* (geniculocalcarine tract) extending from the lateral geniculate body to the visual cortex.
- *Corticopontine* (parietopontine, occipitopontine and temporopontine) fibres.

5. Sublentiform part. It occupies the region beneath the posterior part of the lentiform nucleus and consists of:

- *Auditory radiations,* which originate in the medial geniculate body and terminate in the Heschl's convolutions on the superior surface of the superior temporal gyrus (auditory area).
- *Inferior thalamic radiations* having fibres of reciprocal connections between the medial geniculate body and the temporal lobe.
- *Corticopontine* (parietotemporopontine) fibres.

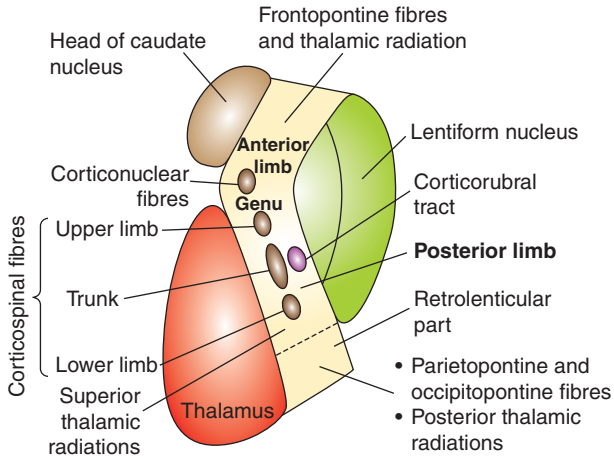


Fig. 10.4-12 Parts of internal capsule; disposition of motor fibres and thalamic radiations passing through it.

Blood supply of internal capsule

The internal capsule is supplied by the branches of middle cerebral, anterior cerebral and posterior cerebral arteries.

APPLIED ASPECTS

In internal capsule, fibres are densely crowded in a narrow area. Pyramidal fibres being compressed in this little space are particularly vulnerable to effects of even a pinpoint vascular lesion.

- Damage to internal capsule from infarction and haemorrhage is a common form or stroke, resulting in loss or decrease in sensations and movements of the opposite half of the body (hemianaesthesia and hemiplegia).
- Most common cause of the hemiplegia is the thrombosis or rupture of one of the striate branches of middle cerebral artery, which passes through the anterior perforated substance to supply the internal capsule. One of the lateral striate arteries, which is the largest of the per forating branches, is said to be particularly prone to such pathological conditions and is commonly called the artery of the cerebral haemorrhage (Charcot's artery) (Fig. 10.4-12) usually all tracts are involved causing complete contralateral hemiplegia with associated sensory loss.
- Thrombosis of the anterior choroidal artery involves the optic radiations producing contralateral hemianopia and hyperacusia.
- Thrombosis of the recurrent branch of anterior cerebral artery (Huebner artery) results in contralateral paralysis of the face and upper limbs on account of the involvement of corticonuclear fibres and adjacent pyramidal fibres for the superior extremity.

Chapter 10.5

Autonomic Nervous System

ANATOMICAL CONSIDERATIONS

- Autonomic nervous system: divisions
- General organization of ANS
- Neurons of ANS
- Physiological anatomy of sympathetic nervous system
- Physiological anatomy of parasympathetic nervous system

PHYSIOLOGICAL CONSIDERATIONS

- Autonomic neurotransmitters and receptors
- Functions of ANS: effects of autonomic nerve impulses on effector organs
- Differences between sympathetic and parasympathetic systems

APPLIED ASPECTS

- Autonomic drugs
- Autonomic failure
- Autonomic function tests

ANATOMICAL CONSIDERATIONS

AUTONOMIC NERVOUS SYSTEM: DIVISIONS

The autonomic nervous system (ANS) collects the information about the changes that take place in the internal environment (i.e. internal viscera), interprets these changes and guides the actions and gets the plan executed with the help of smooth muscles of viscera, cardiac muscles and secretory epithelium of the glandular tissues (which are the effector organs of ANS).

The word autonomous is taken from Greek words, the *autos* meaning 'self' and the *nomos* meaning 'control'. Thus, ANS is an involuntary system. Since it controls the vegetative functions, it is also called vegetative system.

Divisions of ANS

Autonomic nervous system has two main physiological as well as anatomical divisions, sympathetic and parasympathetic, each having a central and a peripheral component.

Sympathetic division, also called thoracolumbar division, consists of thoracic and lumbar chains of sympathetic ganglia.

Parasympathetic division, also called craniosacral division, consists of the ganglia associated with third, seventh, ninth and tenth cranial nerves.

Somatic versus autonomic nervous system

General arrangement of somatic and autonomic nervous system (Fig. 10.5-1) shows that:

- *Afferent (sensory) neuron* of somatic system having cell body in the dorsal root ganglion terminates in dorsal horn, while that of ANS terminates in intermediolateral horns.
- *The interneuron* (connector neuron) of somatic system has cell body in the dorsal horn and terminates in the ventral horn, while that of ANS has cell body in the intermediolateral horn and terminates in the autonomic ganglia.
- *Efferent (motor) neuron* has cell body in the ventral horn and its axon carries impulses of skeletal muscles (effector organ). The post-ganglionic neuron in ANS has cell body outside the CNS in the autonomic ganglion and its axon terminates in the visceral effector.
- There is a single efferent neuron in the somatic system, which extends from CNS to effector organ. While in the ANS, there are two efferent neuron chains between CNS and the effector organ: first efferent neuron (pre-ganglionic neuron) has its cell body in CNS while the second efferent neuron (post-ganglionic neuron) has its cell body outside the CNS in the ganglion (Fig. 10.5-1).
- Somatic motor system innervates the skeletal muscles, while the ANS innervates the smooth muscles, cardiac muscles and secretory glandular epithelium.

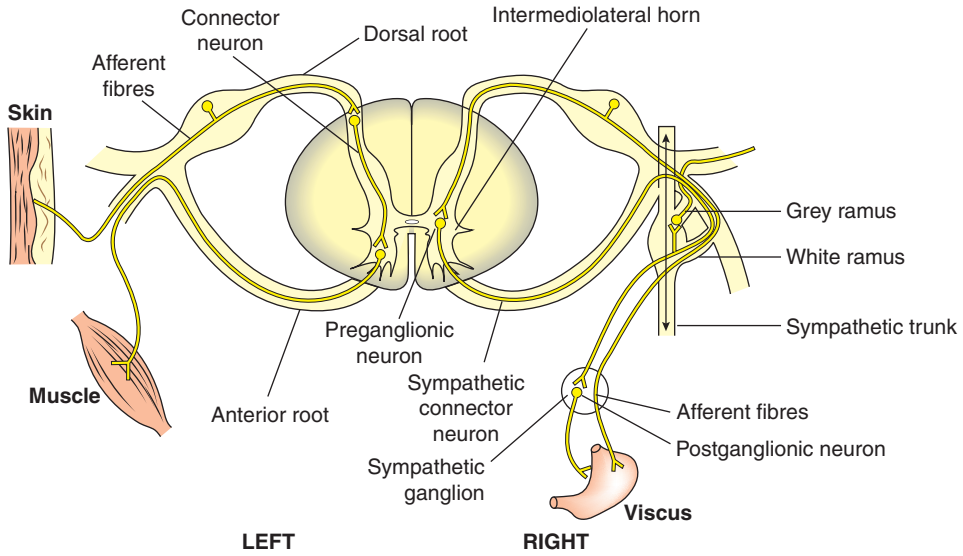


Fig. 10.5-1 General arrangement of somatic part of nervous system (on left) compared with autonomic part of nervous system (on right).

- *Neurotransmitter* released at efferent (motor) neuron ending in the somatic system is acetylcholine, while in ANS the neurotransmitter released between pre and post-ganglionic neuron is also acetylcholine, but that between post-ganglionic and effector organ depends on the component of ANS (see page 766).
- Somatic system activity always causes muscle excitation, while ANS can cause both excitation and inhibition.
- Somatic motor activity is always voluntary, while ANS motor activity is usually involuntary.

GENERAL ORGANIZATION OF THE AUTONOMIC NERVOUS SYSTEM

The ANS is organized as:

I. Autonomic areas in the cerebral hemispheres

The autonomic areas controlling visceral functions located in the cerebral hemisphere are:

- Structures included in limbic system (see page 849),
- Prefrontal cortex,
- Hypothalamus and
- Part of thalamus.

Higher brain centres, such as the limbic cortex, parts of the cerebral cortex, can influence the activity of autonomic nervous system by sending signals to the hypothalamus and lower brain area.

Hypothalamus is the site of integration of somatic, autonomic and endocrine functions. Such integration is essential for maintenance of homeostasis during exposure to stresses like extreme hot, extreme cold, stress of surgical operation, stress of injuries and haemorrhage and so on.

Since hypothalamus plays an important role in the regulation of autonomic activity, it has been called the main ganglion of the ANS. However, it is now known that the limbic cortex is equally important in the regulation of the ANS.

II. Autonomic centres in the brain stem

These are located in the reticular formation and in the general visceral nuclei of cranial nerves.

Autonomic centres in reticular formation

- *Gigantocellular* nucleus and
- *Parvocellular nuclei.*

The effects of their stimulation mediated through connections between the reticular formation and autonomic centres in the brain stem and spinal cord, but the pathways concerned are not well defined.

General visceral nuclei of cranial nerves

These include both general visceral afferent and efferent nuclei.

General visceral afferent nucleus is represented by the *nucleus of solitary tract* present in the medulla. It receives fibres carrying general visceral sensations through the vagus and glossopharyngeal nerves. Through these afferents and through connections with the reticular formation, nucleus plays an important role in the reflex control of respiratory and cardiovascular functions.

Fibres of taste (special visceral afferents) carried by the facial, glossopharyngeal and vagus nerves end in the upper part of the nucleus of the solitary tract, which is sometimes called *gustatory nucleus.*

According to some authorities some general visceral afferents end in the dorsal vagal nucleus.

General visceral efferent nuclei. These nuclei give origin to the pre-ganglionic fibres that constitute the cranial parasympathetic outflow. *The general visceral efferent nuclei* include:

- Edinger–Westphal nucleus (of oculomotor nerve) situated in the mid brain,
- Salivary nucleus (superior and inferior) located in the pons and
- Dorsal nucleus of vagus, present in the medulla.

III. Autonomic centres in the spinal cord

These are located in the intermediolateral grey column of spinal cord at two levels:

1. Neurons present in the thoracic and upper two or three lumbar segments of spinal cord (T_1–L_2) constitute the pre-ganglionic neurons of the sympathetic nervous system (thoracolumbar outflow).
2. Neurons present in the second, third and fourth sacral segments of spinal cord (S_2–S_4) are the pre-ganglionic neurons of the sacral part of parasympathetic system, which along with the cranial part constitute the craniosacral outflow.

IV. Peripheral part of ANS

This is made up of all autonomic nerves and ganglia throughout the body. It is important to stress here that there is no nerve in the body which is totally made of autonomic fibres. Hence, it is not possible to speak of autonomic nerves. In fact, autonomic fibres are intimately related to different cranial and spinal nerves.

NEURONS OF ANS

It is believed that ANS has both afferent and efferent components. The visceral afferents are sometimes called *autonomic afferents.* The autonomic efferent pathway from the spinal cord or cranial nuclei is made of two neurons: the pre-ganglionic and post-ganglionic.

Pre-ganglionic neurons. The cell body of the pre-ganglionic neuron is located in either the brain stem or spinal cord. The axon of this visceral motor neuron projects as a thinly myelinated pre-ganglionic fibre to an autonomic ganglion.

Post-ganglionic neuron. The cell body of the post-ganglionic neuron is located in the autonomic ganglion and sends an unmyelinated axon, the post-ganglionic fibre, to the visceral effector cells.

IMPORTANT NOTE

In general, sympathetic ganglia are located close to the central nervous system, whereas parasympathetic ganglia are located close to the effector tissues. Therefore, sympathetic pathway has short pre-ganglionic fibres and long post-ganglionic fibres, whereas parasympathetic pathway has long pre-ganglionic fibres and short postganglionic fibres.

PHYSIOLOGICAL ANATOMY OF SYMPATHETIC NERVOUS SYSTEM

Sympathetic pre-ganglionic neurons

The cell bodies of the sympathetic pre-ganglionic neurons are located in the intermediolateral horn of the spinal cord from level T_1 to L_2. The myelinated axons of these visceral motor neurons leave the spinal cord via the ventral root and then pass via the white rami communicantes to the paravertebral ganglia of the sympathetic trunk (Fig. 10.5-1). After reaching the sympathetic trunk pre-ganglionic fibres may pass to one of the following three destinations:

- They may terminate in the ganglion at the level of entrance by synapsing with an excitor cell in the ganglion (Fig. 10.5-2).
- They may travel up or down in the sympathetic trunk to terminate in the ganglia located at a higher or lower level (Fig. 10.5-2).
- They may travel through the sympathetic trunk and exit without synapsing via splanchnic nerve and terminate in a pre-vertebral ganglion (Fig. 10.5-2).

Pre-ganglionic fibres that innervate the adrenal medulla travel through the sympathetic trunk, exit without synapsing via greater splanchnic nerve and end directly on the cells of suprarenal medulla. These medullary cells may be regarded as modified sympathetic excitor cells that secrete epinephrine and norepinephrine into the blood stream. These secretory cells of the adrenal medulla are derived embryologically from the nervous tissue and are analogous to the post-ganglionic neurons.

Sympathetic ganglia

Sympathetic ganglia are of three types:

- Paravertebral ganglia,
- Prevertebral or collateral ganglia and
- Peripheral or terminal ganglia.

Paravertebral ganglia. Paravertebral ganglia are arranged as enlargements along the entire length of two sympathetic trunks (right and left placed on either side of vertebral column throughout its length). Paravertebral ganglia of sympathetic trunk are divided into:

- *Cervical ganglia.* These are three in number: superior, middle and inferior.
- *Thoracic ganglia.* These are 11–12 in number.
- *Lumbar ganglia* are four in number.

Note. In all there are 22 or 23 ganglia on each trunk. The inferior cervical ganglion and the first thoracic ganglion are often fused to form a large stellate ganglion.

The two sympathetic trunks end below by joining together to form a single ganglion, the *ganglion impar*.

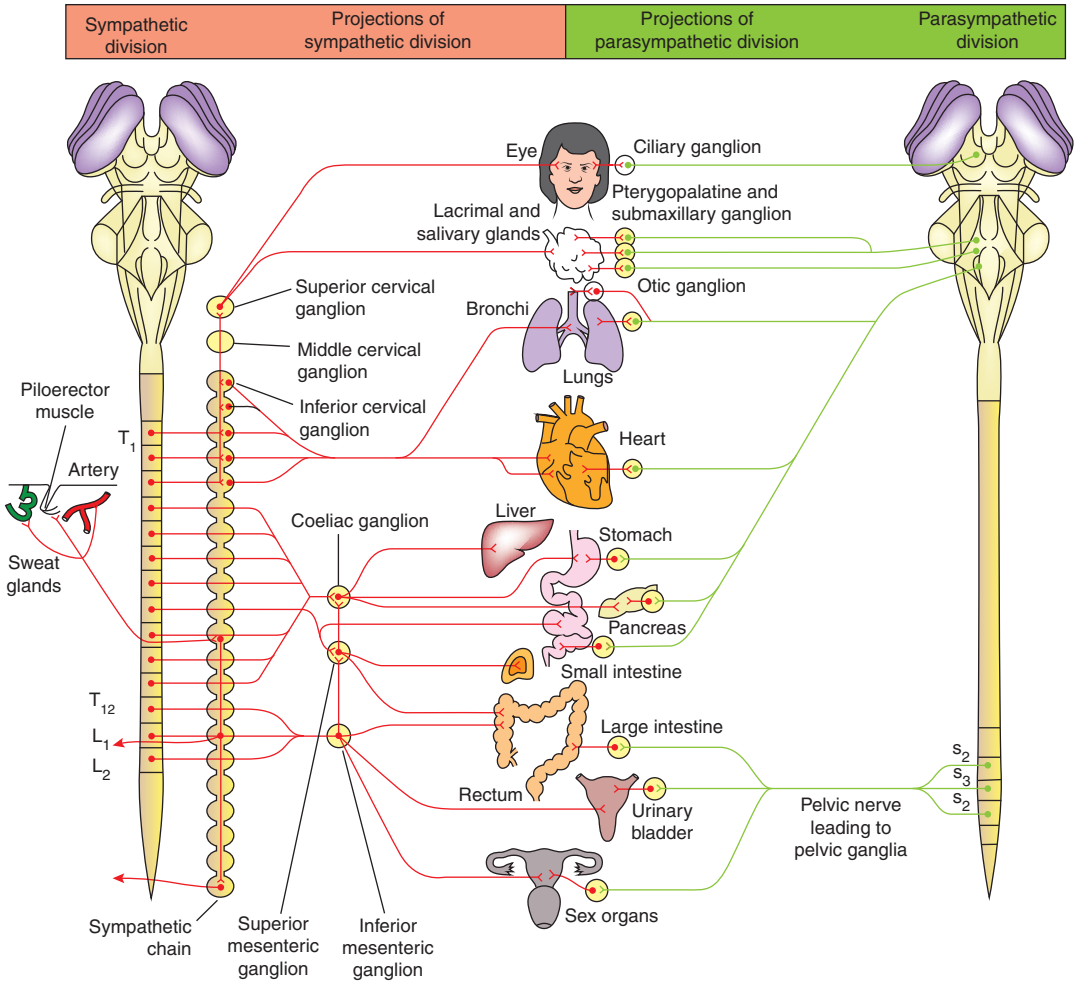


Fig. 10.5-2 Efferent part of autonomic nervous system: parasympathetic (on left) and sympathetic (on right).

Pre-vertebral or collateral ganglia are three in number (coeliac ganglion, inferior mesenteric ganglion and superior mesenteric ganglion).

Peripheral or terminal ganglia are located within or close to structures innervated by them. Heart, bronchi, pancreas and urinary bladder are innervated by the terminal ganglia.

Post-ganglionic sympathetic neurons

Sympathetic post-ganglionic neurons are located primarily in ganglia on the sympathetic trunks. Some are located in the pre-vertebral ganglia and the peripheral autonomic plexus (Figs 10.5-1 and 10.5-2). Axons arising from these neurons behave in one of the following ways:

- *The axons may pass through a grey ramus communicantes* and re-enter ventral root to reach a spinal nerve (Fig. 10.5-1). The grey rami communicantes are grey in colour because the post-ganglionic fibres are unmyelinated fibres. In the spinal nerve, the post-ganglionic fibres travel through its branches to innervate sweat glands and arrectores pilorum muscles of the skin in the region to which spinal nerve is distributed.
- *The axons may reach a cranial nerve* through a communicating branch and may be distributed through it as in the case of spinal nerve.
- *The axons may pass into a vascular branch* and may be distributed to branches of the vessel.
- Some fibres from these plexuses may pass to other structures in the neighbourhood of the vessel.
- *The axons of post-ganglionic neurons arising* in the sympathetic ganglia may travel through vascular branches and through autonomic plexus to reach some viscera (e.g. the heart).
- *The axons of post-ganglionic neurons located in the peripheral autonomic plexus innervate neighbouring viscera.* These fibres often travel to the viscera in plexuses along blood vessels. For example, fibres for the gut travel along the plexuses surrounding the branches of coeliac, superior mesenteric and inferior mesenteric arteries.

Sympathetic afferent fibres

The afferent myelinated fibres travel from the viscera through the sympathetic ganglia without synapsing (Fig. 10.5-1). They enter the spinal nerve via the white rami communicantes

Table 10.5-1 Distribution of pre-ganglionic neurons and post-ganglionic fibres

Segmental level of pre-ganglionic neurons	Area of distribution	Final distribution of post-ganglionic fibres
T_1, T_2	Head and neck	Dilator pupillae muscle, superior and inferior, Muller's muscles of eyelids, blood vessels and sweat glands.
T_3, T_4	Thoracic viscera	Heart, oesophagus, trachea, bronchi and lungs
T_5–T_9	Upper limb	Blood vessels, sweat glands and arrectores pilorum muscles
T_{10}–L_2	Lower limb	Blood vessels, sweat glands and arrectores pilorum muscles.
T_6–T_{12}	Upper abdominal viscera	Gastrointestinal tract, Liver, spleen capsule, adrenal medulla and urinary tract
L_1, L_2	Lower abdominal viscera	Bladder, uterus, fallopian tubes (or testis, vas deferens, seminal vesicles and prostate)
T_1–T_{12}	Thoracic and abdominal parities	Blood vessels, sweat glands and arrectores pilorum muscles

and reach their cell bodies in the posterior root ganglia of the corresponding spinal nerve. The central axons then enter the spinal cord and may form the afferent component of a local reflex arc. Others may pass up to higher autonomic centres in the brain.

Distribution of sympathetic pre-ganglionic neurons and post-ganglionic fibres

The distribution of pre-ganglionic neurons and post-ganglionic fibres is shown in Table 10.5-1 and Fig. 10.5-2.

PHYSIOLOGICAL ANATOMY OF PARASYMPATHETIC NERVOUS SYSTEM

Parasympathetic pre-ganglionic neurons

The parasympathetic fibres form the *craniosacral outflow*, consisting of cranial parasympathetic outflow and sacral parasympathetic outflow.

I. Cranial parasympathetic outflow

The cell bodies of the neurons which give rise to pre-ganglionic parasympathetic fibres are located in the general visceral efferent nuclei. These pre-ganglionic fibres end in the peripheral ganglia associated with the branches of cranial nerves. Post-ganglionic fibres arising in these ganglia supply smooth muscles or glands. Cranial parasympathetic outflow can be further divided into:

- Mid brain or tectal outflow and
- Bulbar outflow.

1. Mid brain or tectal outflow. The general visceral efferent nucleus associated with the mid brain outflow is Edinger–Westphal nucleus.

Edinger–Westphal nucleus. It lies in the mid brain and is closely related to the oculomotor nucleus complex.

- *Pre-ganglionic fibres* arising from the Edinger–Westphal nucleus pass through oculomotor (third cranial) nerve and relay in the ciliary ganglion.
- *Ciliary ganglion* is a peripheral parasympathetic ganglion placed in the course of oculomotor nerve.
- *Post-ganglionic fibres* arising in the ciliary ganglion pass through the short ciliary nerves and supply the sphincter pupillae and the ciliary muscle.

2. Bulbar outflow. The general visceral efferent nuclei associated with bulbar outflow are:

- Superior salivary nucleus,
- Lacrimal nucleus,
- Inferior salivary nucleus and
- Dorsal vagal nucleus.

(i) Superior salivary nucleus. *Pre-ganglionic fibres* arising from the superior salivary nucleus enter the facial (seventh cranial) nerve and ultimately relay in the *submandibular ganglion*. Post-ganglionic fibres pass to the submandibular and sublingual salivary glands to which they are secretomotor.

(ii) Lacrimal nucleus of seventh cranial nerve sends pre-ganglionic fibres to the *pterygopalatine* (sphenopalatine) ganglion. The post-ganglionic fibres reach the lacrimal gland to which they are secretomotor.

(iii) Inferior salivary nucleus sends pre-ganglionic fibres into the glossopharyngeal (ninth cranial) nerve. These fibres relay in the otic ganglion, from where the post-ganglionic fibres go to parotid gland to which they are secretomotor.

(iv) Dorsal (motor) nucleus of the vagus. About 75% of all parasympathetic fibres arise from dorsal nucleus.

Pre-ganglionic fibres travelling in the vagus nerve end in ganglia (or nerve plexuses) closely related to the visceral organs, such as heart, lungs, bronchi, oesophagus, stomach,

small intestine and large intestine up to two-thirds of transverse colon. *Post-ganglionic fibres* arise in these ganglia and run a short course to supply smooth muscles and glands in these organs.

II. Sacral parasympathetic outflow

Pre-ganglionic fibres. Cell bodies of the pre-ganglionic neurons, which constitute the sacral parasympathetic outflow, are located in the intermediolateral grey horn of second, third and fourth sacral segments (S_2, S_3 and S_4) of spinal cord (Fig. 10.5-2). Their axons form the pre-ganglionic fibres which pass out through the ventral spinal root of corresponding nerves. These axons leave the spinal nerves to form the pelvic splanchnic nerves, which end in the pelvic autonomic plexuses.

Post-ganglionic fibres. The post-ganglionic neurons are located in the pelvic autonomic plexuses close to or within the viscera. Their axons (post-ganglionic fibres) run a very short course to supply the concerned pelvic viscera. These fibres also supply the rectum, the sigmoid colon, the descending colon and the left one-third of transverse colon.

Parasympathetic afferent fibres

The afferent myelinated fibres travel from viscera to their cell bodies located either in the sensory ganglia of the cranial nerves or in the posterior root ganglia of the sacrospinal nerves. The central axons then enter the central nervous system and take part in the formation of local reflex arc, or pass to higher centres of the ANS.

The afferent component of ANS is identical to the afferent component of somatic nerves and forms part of the general afferent segment of the entire nervous system. The nerve endings in the autonomic afferent component may not be activated by such sensations as heat or touch but instead by stretch or lack of oxygen. Once the afferent fibres gain entrance to the spinal cord or brain, they are thought to travel alongside, or are mixed with the somatic afferent fibres.

PHYSIOLOGICAL CONSIDERATIONS

AUTONOMIC NEUROTRANSMITTERS AND RECEPTORS

Neurotransmitters of ANS (Fig. 10.5-3)

Parasympathetic fibres

1. Pre-ganglionic fibres: acetylcholine
2. Post-ganglionic fibres: acetylcholine

Sympathetic fibres

1. Pre-ganglionic fibres: acetylcholine
2. Post-ganglionic fibres:

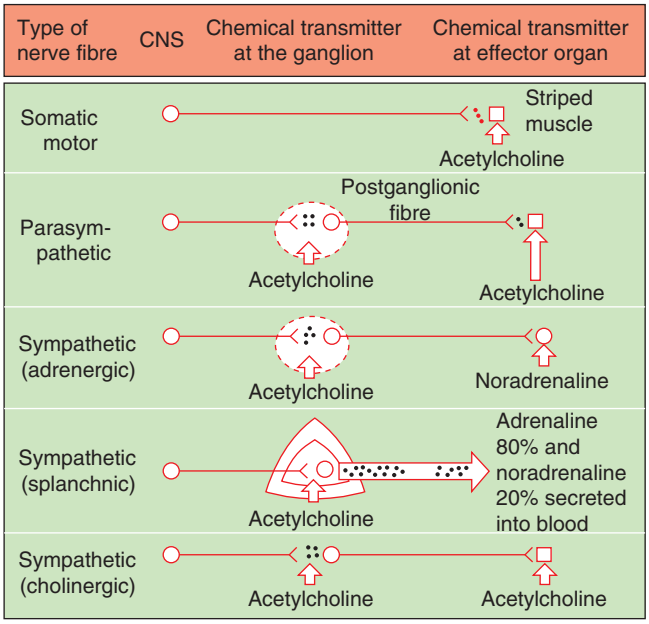


Fig. 10.5-3 Neurotransmitters of peripheral somatic and autonomic nervous system.

- Adrenergic fibres:
  Norepinephrine (mainly), or epinephrine (All post-ganglionic sympathetic fibres other than cholinergic).

- Cholinergic fibres:
  Acetylcholine (the post-ganglionic sympathetic cholinergic nerve fibres supplying sweat glands, blood vessels in heart and skeletal muscles).

Thus:

- All pre-ganglionic fibres (sympathetic as well as parasympathetic) release acetylcholine.
- All post-ganglionic parasympathetic fibres release acetylcholine.
- Most post-ganglionic sympathetic (adrenergic) fibres release norepinephrine.
- A few post-ganglionic sympathetic (cholinergic) fibres release acetylcholine.

Note. For details about neurotransmitters see page 787.

Autonomic receptors

The autonomic neurotransmitters (acetylcholine or norepinephrine) produce their effects on the organs by combining with specific protein molecules known as receptors, which are of following types:

1. Cholinergic receptors

On the basis of their pharmacologic properties, these are of two types:

- Nicotinic receptors and
- Muscarinic receptors.

(i) Nicotinic receptors

Location. These receptors are located in/at:

- Autonomic ganglia of sympathetic and parasympathetic nervous system,
- Neuromuscular junction and
- Adrenal medulla.

The receptors at these locations are similar but not identical.

Activation. Nicotinic receptors are activated by:

- Acetylcholine (Ach) and
- Nicotine.

Effect. These receptors produce excitation.

Blockage. Ganglion blockers (e.g. hexamethonium, trimethaphan) block the nicotinic receptors for Ach in the autonomic ganglia, but not at the neuromuscular junction.

Mechanism of action. Ach binds to α subunit of the nicotinic cholinergic receptors. The nicotinic Ach receptors are also ion channels for Na^+ and K^+.

(ii) Muscarinic receptors

Location. Muscarinic receptors are located in the:

- Heart,
- Smooth muscles (except vascular smooth muscle) and
- Glands.

Activation. These receptors are activated by:

- Acetylcholine (Ach) and
- Muscarine.

Effect produced by their stimulation.

- *Inhibitory in the heart,* e.g. decreased heart rate and decreased conduction velocity in atrioventricular (AV) node.
- *Excitatory* in smooth muscle and glands (e.g. increased gastrointestinal motility and increased secretion).

Blockage. Muscarinic receptors for acetylcholine are blocked by atropine.

Mechanism of action

- *In heart sinoatrial (SA) node,* these receptors cause inhibition of adenylyl cyclase, which leads to opening of K^+ channels, slowing of the rate of spontaneous depolarization and decreased heart rate.
- *In smooth muscle and glands,* these receptors act by formation of inositol 1,3,5-triphosphate (IP_3) and increase in intracellular Ca^{2+}.

2. *Adrenergic receptors*

On the basis of their pharmacologic properties, adrenergic receptors are of two types:

- *Alpha (α)* adrenergic receptors (which are of further two types: α_1 and α_2) and
- *Beta (β)* adrenergic receptors (which are of further three types: β_1, β_2 and β_3).

(i) α_1 receptors

Location. α_1 receptors are located on:

- Vascular smooth muscles of skin and splanchnic regions.
- Gastrointestinal and bladder sphincters and
- Radial muscles of the iris.

Effect. These receptors produce excitation, e.g. contraction or constriction.

Catecholamine sensitivity. α_1 receptors are equally sensitive to norepinephrine and epinephrine, but only norepinephrine is present in concentrations that are high enough to activate α_1 receptors.

Mechanism of action. These receptors act by formation of IP_3 and increase in intracellular Ca^{2+}.

(ii) α_2 receptors

Location. α_2 receptors are located in:

- Pre-synaptic nerve terminals,
- Platelets,
- Fat cells and
- Walls of the gastrointestinal tract.

Effect. Often produce inhibition (e.g. relaxation or dilatation).

Mechanism of action. α_2 receptor causes inhibition of adenylyl cyclase and decrease in cyclic adenosine monophosphate (cAMP).

(iii) β_1 receptors

Location. β_1 receptors are located in the:

- Sinoatrial node,
- Atrioventricular (AV) node and
- Ventricular muscles of the heart.

Effect. These receptors produce excitation (e.g. increased heart rate, increased conduction velocity and increased contractility).

Catecholamine sensitivity. β_1 receptors are sensitive to both norepinephrine and epinephrine, and are more sensitive than the α_1 receptors.

(iv) β_2 receptors

Location. β_2 receptors are located on:

- Vascular smooth muscle of skeletal muscle,
- Bronchial smooth muscle,
- Walls of the gastrointestinal tract and
- Bladder.

Effect. These receptors produce relaxation (e.g. dilation of vascular smooth muscle, dilation of bronchioles and relaxation of bladder wall).

Sensitivity to epinephrine is more than to norepinephrine. These are more sensitive to epinephrine than the α_1 receptors (e.g. when small amounts of epinephrine are released from the adrenal medulla, vasodilation β_2 effect) occurs, when larger amounts of epinephrine are released from the adrenal medulla, vasoconstriction α_1 effect) occurs.

Mechanism of action. Same as for β_1 receptors.

(v) β_3 receptors. These receptors are located on the adipose tissue and causes lipolysis.

Mechanism of action. *B_3 receptors cause increase in cAMP.*

Note. The type of adrenergic receptors present in various organs and the effects produced by their stimulation are depicted in Table 10.5-2.

FUNCTIONS OF AUTONOMIC NERVOUS SYSTEM: EFFECTS OF AUTONOMIC NERVE IMPULSES ON EFFECTOR ORGANS

General principles

- Autonomic nervous system controls the various vegetative functions, which are beyond voluntary control and thus plays an important role in maintaining the constant internal environment (homeostasis).
- Most of the visceral organs have dual innervation, i.e. are supplied by both sympathetic and parasympathetic divisions of ANS. The two divisions produce antagonistic effects on each organ and provide a very fine degree of control over the effector organ. When the fibres of one division supplying to an organ are sectioned or affected by lesion, the effects of fibres from other division on the organ become more prominent.
- Some of the visceral organs are innervated by one division of ANS only; e.g.
  - Uterus, adrenal medulla and most of the arterioles are innervated by sympathetic division only.
  - Glands of stomach and pancreas are innervated by parasympathetic division only.
- In the case of sphincter's muscles, both adrenergic and cholinergic innervations are excitatory, but one supplies the constrictor component of the sphincter and other the dilator.
- *Effects of acetylcholine,* i.e. of localized cholinergic discharge are generally discrete and short lasting, because Ach is rapidly removed from the nerve endings due to high concentration of acetylcholine esterase at cholinergic nerve endings.
- *Effects of norepinephrine* are more prolonged than Ach, as it spreads further. In the blood, epinephrine and dopamine come from the adrenal medulla, while norepinephrine diffuses from the adrenergic nerve endings.

Epinephrine versus norepinephrine (also see page 598).

- Epinephrine acts equally on α and β receptors, and has a special property of stimulating β_2 receptors. While norepinephrine acts mainly on α receptors and also on β_1 receptors but has no action on β_2 receptors.

Effects of stimulation of sympathetic and parasympathetic division of ANS

Responses of effector organs to autonomic nerve impulses are summarized in Table 10.5-2.

DIFFERENCES BETWEEN SYMPATHETIC AND PARASYMPATHETIC SYSTEMS

As summarized in Table 10.5-2, sympathetic and parasympathetic systems produce antagonistic effects on each organ of the body. The main differences between sympathetic and parasympathetic systems are depicted in Table 10.5-3.

APPLIED ASPECTS

A few important considerations about applied aspect of ANS are:

- Autonomic drugs,
- Autonomic failure and
- Autonomic function tests.

AUTONOMIC DRUGS

Autonomic drugs exert their effects by action on the autonomic receptors directly or indirectly. These include:

- Sympathomimetic drugs,
- Sympatholytic drugs or sympathetic blockers,
- Parasympathomimetic drugs and
- Parasympatholytic drugs or parasympathetic blockers.

Sympathomimetic drugs

Sympathomimetic drugs also called adrenaline-like drugs, when administered in the body produce effects similar to the effects of sympathetic nerve stimulation.

Examples of these drugs are adrenaline, noradrenaline, phenylephrine, isoproterenol and albuterol.

Sympathetic blockers

Sympathetic blockers or sympatholytic drugs block the actions of sympathetic neurotransmitters. Mechanisms by which sympatholytic drugs act are:

- Prevention of synthesis and storage of NE, e.g. reserpine.
- Prevention of release of NE, e.g. guanethidine.
- Blockage of α receptors, e.g. phentolamine

Table 10.5-2 Responses of effector organs to sympathetic and parasympathetic stimulation

S. No.	Effector organ	Parasympathetic effect	Sympathetic effect	
			Receptor type	Response
1.	***Eyes***			
	• Dilator pupillae muscle	–	α	Contraction (mydriasis)
	• Sphincter pupillae muscle	Contraction (meiosis)	–	–
	• Ciliary muscle	Contraction (produces accommodation for near vision)	β_2	Relaxes (flattens lens for far vision)
2.	***Heart***			
	• SA node	↓ Heart rate, vagal arrest	β_1 & β_2	↑ Heart rate
	• Atria	↓ Contractility	β_1 & β_2	↑ Contractility
		↓ Conductivity		↑ Conductivity
	• AV node and conduction system	↓ Conduction velocity	β_1 & β_2	↑ Conduction velocity
	• Ventricles	↓ Contractility	β_1 & β_2	↑ Contractility
3.	***Arterioles***			
	• Coronary		α_1 & α_2	Constriction
			β_2	Dilatation
	• Cutaneous and mucosal	No supply	α_1 & α_2	Constriction
	• Skeletal muscle	No supply	α_1	Constriction
			β_2	Dilatation
	• Cerebral	Dilatation	α_1	Constriction
	• Pulmonary	Dilatation	α_1	Constriction
			β_2	Dilatation
	• Abdominal viscera	No supply	α_1	Constriction
	• Renal	No supply	α_1 & α_2	Constriction
			β_1 & β_2	Dilatation
	• Salivary glands	Dilatation	α_1 & α_2	Constriction
4.	***Systemic veins***	No supply	α_1 & α_2	Constriction
			β_2	Dilatation
5.	***Lungs***			
	• Bronchial muscles	Contraction	β_2	Relaxation
			α_1	Inhibition
	• Bronchial glands	Stimulation	β_2	Stimulation
6.	***Salivary glands***	Stimulation (profuse watery secretion)	α_1	Stimulation (thick viscous secretion)
7.	***Stomach***			
	• Motility and tone	Increases	$\alpha_1, \alpha_2, \beta_2$	Decreases
	• Sphincters	Relaxation	α_1	Contraction
	• Secretion	Stimulation	α_2	Inhibition
8.	***Gall bladder***	Contraction	β_1	Relaxation
9.	***Liver***	–	α_1, β_2	Glycogenolysis
10.	***Pancreas***			
	• Exocrine glands	Stimulates secretion	α_1	Inhibits secretion
	• Endocrine glands	–	α_2	Inhibits insulin secretion
			β_2	Stimulates glucagon
11.	***Spleen capsule***		α_1	Contraction
			β_2	Relaxation
12.	***Adrenal medulla***	Secretion of epinephrine and norepinephrine		
13.	***Urinary bladder***			
	• Detrusor muscle	Contraction	β_2	Relaxation (usually)
	• Sphincter	Relaxation	α_1	Contraction

Table 10.5-2 Continued

S. No.	Effector organ	Parasympathetic effect	Sympathetic effect	
			Receptor type	Response
14.	***Uterus***	Variable	α_1 β_2	Contraction (pregnant) Relaxation (non-pregnant)
15.	***Male sex organ***	Erection	α_1	Ejaculation
16.	***Lacrimal glands***	Secretion		
17.	***Skin*** • Pilomotor muscle • Sweat glands	 – Generalized (cholinergic sweating)	 α_1 α_1	 Contraction (erection of hair) Localized (adrenergic) sweating
18.	***Nasopharyngeal glands***	Secretion	–	–
19.	***Adipose tissue***	–	$\alpha_1, \beta_1, \beta_3$	Lipolysis, release of FFA
20.	***Juxtaglomerular cells***	–	β_1	Increased renin secretion
21.	***Pineal gland***	–	β_1	Increased melatonin synthesis
22.	***Skeletal muscles***	–	β_2	Increased glycogenolysis
23.	***Basal metabolic rate***	–	β_2	Increased
24.	***Mental activity***	–		Increased

- Blockage of β receptors, e.g. propranolol, metoprolol, timolol, etc.
- Blockage of transmission of nerve impulse through sympathetic ganglion (ganglion blockers), e.g. hexamethonium and pentolinium.

Parasympathomimetic drugs

Parasympathomimetic drugs, also known as acetylcholine-like drugs, when administered in the body produce effects similar to the effect of parasympathetic nerve stimulation. Depending upon their mechanism of action parasympathomimetic drugs are of following types:

- Drugs acting on muscarinic receptors, e.g. pilocarpine, methacholine.
- Drugs prolonging the action of acetylcholine, e.g. neostigmine and physostigmine, which inhibit the activity of acetylcholine esterase.

Parasympathetic blockers

Parasympathetic blockers, also called parasympatholytic drugs, block the actions of parasympathetic neurotransmitters by blocking the muscarinic receptors. Examples of parasympathetic blockers are atropine, homatropine, scopolamine, cyclopentolate and tropicamide.

AUTONOMIC FAILURE

Types. Autonomic failure is of two types:

- *Primary autonomic failure* from an unexplained (primary) autonomic neuronal degeneration.
- *Secondary autonomic failure* occurs secondary to some general medical disorders. Diabetes mellitus is the most common cause of secondary autonomic dysfunction.

Features of autonomic failure (primary or secondary) are:

- *Cardiovascular features* include tachycardia and orthostatic hypotension.
- *Sudomotor features* are anhidrosis and heat intolerance.
- *Gastrointestinal features* include constipation, occasional diarrhoea and dysphagia.
- *Urinary features* are nocturia, frequency, urgency, incontinence and retention of urine.
- *Reproductive* organ problems include erectile and ejaculation failure.
- *Ocular* features include miosis and enophthalmos.

Horner's syndrome

Horner's syndrome refers to ipsilateral *oculosympathetic paresis* due to any cause. Its common causes are Pancoast's tumour of the lung, malignancy of cervical lymph nodes pressing on the cervical sympathetic chain.

Clinical features of Horner's syndrome are:

- *Ptosis* (drooping down of upper eyelid) due to paralysis of Muller's muscle of upper eyelid.
- *Miosis* (small pupil) due to paralysis of dilator pupillae muscle.
- *Facial anhidrosis,* i.e. reduced sweating on the ipsilateral face and neck.

Table 10.5-3 Main differences between sympathetic and parasympathetic system

Feature	Sympathetic system	Parasympathetic system
Location	Cell bodies of pre-ganglionic neurons are located in intermediolateral horn of T_1–L_2 or L_3 spinal segments, so also called thoracolumbar outflow.	Cell bodies of pre-ganglionic neurons are located in: • *Cranial nuclei* associated with third, seventh, ninth and tenth cranial nerves (cranial outflow), and • Intermediolateral horn of S_2–S_4 spinal segments (sacral outflow). So, it is also called craniosacral outflow.
Components and ganglia	• Components are consolidated • Ganglia are linked up to form a chain	• Components are isolated. • Ganglia remain isolated.
Pre-ganglionic fibres	Are short, myelinated and end in paravertebral or prevertebral ganglia	Are long, myelinated and end on short post-ganglionic neurons located on or near the viscera.
Post-ganglionic fibres	• Long • Non-myelinated	• Short • Myelinated
Neurotransmitter • Pre-ganglionic fibres • Post-ganglionic fibres	 • Cholinergic • Mostly adrenergic	 • Cholinergic • Cholinergic
Area of effect	Pre-ganglionic fibres branch, enter several ganglia and transmit nerve impulse to many post-ganglionic fibres. So, sympathetic activity is spread over many segments	Pre-ganglionic fibres do not branch, each enters a single ganglion and transmits nerve impulses to a single post-ganglionic fibre. Therefore, parasympathetic activity is localized, i.e. target is usually a single organ or system.
Functions	Mass sympathetic discharge usually occurs in threatening situation, i.e. it prepares the individual to cope with the emergency. It causes **flight or fight** reactions characterized by: • Dilatation of pupil • Increased heart rate • Increased blood pressure (providing better perfusion of the vital organs and muscles) • Constriction of cutaneous arterioles (which limits blood loss from wounds, if any) • Increased alertness and arousal due to decreased threshold in the reticular formation • Increased blood glucose and FFA levels (supplying more energy). Because of these actions sympathetic system is also sometimes called *catabolic nervous system*	Unlike sympathetic nervous system, the functions of parasympathetic system are discrete and each function is separately regulated. This system is concerned with vegetative aspect of day-to-day living. For example, its action favours: • Digestion and absorption of food, increased activity of intestinal musculature and increased gastric secretion and pyloric relaxation. • Micturition, • Pupillary constriction and • Bradycardia Since, parasympathetic system decreases the rate of metabolism, it is also called *anabolic nervous system*.

AUTONOMIC FUNCTION TESTS

A. Tests of cardiovascular autonomic function

1. Valsalva's manoeuvre. After closing both the nostrils, patient is made to blow into a tube connected to sphygmomanometer and maintain air pressure at 40 mm Hg for 15 s. The ECG recording is done during 15 s following the Valsalva manoeuvre. Normal response consists of tachycardia during strain and bradycardia after release (Fig. 10.5-4) (see page 283).

Valsalva ratio. The ratio between longest R–R interval (after the strain) and shortest R–R interval (during the strain) is known as the Valsalva ratio. Normal Valsalva ratio is > 1.20. In autonomic neuropathy, Valsalva ratio is < 1.20.

2. Heart rate variation during deep breathing. While recording ECG, patient is asked to inhale deeply for 5 s followed by exhalation for 5 s alternately for six times. The ratio between longest R–R interval during expiration and the shortest R-R interval during inspiration (E/1 ratio) in

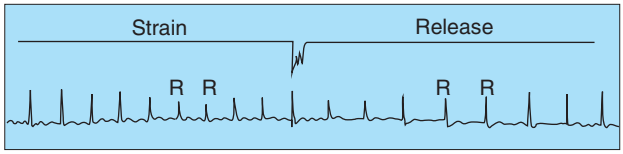


Fig. 10.5-4 Effect of Valsalva manoeuvre on heart rate.

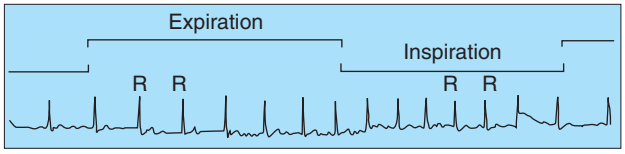


Fig. 10.5-5 Heart rate variations during deep breathing.

each respiratory cycle is calculated and averaged for the total record. Normal ratio is >1.20 (Fig. 10.5-5). In autonomic dysfunction, E/1 ratio is < 1.20.

3. Heart rate response to standing. Normally, a change of posture from supine to standing results in mild increase in heart rate (HR). The ratio between the HR on standing and in supine posture is >1.04. In autonomic neuropathy, this ratio is 1.00, i.e. there occurs no change in HR with posture change.

4. Blood pressure response to standing. Normally, a change posture from supine to standing leads a slight fall in systolic blood pressure (SBP) which is never more than 10 mm Hg. In autonomic dysfunction, this fall in SBP on change of posture is 30 mm Hg or even more. This is called *orthostatic hypotension*.

5. Blood pressure response to sustained hand grip. Patient is asked to maintain hand grip or a hand grip dynamometer at 30% of the maximum voluntary contraction for 5 min. Blood pressure is recorded just before and at the end of hand grip. Normally, the diastolic blood pressure (DBP) shows an increase by more than 15 mm Hg. In autonomic dysfunction, the rise in DBP is always less than 10 mm Hg.

B. Test of sudomotor function

Evaluation of sweating response to heat exposure tests the sudomotor functions. This test is performed by exposing the patient to electric heater till his body temperature is raised by 1 °C; and the sweating response is studied by demarcating the area of sweating with the help of iodine starch or alizarin red, or quinizarin powders, which change colour when moist.

C. Tests of pupillary functions

Pupillary function tests are specifically useful in detecting sympathetic denervation of iris (e.g. in Horner's syndrome). Commonly performed tests are:

1. Cocaine test. Cocaine prevents the reuptake of NE at the adrenergic synapse and thus when 4% cocaine is instilled in both eyes, the normal pupil will dilate but the Horner's pupil will not.

2. Adrenaline test. When adrenaline 1 in 1000 strength or 1% noradrenaline is instilled in both eyes, Horner's pupil dilates more than the normal due to denervation hypersensitivity in the involved eye.

D. Tests of bladder function

In autonomic dysfunction, a cystometrogram reveals:

- Absence of accommodation of urinary bladder in response to bladder filling and
- Absent or poor voluntary bladder contraction when asked to micturate. The bladder capacity may be increased to 1 L in advanced cases of autonomic neuropathy.

Chapter 10.6

Meninges, Cerebrospinal Fluid, Blood–Brain Barrier and Cerebral Blood Flow

MENINGES OF THE BRAIN

The brain is enclosed within the cranial cavity by three concentric connective tissue layers: pia mater, arachnoid mater and dura mater, which constitute the meninges of the brain (Fig. 10.6-1).

Pia mater

Pia mater, covering closely and continuously the external surface of the brain, is a thin and highly vascular membrane. Folds of pia mater enclose tufts of capillaries called choroid plexuses to form tela choroidea in relation to the ventricles of brain.

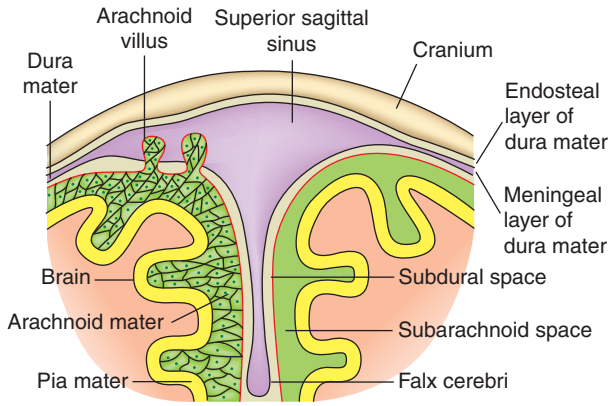


Fig. 10.6-1 Meninges of brain.

Arachnoid mater

Arachnoid mater is connected to the pia mater by many filamentous fibres. Subarachnoid space between these two layers is filled with cerebrospinal fluid (CSF).

Dura mater

Dura mater is composed of two layers: outer endosteal and inner meningeal. These are fused except where folds form (e.g. falx cerebri) or venous sinuses (e.g. superior sagittal sinus) are enclosed between them. *Subdural space* separates the dura mater from the arachnoid mater. The arachnoid mater has minute protrusions (*arachnoid villi*), which pass through fenestrae in the dura mater and project into the venous sinuses to allow escape of CSF into the venous sinuses.

CEREBROSPINAL FLUID

Cerebrospinal fluid cushions the brain and along with blood–brain barrier, the buffering function of neuroglia and regulation of central nervous system (CNS) circulation controls the extracellular environment of neurons. Within the substance of brain in the *ventricular system,* there are series of spaces filled with CSF.

Composition, volume and pressure of CSF

Composition of CSF. The extracellular fluid within the CNS communicates directly with the CSF. Thus, the composition

of CSF indicates the composition of the extracellular environment of the neurons in the brain and spinal cord. The composition of CSF vis-a-vis blood is depicted in Table 10.6-1. The CSF differs from blood in having a lower concentration of K^+, glucose, and protein and a higher concentration of Na^+ and Cl^-. Cerebrospinal fluid normally lacks blood cells. The increased concentration of Na^+ and Cl^- enables the CSF to be isotonic to blood, despite the much lower concentration of proteins in the CSF.

CSF volume and pressure. The cranial cavity contains about 140 mL CSF, 100 mL blood and 200 mL of extracellular fluid in the brain which weighs about 1350 g. Thus, the extracellular fluid space in the cranial cavity totals approximately 440 mL.

Table 10.6-1 Composition of CSF vis-a-vis blood

Constituent	Lumbar CSF	Blood
Na^+ (mEq/L)	148	136–145
K^+ (mEq/L)	2.9	3.5–5
Ca^{2+} (mEq/L)	2.3	4.7
Cl^- (mEq/L)	120–130	100–106
HCO_3^- (mEq/L)	25.1	24.8
Glucose (mg/dL)	50–75	70–100
Protein (mg/dL)	15–45	6.8×10^3
pH	7.3	7.4
Osmolality (mOsm/kgH_2O)	289	289

The volume of CSF within the cerebral ventricles is approximately 40 mL, and that in the subarachnoid space is about 100 mL. The pressure in the CSF column is about 120–180 mm H_2O when a person is recumbent. Rate of CSF formation (about 0.35 mL/minute) is independent of CSF pressure as well as systemic blood pressure.

Formation, circulation and absorption of CSF

Formation of CSF. The CSF is mainly formed by the choroidal plexuses, which are covered by specialized ependymal cells. The *choroidal plexuses* are located in the cerebral ventricles (lateral, third and fourth). About 500 mL of CSF is secreted per day.

Circulation of CSF (Fig. 10.6-2). Cerebrospinal fluid formed in the lateral ventricles passes through the interventricular foramina (of Monro) into the third ventricle. Thence the fluid flows through the cerebral aqueduct (of Sylvius) into the fourth ventricle. From fourth ventricle, some CSF passes into the central canal of spinal cord, but most escapes into the subarachnoid space (surrounding the brain and spinal cord) through the median aperture (foramen of Magendie) of fourth ventricle and the two lateral apertures of fourth ventricle (foramina of Luschka).

Subarachnoid cistern refers to the regions where subarachnoid space is distended to form pools of CSF. An example is the *lumbar cistern,* which surrounds the lumbar

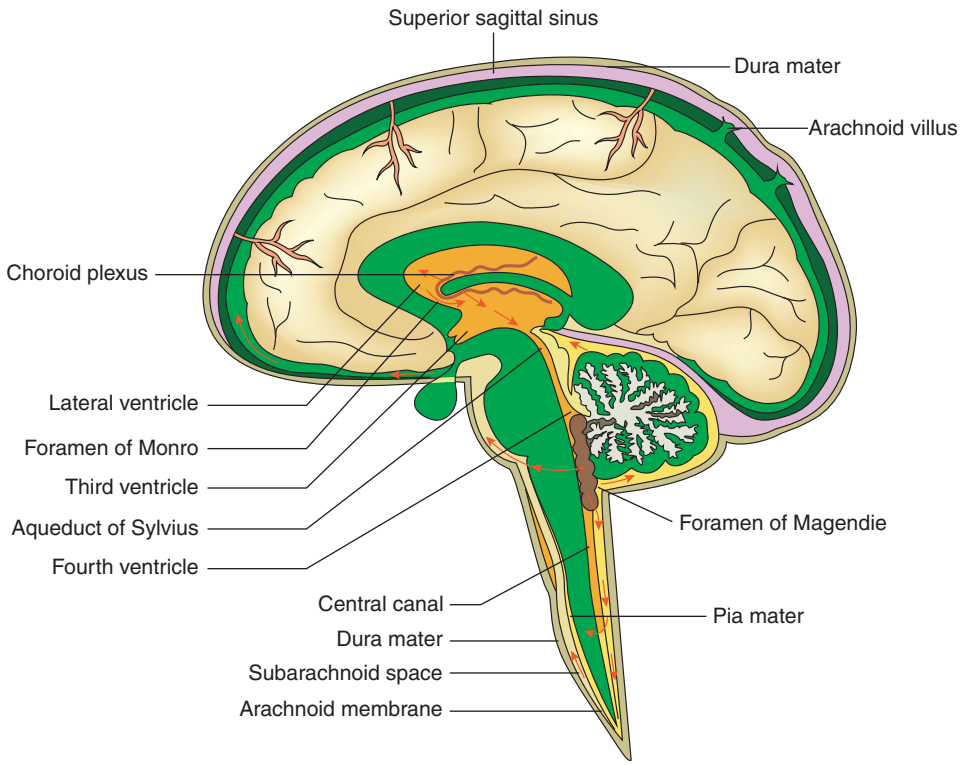


Fig. 10.6-2 Circulation of cerebrospinal fluid. Arrows indicate the direction of flow.

and sacral spinal roots below the level of termination of spinal cord. The lumbar cistern is the target for lumbar puncture, a procedure used clinically to sample the CSF.

Absorption of CSF. A large part (80%) of CSF is removed by bulk flow through the valvular *arachnoid villi* into the dural venous sinuses in the cranium. Unlike rate of formation, the absorption rate of CSF is a direct function of the CSF pressure because:

- Hydrostatic pressure of CSF is more than the venous pressure in venous sinuses.
- High plasma protein levels by their osmotic effect favour CSF absorption.

A small part (20%) may pass along the sheaths of cranial nerves and drains into the cervical lymphatics and perivascular spaces.

Functions of CSF

- Protection to CNS by acting as a 'water-jacket' as it absorbs shock in the event of blow.
- Removal of waste products of brain metabolism.
- Regulates extracellular environment for the neurons of central nervous system.
- Transports hormones and hormone releasing factors.

Clinical applications

Hydrocephalus

Hydrocephalus refers to an abnormal accumulation of CSF in the cranium.

Causes of hydrocephalus include:

- Obstruction to CSF circulation,
- Excessive production of CSF and
- Interference with absorption of CSF.

Types of hydrocephalus are:
1. Internal or non-communicating hydrocephalus occurs when obstruction is within the ventricular system or in the roof of fourth ventricle. It results in the dilatation of the ventricles.

2. External or communicating hydrocephalus occurs when obstruction is in subarachnoid space or arachnoid villi. In it excess fluid is mainly in the subarachnoid space. The arachnoid granulations often suffer from a moderate obstruction in patients suffering from cerebral meningitis or haemorrhage into the subarachnoid space.

Lumbar and cisternal puncture

Lumbar puncture refers to the tapping of CSF from the lumbar cistern. Cerebrospinal fluid examination is required in many disorders of CNS. It is performed by inserting a needle in between the L_2 and L_3 or L_3 and L_4 vertebrae into the subarachnoid space within vertebral canal, as there is no risk of damage to spinal cord as it ends at the level of first lumbar vertebra.

Cisternal puncture refers to the tapping of CSF from the cisterna magna. To do this, a needle is passed through the posterior atlanto-occipital membrane forwards and upwards to a depth of 4.5 cm from the surface.

IMPORTANT NOTE

Removal of CSF during lumbar puncture sometimes causes severe headache afterwards. This happens due to stimulation of pain fibres due the traction effect.

Measurement of CSF pressure

Pressure of CSF can be measured during lumbar puncture by connecting a glass tube to a spinal needle. Spinal fluid is allowed to rise in the tube, the level of CSF height (mm) in the glass tube above the level of spinal needle will give CSF pressure in cm H_2O.

BLOOD–BRAIN BARRIER AND BLOOD–CSF BARRIER

Blood–brain barrier

Blood–brain barrier restricts the movement of large molecules and highly charged ions from the blood into the brain and spinal cord. It is formed by CNS capillary endothelial cells, their intercellular junctions and a relative lack of vesicular transport. Most substances that must cross the blood–brain barrier are not lipid soluble and therefore cross by specific carrier-mediated transport system.

Some areas of the brain do not have a blood–brain barrier, e.g. posterior pituitary and circumventricular organs. The absence of blood–brain barrier in these regions is consistent with their physiological functions. These leaky regions are isolated from the rest of the brain by specialized ependymal cells called *tanycytes.*

Disruption of blood–brain barrier occurs in a variety of pathological situations, such as brain tumours and bacterial meningitis, etc. This fact can be exploited radiologically by introducing into the circulation a substance that normally cannot penetrate the blood–brain barrier. If the substance can be imaged, its leakage into the region occupied by the brain tumour can be used to demonstrate the distribution of tumour.

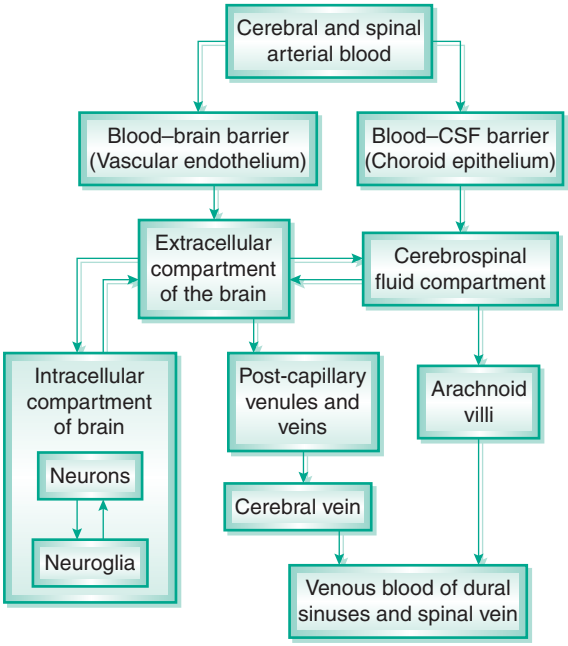


Fig. 10.6-3 Structural and functional relationship between intracranial fluid compartments and blood–brain and blood–CSF barriers. The tissue elements indicated in parentheses form the barrier. Arrows indicate direction of fluid flow under normal conditions. Substances entering the neurons and glial cells (i.e. intracellular compartments) must pass through the cell membrane.

Blood–CSF barrier

The capillaries that traverse the choroidal plexuses are freely permeable to plasma solutes. However, a barrier (blood–CSF barrier) exists at the level of epithelial cells that make up the choroid plexuses. This barrier is responsible for carrier-mediated active transport.

Relationship between intracranial fluid compartments and the blood–brain barrier and blood–CSF barrier is shown in Fig. 10.6-3.

CEREBRAL BLOOD FLOW

Functioning of the brain is closely related to the level of cerebral blood flow. Total cessation of blood flow to the brain causes unconsciousness within 5–10 s because of the decrease in oxygen delivery and the resultant cessation of metabolic activity.

Normal cerebral blood flow in an adult averages 50–65 mL/100 g, or about 750–900 mL/min. Thus, the brain receives approximately 15% of the total resting cardiac output.

Details of cerebral blood flow are given at page 272.

Chapter 10.7

Synaptic Transmission

SYNAPSE
- **Types of synapses**
  - Anatomical types
  - Physiological types

CHEMICAL SYNAPSE
- **Structure**
- **Process of chemical synaptic transmission**
- **Inhibition at synapses**
- **Properties of synaptic transmission**

NEUROTRANSMITTERS
- **Small molecule neurotransmitters**
  - Acetylcholine
  - Biogenic amines
  - Amino acid neurotransmitters
- **Neuropeptide transmitters**
  - Neuroactive peptides
  - Pituitary peptides
  - Peptides acting on the gut and brain

SYNAPSE: DEFINITION AND TYPES

DEFINITION

The synapse is the anatomic site where the nerve cells communicate among themselves. There is no anatomical connection or continuity between different neurons. They are connected only functionally. So, synapse is the functional junction between two neurons.

TYPES OF SYNAPSES

A. Anatomical types

Depending upon the manner an axon terminates on the other neurons, the synapses can be of following types (Fig. 10.7-1):

1. Axo-dendritic synapse. (Fig. 10.7-1A) is the synapse between axon of a neuron with dendrite of another neuron. It is the most common type of synapse. Synapse on dendrites may be located on the spines or on the smooth areas between spines.

2. Axo-somatic synapse. (Fig. 10.7-1B) refers to the synapse between axon of a neuron with the soma (body) of another neuron.

3. Axo-axonic synapse. (Fig. 10.7-1C) is the synapse between axon of a neuron with axon of another neuron. It is a less common type of synapse. An axo-axonal synapse may be located either on the initial segment (of the receiving axon) or just proximal to an axon terminal.

In some parts of the brain (e.g. thalamus), some synapses are seen in which the pre-synaptic element is dendrite instead of an axon. Such synapses may be *dendro-axonic* or *dendro-dendritic*. In yet others, the soma of the neuron may synapse with the soma of a neuron (*somato-somatic synapse*) or with a dendrite (*somato-dendritic synapse*).

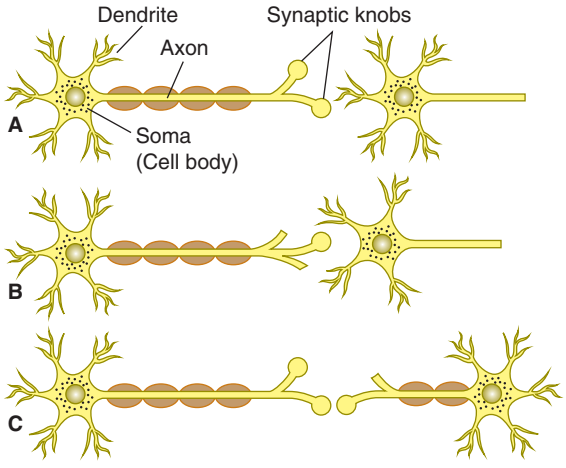


Fig. 10.7-1 Types of synapses depending on the manner, an axon terminates on the other neuron: A, axodendritic synapse; B, axosomatic synapse and C, axo-axonic synapse.

B. Physiological types

Depending upon the process of transmission of impulse, the synapses can be classified as:

1. Chemical synapses are those in which transmission is carried out by *neurotransmitter*. Most synapses in human nervous system are of this type. Chemical synapses conduct information only in one direction. These synapses are more vulnerable to fatigue on repeated stimulation (*synaptic fatigue*) and to the effects of hypoxia and pH changes. Chemical synaptic transmission is definitely slower than the velocity of nerve conduction resulting in the synaptic delay.

2. Electrical synapses are those in which transmission occurs through gap junctions. Transmission at electrical synapses is essentially electrotonic conduction between two neurons. It is similar to the process of nerve conduction. The electrical synapses can conduct in both directions. The speed of transmission at electrical synapses is the same as that of nerve conduction.

Electrical transmission is seen in a few locations (e.g. within the retina and olfactory bulb) in human nervous system. It is found mainly in invertebrates and lower vertebrates.

3. Conjoint synapse refers to a synapse where both the chemical and electrical transmission co-exist.

CHEMICAL SYNAPSE

STRUCTURE OF A CHEMICAL SYNAPSE

As mentioned earlier, the synapse is the functional junction between two neurons. A typical chemical synapse between the axon of one neuron and dendrite of other neuron exhibits following characteristics (Fig. 10.7-2):

Synaptic knob or button. As the axon of neuron approaches the synapse, it loses the myelin sheath and divides into a number of fine branches which end in small swellings called the synaptic knobs or synaptic buttons, which make synapse with the soma or dendrite of the post-synaptic neuron. Each synaptic knob contains large number of mitochondria and synaptic vesicles containing neurotransmitter. Mitochondria provide ATP required for the synthesis of neurotransmitter. The circular synaptic vesicles contain excitatory neurotransmitter and flat or elongated vesicles contain inhibitory neurotransmitters. Besides the neurotransmitter, the vesicles also contain other protein to bind the neurotransmitter to the vesicle. The microtubules present in the synaptic knob transport the vesicles along the axons up to the pre-synaptic grid.

Pre-synaptic membrane refers to the axonal membrane lining the synaptic knobs. On the inner aspect of pre-synaptic membrane are present *zones of dense cytoplasm,* which presumably forms a *pre-synaptic vesicular grid* for organized channeling of the vesicles to the pre-synaptic membrane at site opposite to the receptors on the post-synaptic membrane.

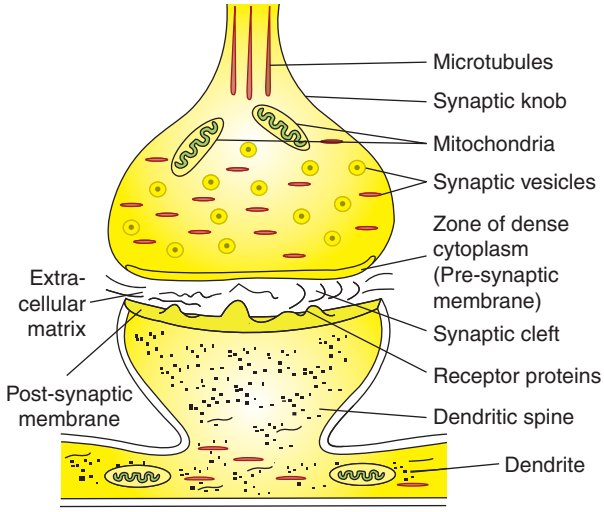


Fig. 10.7-2 Structure of a chemical synapse.

Synaptic cleft is a small gap (20–40 nm wide) between the pre- and post-synaptic membranes. It is filled by the extracellular fluid (ECF) containing some glycoproteins. The extracellular matrix may be acting as an adherent between synaptic neurons.

Post-synaptic process is the name given to the region of receiving neuron (e.g. dendritic spine) where the synaptic knob synapses.

Post-synaptic membrane is the membrane lining the post-synaptic process. On the inner aspect of post-synaptic membrane is present a *zone of dense cytoplasm,* which constitutes the active zone of a synapse. Post-synaptic membrane contains large number of receptor proteins, which protrude outwards in the synaptic cleft. Neurotransmitter released in the synaptic cleft binds with these receptor proteins to cause the effect.

Receptor proteins are of two types:
1. Ion channel receptor proteins. These line the ion channels (Na^+, K^+, Cl^-, etc.) and the neurotransmitter released in the cleft causes opening of the channels by reacting with these receptor proteins.

2. Enzymatic type of receptor proteins. The neurotransmitter released in the cleft reacts with enzymatic type of receptor proteins and causes following effects:

- Activation of cellular gene for manufacture of additional receptor protein channels in the membrane.

- Activation of protein kinase, which decreases the number of receptor protein channels in the membrane.

Thus, there occurs alteration in the reactivity of the neuron to the transmitter. Such effects are called *synaptic modulator effects*.

TYPES OF CHEMICAL SYNAPSES

On the basis of ultrastructure and neurotransmitter present the two types of chemical synapses have been distinguished by Golgi: Type I or asymmetric synapses and type II or symmetric synapses. Their features are summarized in Table 10.7-1.

Table 10.7-1 Features of two types of chemical synapses

Feature	Type I or asymmetric synapses	Type II or symmetric synapses
• Structure	Asymmetric	Symmetric
• Synaptic cleft	Wider (about 30 nm)	Narrower (about 20 nm)
• Thickening of post-synaptic membrane	Marked	Less marked
• Dense extracellular material in the synaptic cleft	Present	Absent
• Shape of vesicles	Small spherical and – Dense cored	Flat or elongated
• Neurotransmitters released	Acetylcholine, glutamate or serotonin is released by spherical vesicles – Dense cored vesicles release noradrenaline, adrenaline or dopamine	GABA, Glycine
• Type of effect	Mostly excitatory	Mostly inhibitory
• Type of synapse	Usually axodendritic	Usually axosomatic

PROCESS OF CHEMICAL SYNAPTIC TRANSMISSION

Most synapses within the central nervous system (CNS) use chemical transmitters. The sequence of events which occur during chemical synaptic transmission are:

- Release of neurotransmitter.
- Development of the excitatory post-synaptic potential (EPSP) or inhibitory post-synaptic potential (IPSP).
- Removal of neurotransmitter from the synaptic cleft.
- Development of action potential.

A. RELEASE OF NEUROTRANSMITTER

- When the nerve impulse (action potential) travelling in a nerve fibre (axon) reaches the nerve terminal (synaptic knobs), there occurs depolarization of the pre-synaptic terminal.
- As a result of depolarization, the voltage-gated Ca^{2+} channels present on the pre-synaptic membrane open up increasing its permeability to Ca^{2+} ions. Consequently, the Ca^{2+} ions present in the ECF of synaptic cleft enter the axon terminal.
- The elevated Ca^{2+} levels in the cytosol of axon results in marked increase in exocytosis of vesicles releasing neurotransmitter into the synaptic cleft. Most commonly, the synaptic vesicles discharge their contents through a small hole in the cell membrane, then opening gets sealed and the main vesicle stays in the cell. This is called *kiss and run* discharge of neurotransmitters. Usually, only one type of neurotransmitter is released from all the terminals of a single neuron. This was first *propounded* by Dale and is called *Dale's phenomenon*.
- After being released from the pre-synaptic terminal, the transmitter diffuses across the synaptic cleft and binds to the post-synaptic receptors. The time lapse (less than 1 ms) occurring between the arrival of nerve impulse at the pre-synaptic terminal and the effect of neurotransmitter on post-synaptic membrane is called *synaptic delay*.

B. DEVELOPMENT OF EXCITATORY POST-SYNAPTIC POTENTIAL AND INHIBITORY POST-SYNAPTIC POTENTIAL

Excitatory post-synaptic potential

Recording of EPSP. The EPSP is a local response. It can be studied by inserting a microelectrode into ventral horn cell of the spinal cord and stimulating the sensory nerve fibres in the dorsal root (Fig. 10.7-3A).

Excitatory post-synaptic potential, i.e. depolarization of the post-synaptic membrane is produced by the excitatory neurotransmitters. The most common excitatory neurotransmitter within the CNS is *glutamate*. The magnitude of the EPSP is 8 mV. The depolarization starts with a latency of 0.5 ms, rises to its peak in 2.0 ms and then declines with a half-life of 4.0 ms (Fig. 10.7-3B).

Ionic basis of EPSP. The excitatory neurotransmitter binds with a specific receptor protein and opens the ligand-gated Na^+ or Ca^{2+} channels on the post-synaptic membrane. As a result, the Na^+ diffuse inward and depolarize the membrane. However, since a very small area of post-synaptic membrane

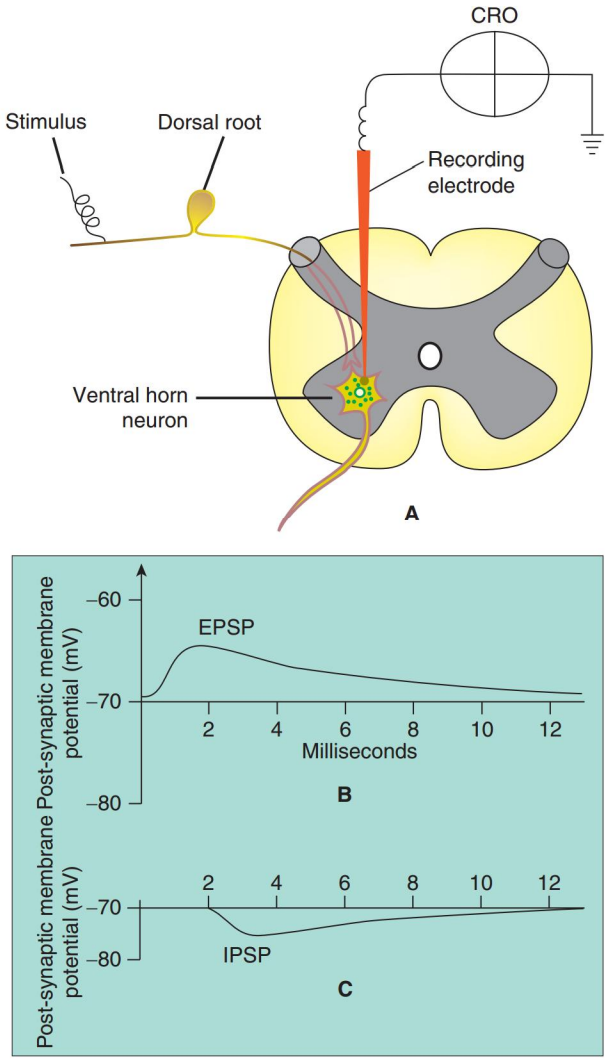


Fig. 10.7-3 Excitatory post-synaptic potential (EPSP) and inhibitory post-synaptic potential (IPSP): A method of recording of EPSP and IPSP; B, the record of EPSP and C, the record of IPSP.

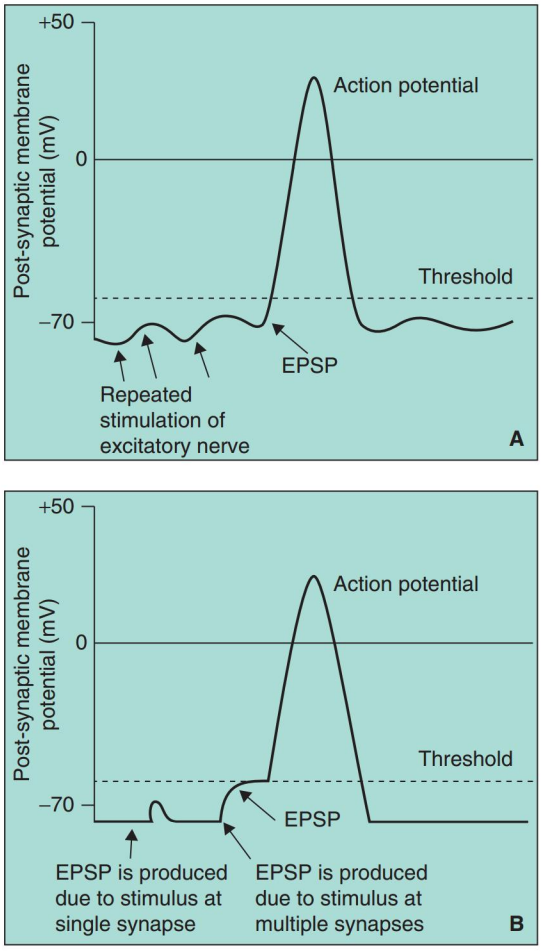


Fig. 10.7-4 Excitatory post-synaptic potential (EPSP) summation resulting in action potential (AP) when threshold is reached: A, temporal summation and B, spatial summation.

develops increased Na^+ permeability, the amount of Na^+ influx is able to produce only a brief depolarization followed by a slower decline to the resting potential.

Conduction of EPSP. The EPSP does not transmit over the cell. However, it can depolarize the adjacent membrane. This occurs passively due to local currents which are set up.

Summation of EPSP. The EPSP is a graded response. It does not follow all or none law like action potential. It shows temporal and spatial summation.

Temporal summation occurs when repeated stimuli are applied at very short intervals (i.e. before the EPSP due to previous stimulus has decayed). The next stimulus adds to the previous post-synaptic potential producing a large response (Fig. 10.7-4A).

Spatial summation occurs when post-synaptic membrane receives impulses from a large number of pre-synaptic terminals simultaneously. The activity in one synaptic knob is said to facilitate in another to reach the firing level. The effect of all the impulses is added up and enough transmitter substance is released to cause a greater response (Fig. 10.7-4B).

Both types of summations occur simultaneously in the neuronal pool. When the temporal or more commonly spatial summation brings the membrane potential of the cell to the firing level, an action potential is fired and propagated in the post-synaptic neuron.

Inhibitory post-synaptic potential

Inhibitory post-synaptic potential, i.e. hyperpolarization of the post-synaptic membrane is produced by the inhibitory neurotransmitters released in the synaptic cleft. The most common inhibitory neurotransmitter within the CNS are *glycine* and γ-aminobutyric acid (GABA).

Ionic basis of IPSP. The inhibitory transmitter released at the synaptic cleft causes opening of either K^+ channels or Cl^- channels in the post-synaptic membrane, leading to diffusion of large number of K^+ ions from the neuron to the ECF or large number of Cl^- ions to diffuse to the interior of

the neuron. This causes post-synaptic membrane potential to become more negative (*hyperpolarization*). This change in potential is called IPSP.

Value of IPSP. The magnitude of IPSP is –2 mV. The hyperpolarization has a latency of 2.0 ms, attaining its maximum at 4 ms and then returning towards the resting membrane potential (RMP) with a half-life of 3 ms (Fig. 10.7-3C).

Recording of IPSP can be made by a technique similar to that of the recording of EPSP (Fig. 10.7-3A).

Summation of IPSP. Spatial and temporal summation also occurs, as seen with EPSP (Fig. 10.7-4). This type of inhibition is called post-synaptic (or direct) inhibition.

IMPORTANT NOTE

Slow post synaptic potentials (both IPSP and EPSP) have been described in autonomic ganglia, cardiac and smooth muscle, and cortical neurons. These potentials have long latency of 100–500 ms and last for longer duration.

C. INACTIVATION OF NEUROTRANSMITTER FROM THE SYNAPTIC CLEFT

The neurotransmitter released in the synaptic cleft from the pre-synaptic terminal is soon inactivated in one of the three ways:

- Diffusion of the transmitter out of the cleft, or
- Enzymatic degradation of the transmitter, e.g. dissociation of acetylcholine by acetylcholinesterase or
- Active transport back into the pre-synaptic terminal (transmitter re-uptake), e.g. active re-uptake of norepinephrine at sympathetic post-ganglionic nerve endings.

The inactivation of the neurotransmitter is essential so that in response to a single electrical impulse, there is release of a transient pulse of the neurotransmitter in the synaptic cleft. Persistence of the transmitter in the synaptic cleft would have produced prolonged stimulation of the post-synaptic neuron in response to a single electrical impulse in the pre-synaptic neuron.

D. DEVELOPMENT OF ACTION POTENTIAL

The development of action potential (AP) from EPSP can be considered in three steps:

- Synaptic integration,
- Generation of initial segment spike and
- Generation of propagated signals, i.e. action potential.

Synaptic integration. Synaptic integration refers to the phenomenon of summation (temporal as well as spatial as described above) of both EPSP and IPSP produced at the post-synaptic membrane. It is the net algebraically summated potential, which determines whether synaptic transmission will occur or not.

IMPORTANT NOTE

The soma of the neuron acts as an integrator that permits grading and adjustment of neural activity for normal function.

Generation of initial segment spike. The summated potential (EPSP and IPSPs) produced by the excitatory and inhibitory neurotransmitters spread passively to the initial segment, which comprises axon hillock and the proximal part of the unmyelinated nerve fibres. If the summated potential is large enough to depolarize the initial segment of neuron to threshold level of about 6–10 mV (the threshold of initial segment is lowest as compared to the other parts of the nerve fibre), a spike potential called the *initial spike* (IS) is generated (Fig. 10.7-5). The magnitude of IS is 30–40 mV from the threshold level.

Generation of propagated signals, i.e. action potential. The IS spike requires a relatively low degree of depolarization for its own production (due to low threshold value of initial segment), but once initiated, itself produces a further depolarization of 30–40 mV by opening the voltage-gated channels on the axon hillock (the sodium channels are plenty in axon hillock than in any other part of the soma). Thus the IS spike, in turn, triggers the generation of the AP spike (Fig. 10.7-5). Once generated the AP travels in both directions, i.e. peripherally in the axon as a nerve impulse and also retrogradely over the cell membrane

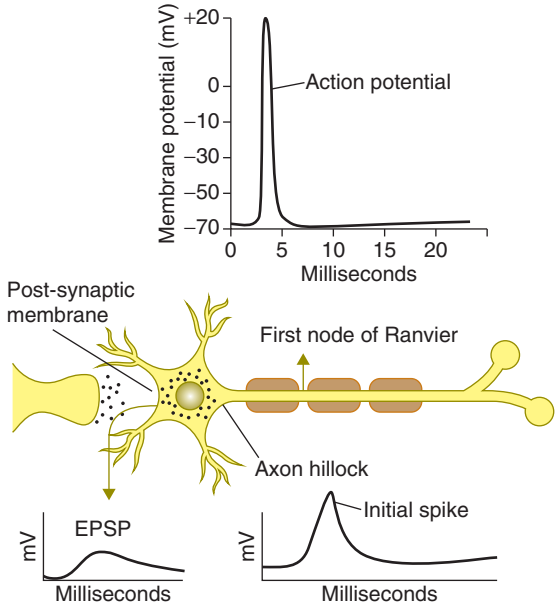


Fig. 10.7-5 Summated post-synaptic potential producing initial spike (IS) and action potential (AP).

of soma and dendrites. This backward conducted AP is called the *SD spike*. The SD spike helps to clear the existing EPSP so that the cell is ready to react to another set of stimuli.

INHIBITION AT SYNAPSES

Four different types of inhibitions known to occur at synapses in the CNS are:

- Post-synaptic inhibition,
- Pre-synaptic inhibition,
- Feedback inhibition and
- Feed forward inhibition.

1. POST-SYNAPTIC INHIBITION

The post-synaptic inhibition, i.e. inhibition of the post-synaptic membrane can occur by following mechanisms:

(i) Direct post-synaptic inhibition by development of inhibitory post-synaptic potential. (as described above, page 780) occurs due to release of inhibitory neurotransmitters.

(ii) Post-synaptic inhibition due to refractory period
Sometimes, the post-synaptic membrane can be refractory to the excitation because it has just fired and is in its refractory period, i.e. existing EPSP has not been still cleared by the SD spike.

2. PRE-SYNAPTIC INHIBITION

This is also known as indirect inhibition as IPSP is not produced. In pre-synaptic inhibition, the excitability of post-synaptic cell is not diminished, whereas in post-synaptic inhibition the IPSP reduces the effectiveness of all excitatory input to a cell. Pre-synaptic inhibition allows a particular excitatory input to be inhibited without affecting the ability of other excitatory synapses to fire the cells. Pre-synaptic inhibition occurs because of the failure of the release of excitatory neurotransmitter substance from the pre-synaptic axon terminal. This occurs in synapses where an inhibitory neuron (neuron C in Fig. 10.7-6) synapses with the afferent fibres of an excitatory neuron (neuron A in Fig. 10.7-6) before the latter synapses with the afferent neuron (neuron B in Fig. 10.7-6). In other words, the pre-synaptic inhibition occurs because of axo-axonic synapse. There are two mechanisms by which pre-synaptic release of neurotransmitter is decreased.

(i) By opening Cl^- channels of pre-synaptic terminal. The inhibitory neuron (neuron C in Fig. 10.7-6) releases an inhibitory neurotransmitter (i.e. GABA), which binds to GABA-gated Cl^- channels on the pre-synaptic neuron terminal (neuron A in Fig. 10.7-6). Increase in Cl^- permeability results in hyperpolarization of the pre-synaptic axon terminal (neuron A). When an AP arrives at the pre-synaptic terminal, the size of AP is reduced because of the increased Cl^- conductance. Because of the smaller size of AP, less Ca^{2+} enters the nerve terminal and thus the amount of excitatory neurotransmitter released is markedly decreased.

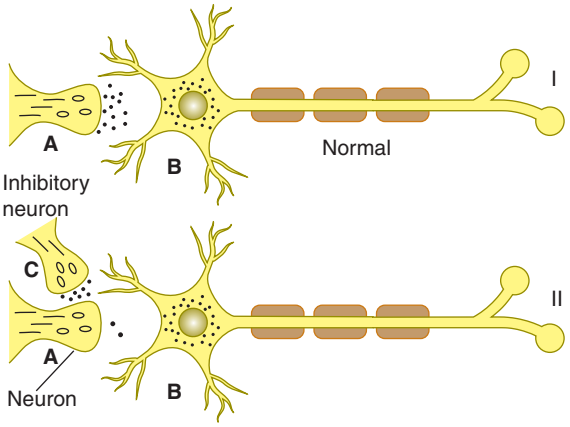


Fig. 10.7-6 Pre-synaptic inhibition produced by an inhibitory neuron (C), which synapses with pre-synaptic axon terminal (A), i.e. by axo-axonic synapse. I: Normal excitatory neurotransmitter released by pre-synaptic terminal (A) and II: reduced excitatory neurotransmitter released by pre-synaptic terminal (A) due to the effect of inhibitory neuron (C).

(ii) By activation of G protein. When the inhibitory transmitter GABA released from the inhibitory neuron (neuron C in Fig. 10.7-6) binds to a receptor called a *GABA receptor*, it activates a *G protein*. The G protein aids in reducing the amount of excitatory neurotransmitter released from the pre-synaptic terminal (neuron A) by acting in one of the two ways:

By opening K^+ channels. The G proteins may open K^+ channels that reduce the size of AP reaching the nerve terminal by hyperpolarizing the pre-synaptic nerve terminal.

By directly blocking the Ca^{2+} channels. The G protein may directly block the opening of Ca^{2+} channels that normally occurs when the AP reaches the nerve terminal; consequently, less Ca^{2+} enters the pre-synaptic terminal and the amount of excitatory neurotransmitter release is diminished.

3. FEEDBACK INHIBITION

The feedback inhibition, also known as *Renshaw cell inhibition,* is known to occur in spinal alpha motor neurons through an inhibitory inter-neuron (the Renshaw cell, Fig. 10.7-7). In feedback (or recurrent) inhibition, a neuron inhibits those very neuron(s) that excite it. In other words, a neuron is inhibited by its own output (that is why it is called negative feedback inhibition). In this way, firing of an action potential by a motor neuron of the spinal cord is

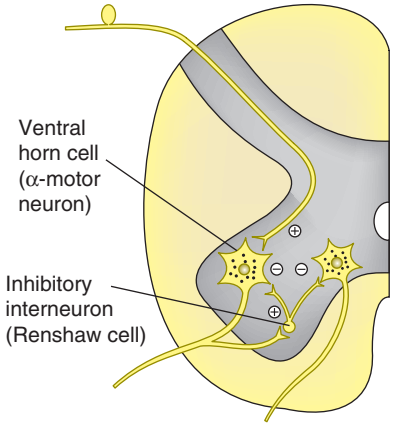


Fig. 10.7-7 Renshaw cell when excited by a recurrent branch of an alpha motor neuron produces feedback inhibition of the soma of the same and other motor neurons.

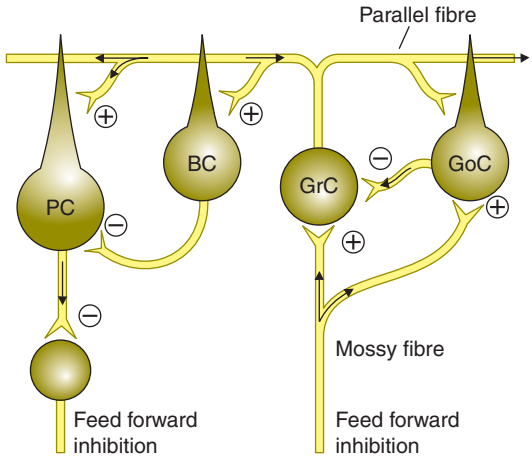


Fig. 10.7-8 Feed forward inhibition of Purkinje cell (PC) by basket cell (BC). Note both Purkinje cell and basket cell are excited by the granule cell (GrC).

followed by a phase of hyperpolarization (inhibition) of not only the same motor neuron, but also many others in the neighbourhood. The feedback inhibition is thus basically a post-synaptic inhibition but is classified separately because the inhibitor Renshaw cells are activated by collateral of the ventral horn cell rather than an afferent neuron. This type of feedback inhibition is also seen in other parts of CNS as well. It serves to limit the excitability of the motor neurons.

4. FEED FORWARD INHIBITION

Feed forward inhibition is seen in cerebellum. In this type of inhibition, a neuron is connected through two pathways: one excitatory and other inhibitory. For example, in cerebellum, the granule cell (GrC) excites Purkinje cells, which is soon inhibited by the basket cell, which in turn was also excited by the granule cell (Fig. 10.7-8). This type of arrangement in the cerebellum limits the duration of excitation produced by any given afferent volley, i.e. allows a brief and precisely timed excitation.

SIGNIFICANCE OF SYNAPTIC INHIBITION

In the CNS, the synaptic inhibition offers a type of restriction over neurons and muscles to react properly and appropriately. Thus, the inhibition helps to select exact number of impulses and to omit or block the excess ones. When the inhibitory system at synaptic level is destroyed, for example by a poison like strychnine, there occurs a continuous and convulsive activity even with a slight stimulation.

IMPORTANT NOTE

In the nervous disorders like parkinsonism, the inhibitory system is impaired resulting in rigidity.

PROPERTIES OF SYNAPTIC TRANSMISSION

Some characteristic features of synaptic transmission are described briefly:

1. One-way conduction. The chemical synapse allows only one-way conduction of an impulse, i.e. from the pre-synaptic to the post-synaptic neuron and never in the opposite direction. This is called *law of dynamic polarity or Bell–Magendie law.*

Cause. One-way conduction occurs because only the pre-synaptic nerve terminals contain the chemical neurotransmitter, whereas the post-synaptic membrane contains the specific receptor sites. Therefore, an impulse conducted antidromically in an axon dies out at the soma due to the absence of the chemical transmitter in the cell body.

Significance. The axons can conduct impulse in either direction with equal ease. However, the synapses act like a valve and are responsible for the orderly conduction of impulse in one direction only.

2. Synaptic delay. Synaptic delay refers to a time lapse, which occurs between arrival of nerve impulse at the pre-synaptic terminal and its passage to the post-synaptic membrane. Normally, synaptic delay occurs by approximately 0.5 ms (almost always less than 1 ms).

Causes of synaptic delay include time taken for:

- Release of neurotransmitter,
- Diffusion of transmitter through synaptic cleft to post-synaptic membrane,
- Action of neurotransmitter to bind with receptors on the post-synaptic membrane and to cause the opening of ion channels and
- Diffusion of ions causing changes in resting membrane potential (i.e. development of EPSP or IPSP).

Significance. When an impulse passes through a chain of neurons, it is delayed at every synapse. The synaptic delay is one of the causes for the latent period of the reflex activity. The number of neurons involved in the reflex can be estimated from the duration of reaction time of a reflex action.

3. Summation property of synapse. A synapse exhibits the property of both temporal and spatial summation of EPSP and IPSP (see page 780).

Significance. Excitation of a single pre-synaptic terminal almost never excites (or inhibits) the post-synaptic neuron as sufficient neurotransmitter is not released to raise EPSP to a threshold level. Therefore property of summation is essential for the stimulation of post-synaptic membrane either by the simultaneous stimulation of large number of pre-synaptic terminals on a post-synaptic neuron (spatial summation) or by repeated stimulation of a pre-synaptic terminal (temporal summation).

4. Convergence and divergence property is present in a chemical synapse.

Convergence refers to a phenomenon of termination of signals from many sources (i.e. many pre-synaptic neurons on a single post-synaptic neuron). Information coming from the large number of pre-synaptic neurons is integrated to decide the onward effect. For example, ventral horn cells of the spinal cord receive convergent signals from the corticospinal tract, reticulospinal tract, rubrospinal tract and sensory afferent from the dorsal root, etc. (Fig. 10.7-9).

Divergence. One pre-synaptic neuron may terminate on many post-synaptic neurons. Thus single impulse is converted to a number of impulses going to a number of post-synaptic neurons, which may travel in the same tract or into multiple tracts (Fig. 10.7-10). This causes magnification and therefore helps in amplification of an impulse. This phenomenon is known as divergence.

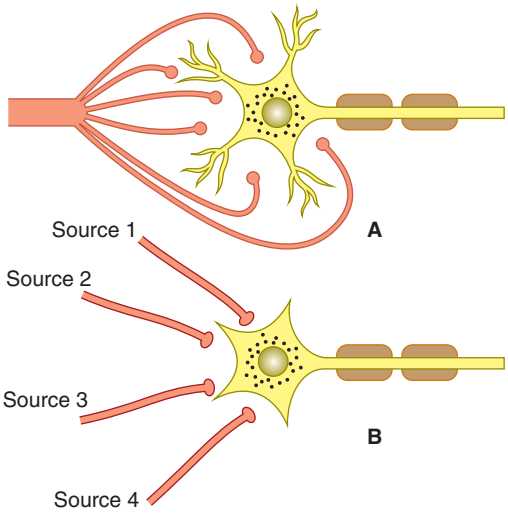


Fig. 10.7-9 Convergence of signals on a single neuron: A, convergence from a single source and B, convergence from multiple sources (e.g. ventral horn cell of spinal cord).

Note. Property of convergence and divergence plays an important role in occlusion and facilitation phenomena.

5. Occlusion phenomenon. The term occlusion describes the situation in which response to stimulation of two pre-synaptic neurons is less than the sum total of the response obtained when they are stimulated separately. For example, when two pre-synaptic neurons (say A and B) are stimulated separately, each stimulates 10 post-synaptic neurons (making a total of 20); but when stimulated simultaneously they stimulate less than 20 post-synaptic neurons (say 15). This happens because of the fact that some post-synaptic neurons are common to both the pre-synaptic neurons (Fig. 10.7-11). Thus occlusion is due to overlapping of afferent fibres in their central distribution.

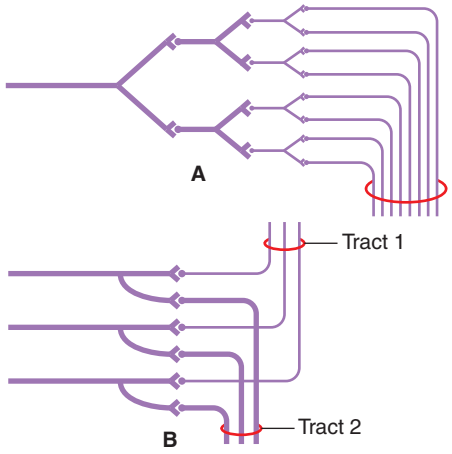


Fig. 10.7-10 Phenomenon of divergence: A, divergence in same tract and B, divergence into multiple tracts.

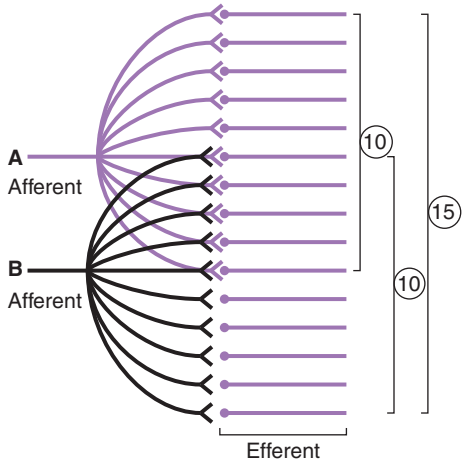


Fig. 10.7-11 Occlusion phenomenon: stimulation of afferent neuron A and B each excites 10 efferent neurons. Simultaneous stimulation of neuron A and B together excite 15 efferent neurons because five efferent neurons are common to both.

6. Subliminal fringe effect. An afferent nerve fibre divides into many hundred branches. Of these, a large number may terminate on one efferent neuron, while a smaller number terminate on other efferent neuron lying nearby. When afferent neuron is stimulated, the efferent (post-synaptic) neuron which has many pre-synaptic terminals is excited to threshold level and AP is fired. Others in the peripheral zone (fringe area) are excited to subthreshold level only, i.e. their excitability is increased but an AP is not fired. This is known as *subliminal fringe effect* (subliminal means below threshold and fringe means border). Thus, the post-synaptic neurons that are fired are said to be in *discharging zone* and those which are not fired are said to be in *subliminal fringe* (i.e. not in the discharging zone).

Because of the subliminal fringe effect, the response obtained by the simultaneous stimulation of two pre-synaptic neurons is greater than the sum total response obtained when they are separately stimulated. This is exactly opposite to occlusion and can be explained as below. Suppose separate stimulation of afferent neurons 'A' and 'B' each causes depolarization of five efferent neurons and subliminal fringe effect in two efferent neurons, then total of 10 efferent neurons are stimulated. But when neurons A and B are stimulated simultaneously, number of post-synaptic neurons stimulated is more (say 12) (Fig. 10.7-12). This is because of the fact that two efferent neurons, which are excited subliminally both by the neuron A and B summate to produce threshold stimulation. This is another example of spatial summation.

IMPORTANT NOTE

Inhibitory impulses also show temporal and spatial summation and subliminal fringe.

Physiological significance. As a result of summation, occlusion and subliminal fringe effect, the patterns of impulses in peripheral nerves are usually altered as they pass through synapses on the way to brain. One such effect is *phenomenon of referred pain* (see page 807).

7. Facilitation. When pre-synaptic axon is stimulated with several consecutive individual stimuli, each stimulus may evoke a larger post-synaptic potential than that evoked by previous stimulus. This phenomenon is known as facilitation.

Mechanism. Each succeeding stimulus increases the duration of action potential in the pre-synaptic neuron, so the voltage-gated Ca^{2+} channels can remain open for a prolonged period liberating more neurotransmitter by exocytosis from the pre-synaptic neuron. In facilitation therefore, normally subliminal stimulus from a pre-synaptic neuron primes the post-synaptic neuron so that another subliminal stimulus can evoke a discharge from the post-synaptic neuron. Hence first stimulus is supposed to facilitate the

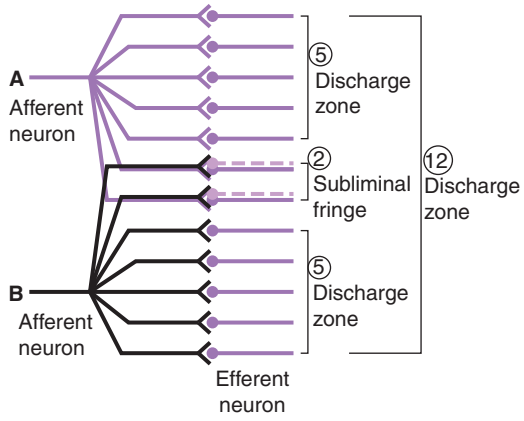


Fig. 10.7-12 Subliminal fringe effect: stimulation of afferent neuron A and B each excites five efferent neurons and subliminal fringe effect on two efferent neurons (which are common to both A and B neurons). Simultaneous stimulation of neuron A and B together excites 12 efferent neurons because the subliminal fringe effect on two neurons gets summated to produce threshold stimulation.

effect due to prolonged exposure of post-synaptic neuron to the neurotransmitter.

8. Synaptic fatigue. When the pre-synaptic neuron is stimulated separately, the rate of impulse discharge in the post-synaptic neuron is initially high but within a few seconds there occurs a gradual decrease and finally disappearance of the post-synaptic response. This phenomenon is called *synaptic fatigue* or *habituation*. Fatigue is a temporary phenomenon. Therefore fatigue and recovery from fatigue constitute an important short-term mechanism for modulating sensitivities of different neuronal circuits.

Mechanism. Fatigue mainly occurs due to *exhaustion of chemical neurotransmitter*, as at high rate of impulse transmission, the synthesis of chemical transmitter fails to keep pace with rate of release at pre-synaptic terminals. Other factors contributing to fatigue are:

- Progressive decreased release of neurotransmitter due to a gradual inactivation of Ca^{2+} channels, which decrease the intracellular Ca^{2+},
- Accumulation of waste products and
- Refractiveness of post-synaptic membrane to transmitter substance.

9. Synaptic plasticity and learning. Plasticity refers to the capability of being easily moulded or changed. Synaptic transmission can be increased or decreased on the basis of past experience. The changes in the synaptic transmission can occur due to alterations at pre-synaptic or post-synaptic location. Plastic changes in synaptic transmission known are:

- Post-tetanic potentiation,
- Long-term potentiation,

- Synaptic fatigue or habituation (see page 785),
- Sensitization and
- Long-term depression.

(i) Post-tetanic potentiation. When a pre-synaptic neuron is stimulated with a single stimulus, followed by stimulation with a volley of stimuli (says 100/s) for 2 s and then again with a single stimulus; the second stimulus evokes a larger post-synaptic response than the first stimulus. The phenomenon is called post-tetanic potentiation. This occurs due to the fact that brief tetanizing stimuli in the pre-synaptic neuron result in an increase in intracellular Ca^{2+} due to increased Ca^{2+} influx (Fig. 10.7-13). It is a form of synaptic facilitation (see page 785).

(ii) Long-term potentiation. When the post-tetanic potentiation gets much more prolonged and lasts for days, it is called long-term potentiation. It occurs due to an increase in the intracellular Ca^{2+} in the post-synaptic neuron rather than the pre-synaptic neuron. This phenomenon commonly occurs in the hippocampus.

(iii) Sensitization. Sensitization refers to a prolonged occurrence of increased post-synaptic responses after a stimulus is paired once or several times with a noxious stimulus. It is basically *pre-synaptic facilitation* of an impulse that occurs due to Ca^{2+}-mediated changes in adenylyl cyclase that results in a greater production of cAMP.

(iv) Long-term depression (LTD) is opposite to long-term potentiation. It is characterized by a decrease in synaptic conduction that occurs due to slow stimulation of pre-synaptic neurons and associated with slow and decrease Ca^{2+} influx. It was first noted in hippocampus and in cerebellum LTD of climbing fibres causing decreased firing of parallel fibres (see page 717).

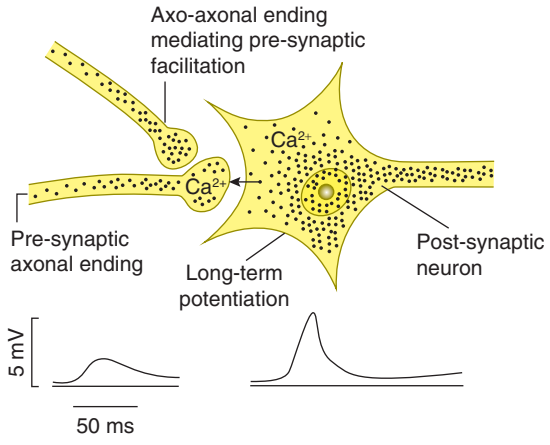


Fig. 10.7-13 Synaptic plasticity: pre-synaptic and post-synaptic sites producing changes in the strength of synaptic transmission.

10. Reverberation. Reverberation refers to the phenomenon of passage of impulse from pre-synaptic neuron and again back to pre-synaptic neuron to cause a continuous stimulation of pre-synaptic neuron. Nervous system is a network of fibres and in this network, it is possible that a branch of axon of a neuron may establish connection with its own dendron. This causes reverberation of impulse through same circuit again and again (Fig. 10.7-14). This is prevented to some extent by phenomenon of fatigue.

11. Reciprocal inhibition. Reciprocal inhibition refers to a phenomenon in which an afferent signal activates an excitatory neuron to a group of muscles and simultaneously activates inhibitory signals to other, usually antagonistic muscles. For example, during flexion of a joint the afferent stimulus causes excitation of the neurons supplying the flexor muscles of the joint and at the same time a branch of afferent fibre excites an inhibitory inter-neuron, which synapses with the motor neuron supplying the extensor muscles of the joint (Fig. 10.7-15), see page 830.

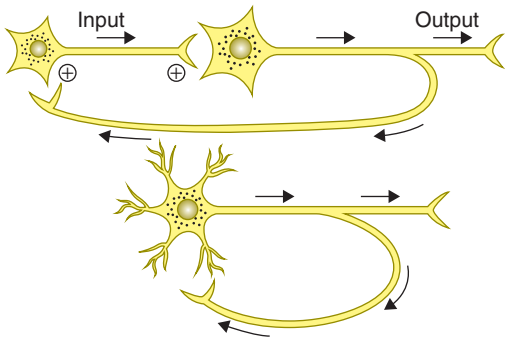


Fig. 10.7-14 Reverberating circuit.

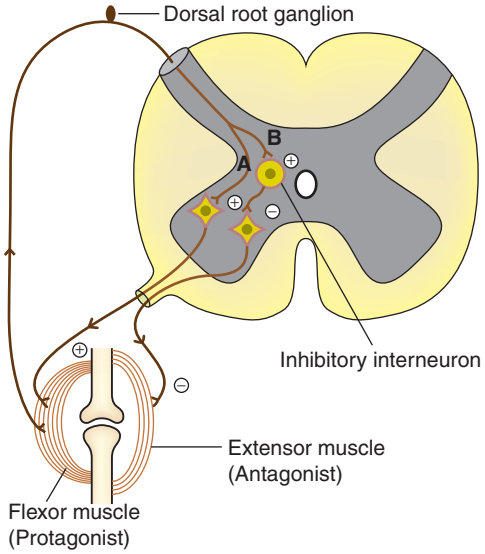


Fig. 10.7-15 Neuronal arrangement of reciprocal inhibition. An afferent stimulus producing contraction of flexors of a joint (through A) and causes inhibition of extensors (through B) by intervention of an inhibitory neuron.

12. After-discharge. After-discharge of a synapse refers to a phenomenon in which a single instantaneous input results into sustained output signals (i.e. a series of repetitive discharges). Input signals only for 1 ms and output signal lasts for many milliseconds.

13. Effect of acidosis and hypoxia. The CNS neurons cannot sustain oxygen lack. Synaptic transmission is particularly vulnerable to the effect of acidosis and hypoxia. This may explain why the first site of fatigue of the synaptic chain is located in the brain.

NEUROTRANSMITTERS

DEFINITION

Neurotransmitters are the chemical substances which are responsible for transmission of an impulse through a synapse.

Criteria for a neurotransmitter. A chemical substance to be qualified as a neurotransmitter should fulfill following criteria:

- A neurotransmitter should be synthesized by pre-synaptic neurons and stored in the vesicles, which are present in axon terminal. The synthesizing enzymes should be present in the nerve at storage site.
- A neurotransmitter should be released on stimulation of nerve.
- A neurotransmitter travels a very small distance between pre-synaptic membrane and post-synaptic membrane.
- A neurotransmitter is associated with an enzyme or enzyme system for its inactivation.
- A neurotransmitter when applied extrinsically should mimic the effects of the nerve stimulation.
- Drug which modifies the response to nerve stimulation should also modify the proposed transmitter action in a similar way.

Extended definition of neurotransmitter. It also includes, in addition to the principal neurotransmitters, following chemical substances:

Neuromediators or neurohormones. These chemical substances are synthesized in neurons and poured into the blood stream through terminals resembling synapses in structure. Similar chemical substances are also poured into the cerebrospinal fluid or into the intercellular spaces to influence other neurons in a diffuse manner.

Neuromodulators are the chemical substances, which are associated with synapses but do not influence synaptic transmission directly, but influence the effects of transmitters or of neuromediators. Several peptides found in the nervous system probably act as neuromodulators. These include substance P, vasoactive intestinal polypeptide (VIP), somatostatin, cholecystokinin and many others.

CLASSIFICATION

At present more than 50 substances have been reported to fulfil the criteria as neurotransmitter. Generally, these substances can be classified in two ways:

Biochemical classification

Biochemically, neurotransmitter can be divided into two groups:

A. Small molecule neurotransmitters. These act rapidly and cause acute response. These are synthesized and packed into synaptic vesicles in the axon terminal. Important small molecule neurotransmitters are:

I. Acetylcholine (Ach)
II. Biogenic amines. These include:
- Catecholamines: epinephrine (EP), norepinephrine (NE), dopamine (DA).
- Serotonin (5 hydroxytryptamine, 5HT) and
- Histamine.

III. Amino acid neurotransmitters. These include:
- Gamma-aminobutyric acid (GABA),
- Glycine,
- Glutamic acid or glutamate, and
- Aspartic acid or aspartate.

B. Neuropeptide transmitters. These are slowly acting and have prolonged effect. These neurotransmitters include:

- Neuroactive peptides.
- Pituitary peptides.
- Peptides acting on the gut and brain.
- Neuropeptides from other tissues.

Physiological classification

Some of the neurotransmitters cause excitation of post-synaptic neurons while others cause inhibition. Thus, physiologically neurotransmitters can be divided into two groups:

- Excitatory neurotransmitters and
- Inhibitory neurotransmitters.

PRINCIPAL NEUROTRANSMITTERS

A. SMALL MOLECULE NEUROTRANSMITTERS

I. ACETYLCHOLINE

Acetylcholine (Ach) is a principal neurotransmitter released by cholinergic neurons in the nervous system.

Cholinergic neurons

Cholinergic neurons, i.e. neurons which secrete Ach at their nerve endings include:

- Nerve endings at the neuromuscular junction,
- Pre-ganglionic parasympathetic nerves,

- All pre-ganglionic sympathetic nerves,
- All post-ganglionic parasympathetic nerves,
- Post-ganglionic sympathetic cholinergic nerves those which innervate:
  - Sweat glands and
  - Skeletal muscle blood vessels (sympathetic vasodilator nerves),
- Endings of some amacrine cells of the retina and
- Many parts of the brain (especially cerebral cortex, thalamus and forebrain nuclei). Ach is specifically released by large pyramidal cells and many neurons of basal ganglia.

Cholinergic receptors

Ach receptors are of two types: nicotinic and muscarinic (for details see page 767).

Ach synthesis, storage, release and removal

- *Ach synthesis and storage.* Ach is synthesized within the mitochondria in the pre-synaptic terminal from acetyl coenzyme A (CoA) and choline by a reaction catalyzed by the enzyme choline acetyltransferase. After formation, Ach is stored temporarily in synaptic vesicles with ATP and proteoglycan for later release.
- Release of Ach by the nerve terminal (see page 779).
- Removal of Ach from synaptic cleft (see page 781).

Actions of Ach

Acetylcholine is very quick in action and in most instances it is excitatory. It produces the excitatory function of synapse by opening the ligand-gated Na^+ channels. At very few places (vagus supplying heart) Ach acts as an inhibitory transmitter.

Muscarinic versus nicotinic actions of Ach on the post-synaptic receptors are summarized in Table 10.7-2.

Role of Ach in function of brain Both muscarinic and nicotinic receptors are found in the CNS, however, most of these are muscarinic. The cell bodies of cholinergic neurons in the brain are concentrated in relatively few areas, but their axons are widely distributed.

- Cholinergic neurotransmission has been most thoroughly studied in cerebral cortex where it acts as an excitatory neurotransmitter. The release of Ach in the cortex is proportional to the level of cortical excitability, being increased by a variety of convulsants and decreased by anaesthesia.
- Its role in memory and malfunction in Alzheimer's disease has aroused a considerable interest recently in anticholinesterase drugs like tacrine and donepezil.
- Cholinergic projections are also involved in motivation, perception and cognition (see page 854).
- Cholinergic neurons in the pons and lateral tegmental nuclei project through the pons-mid brain reticular formation to the thalamus and are involved in the attention and arousal function of RAS (reticular activating system, see page 865).
- In the basal ganglia, Ach is a principal excitatory neurotransmitter (see page 726).
- The ponto-geniculo-occipital spike system responsible for REM sleep (rapid eye movement sleep) is cholinergic (see page 867).

Table 10.7-2 Muscarinic versus nicotinic actions of Ach

Features	Muscarinic action	Nicotinic action
1. Site of action	• Post-synaptic membranes in cardiac muscles, smooth muscles and glandular cells	• All autonomic ganglia • Neuromuscular junctions in skeletal muscles
2. Characteristics of action	• Actions resemble those of mushroom poison muscarine • Actions are slow in onset • Actions are prolonged	• Action resembles the drug nicotine • Actions are quick in onset • Actions are of brief duration
3. Actions are antagonised by	• Atropine which combines with Ach receptors at the sites of muscarinic action	• Hexamethonium at autonomic ganglia and • Tubocurarine at skeletal muscles

II. BIOGENIC AMINES

1. Catecholamines

- Catecholamines include epinephrine (adrenaline), norepinephrine (noradrenaline) and dopamine.
- Epinephrine is not a common neurotransmitter in the brain or peripheral nervous system but is the major hormone secreted by the adrenal medulla.
- The biogenic amines are the basic *by R-NH_2*.
- Catecholamines contain a catechol ring (a six-sided carbon ring with adjacent hydroxyl group) and an amine ring.

Biosynthesis, metabolism and excretion

The three catecholamines, EP, NE and DA, are synthesized from the amino acid phenylalanine (Fig. 10.7-16). For details see page 595.

Norepinephrine and epinephrine

Adrenergic neurons refer to those neurons which either secrete epinephrine (adrenaline) or norepinephrine (noradrenaline) at their nerve endings.

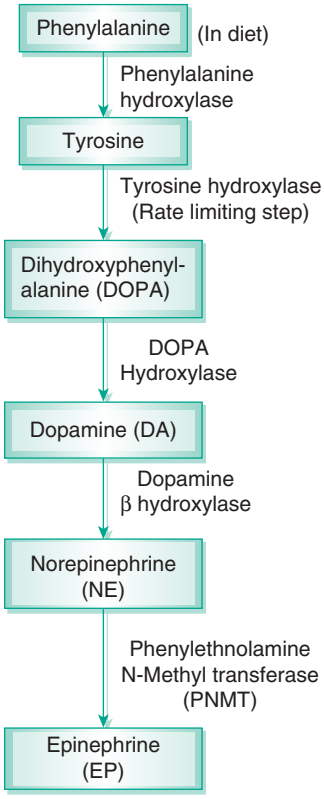


Fig. 10.7-16 Synthesis of catecholamines from the amino acid phenylalanine.

Epinephrine is produced almost exclusively in the adrenal medulla, with a small amount being synthesized in brain.

Norepinephrine is released by the adrenal medulla and following noradrenergic nerve endings in peripheral and central nervous system:

- *Post-ganglionic sympathetic neurons*. It is the primary neurotransmitter released from the post-ganglionic sympathetic neurons except those supplying the sweat glands and blood vessels of skeletal muscles.
- Neurons of cerebral cortex and hypothalamus
- Noradrenergic neurons of pons and medulla oblongata constitute two major systems: *locus coeruleus system* and *lateral tegmental system*. From these neurons, the axons descend to *spinal cord* and *cerebellum*, and ascend to many other parts of the brain.

Distribution of noradrenergic neurons in the CNS is shown in Fig. 10.7-17.

Actions. For detailed action of epinephrine and norepinephrine on different systems of the body, see page 597.

- In general, norepinephrine is mainly excitatory neurotransmitter, only at few places it is inhibitory.
- Epinephrine and norepinephrine produce different effects due to the existence of two types of adrenergic receptors: α and β (each further subdivided into α_1-, α_2-, β_1-, β_2- and β_3-, respectively).

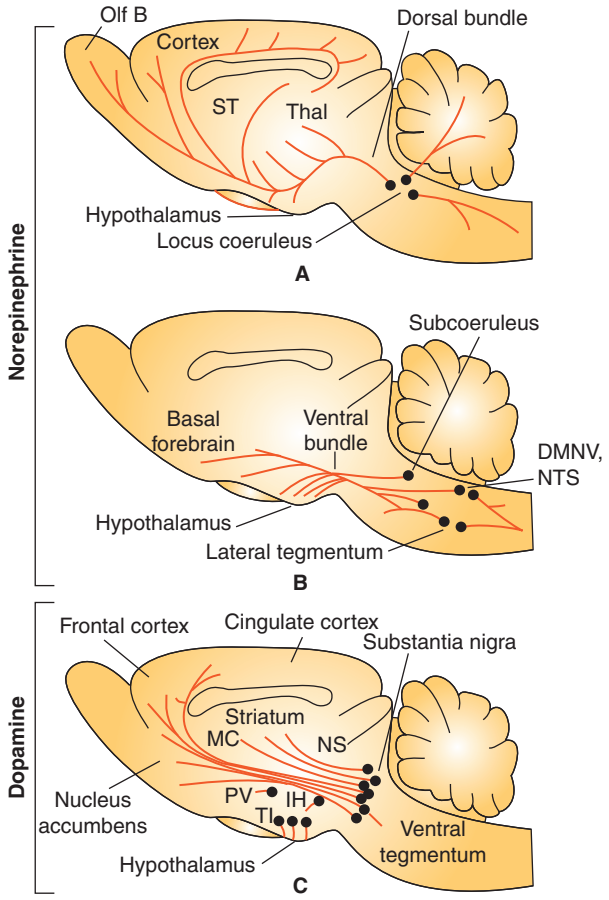


Fig. 10.7-17 Aminergic pathways in the central nervous system. Two principal noradrenergic systems. Locus coeruleus (A) and lateral tegmental (B) and dopaminergic pathway (C). (Olf B = Olfactory bulb; ST = stria terminalis; Thal = thalamus; DMNV = dorsal motor nucleus of vagus; NTS = nucleus tractus solitarius (nucleus of solitary tract); NS = nigrostriatal system; PV = periventricular system; MC = mesocortical system.)

- Epinephrine acts equally on both α and β-receptors, while norepinephrine acts on α receptors (see page 598).
- Their receptor action is linked for the second messenger cAMP and cGMP, etc.

Regulation of NE release. There is a pre-synaptic regulation of NE release mediated by pre-synaptic α-receptors and a positive feedback mechanism mediated by pre-synaptic β-receptors. Such combined effects control the need oriented release of neurotransmitter.

Removal and metabolism of NE. Norepinephrine is removed from the synapse by *reuptake* or is *metabolized* in the pre-synaptic terminal by monoamine oxidase (MAO) and catechol-O-methyl transferase (COMT). The metabolites are: dihydroxymandelic acid, normetanephrine and 3-methoxy-4-hydroxy-phenyl-glycol. For details see page 597, Fig. 8.5-9.

Functional role of norepinephrine in CNS

- Due to its excitatory effects, NE is believed to be involved in dreams, arousal and elevation of mood. Therefore the drugs that increase extracellular NE in the brain elevate mood and drugs that decrease it cause depression.
- Noradrenergic neurons of hypothalamus are involved in the regulation of the secretion of ADH, oxytocin and hypothalamichypophyseotropic hormones (that in turn regulate the secretion of anterior pituitary).
- Noradrenergic neurons suppress ACTH secretion by inhibiting the activity of the neurons, which synthesize and secrete corticotropin-releasing factor.
- Norepinephrine is an inhibitory transmitter in the thalamus, cerebral cortex and cerebellar cortex.

Dopamine Dopamine is naturally acting precursor of NE.

Dopamine receptors. Dopamine acts on three types (D_1–D_3) of dopamine receptors.

- *D_1 receptors* activate adenylyl cyclase via Gs protein and
- *D_2 receptors* inhibit adenylyl cyclase via Gi protein. The brain contains more of *D_2* receptors.
- *D_3 receptors* are localized to nucleus accumbens.

Dopaminergic neurons have their cell bodies in the mid brain. They project from the substantia nigra to the (Fig. 10.7-17):

- Striatum (nigrostriatal tract),
- Olfactory tubercle,
- Nucleus accumbens and
- Limbic system area.

Note. Highest concentration of dopamine is present in the basal ganglia, limbic system and chemoreceptor trigger zone (CTZ) in the medulla. It does not cross blood-brain barrier.

Metabolism of dopamine. Dopamine is metabolized by MAO and COMT (details see on page 789).

Functional roles of dopamine in CNS

1. *Control of movements.* Dopamine serves as a central neurotransmitter for control of movement of the corpus striatum, which modulates muscle tone and voluntary movements by influencing extrapyramidal motor system. Deficiency of dopaminergic neurons (nigrostriatal tract) produces parkinsonism. For details see page 730.

2. *Induction of vomiting.* Dopamine also mediates the activity of CTZ and is responsible for the induction of all vomitings other than that of the vestibular origin.

3. *Inhibition of prolactin secretion and stimulation of GnRH.* Dopamine is released in the hypothalamus and causes:

- Inhibition of prolactin secretion and
- Stimulation of GnRH release.

4. *Retina* also contains some inhibitory dopaminergic neurons.

5. *Dopamine and schizophrenia.* Schizophrenia type of psychosis involves increased levels of D_2 receptors. For details see page 856.

- Amphetamine which stimulates secretion of NE and DA produces schizophrenia, when administered in high doses.
- Several drugs used as tranquillizers reduce the content of dopamine in the brain neurons and are effective in the relief of the schizophrenia.

Note. It is possible that there exist different co-transmitters for dopamine in different areas which affects its activity differently.

2. Serotonin

Synthesis and metabolism. Serotonin (5 hydroxytryptamine, 5HT) is synthesized from tryptophan (an essential amino acid). It is inactivated by MAO to form 5-hydroxy indole acetic acid (5-HIAA), which is excreted in urine.

Note. In the pineal gland, 5HT is converted into melatonin.

Sites of secretion. 5HT is present in brain and non-neural cells.

In the brain, serotonergic neurons have their cell bodies in the brain stem and they project to portions of hypothalamus, limbic system, neocortex and spinal cord.

Non-neural cells that contain serotonin are blood platelets (highest concentration, mast cells and gastrointestinal tract enterochromaffin cells and myenteric plexus).

Serotonin receptors. There are seven groups of serotonin receptors ($5HT_1$ to $5HT_7$), each group has further subgroups from A to F (that is $5HT_{1A}$, $5HT_{1B}$ and so on). 5HT receptors mostly coupled with G-proteins and affect adenylyl cyclase and phospholipase C.

Effects of 5HT. General considerations are:

- Some 5HT receptors are excitatory, while others are inhibitory. In general, 5HT has an excitatory effect on motor pathways and an inhibitory effect on the sensory pathways.
- Effects of 5HT generally have a slow onset, indicating that it works as a neuromodulator (which often modify the post-synaptic cell's response to specific neurotransmitters).
- The activity of serotonergic neurons is lowest or absent in sleep and highest during states of alert wakefulness, increased 5HT activity increases motor responsiveness and suppresses sensory systems to screen out distracting stimuli.

Functional roles of 5HT in CNS

1. Regulation of carbohydrate intake and hypothalamic-releasing hormones. Serotonergic pathways function in the regulation of carbohydrate intake and hypothalamic-releasing hormones, and they have been implicated in alcoholism and other obsessive compulsive disorders. Norepinephrine and 5HT both are involved in food intake and control of body temperature.

2. Pain inhibition. Serotonin inhibits impulses of pain sensation in posterior grey horn of spinal cord. The presence of descending serotonergic neurons in the brain stem and spinal cord is essential for the analgesic action of morphine. For details see page 810.

3. Hallucinations and 5HT. Lysergic acid diethylamide (LSD), the most potent hallucinogenic drug known, activates the serotonergic neurons.

IMPORTANT NOTE

Several chemical substances related to 5HT, such as *psilocybin*, a hallucinogenic agent, found in some mushrooms have potent psychic effects.

4. Depression of mood (see page 855).

3. Histamine

Histamine is formed by the decarboxylation of the amino acid histidine (Fig. 10.7-18).

Sites of secretion. Histamine is secreted in brain and non-neural cells.

- *In the brain,* histaminergic neurons have their cell bodies mainly in the posterior hypothalamus and their axons project to all parts of the brain including the cerebral cortex and spinal cord.
- *Non-neural cells* that contain histamine are of gastric mucosa and heparin containing mast cells.

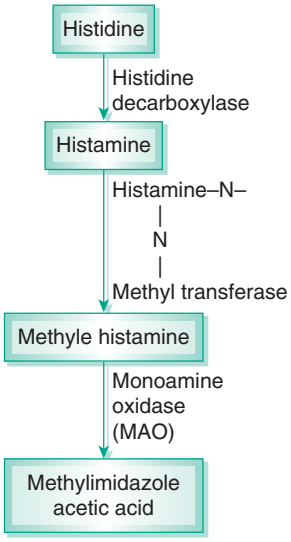


Fig. 10.7-18 Synthesis and catabolism of histamine.

Histamine receptors are of three types: H_1, H_2 and H_3. All the three types of receptors are found in the brain and peripheral tissues:

- *H_1 receptors* activate phospholipase C,
- *H_2 receptors* increase intracellular cAMP and
- *H_3 receptors* are pre-synaptic and they mediate inhibition of the release of histamine via G-protein.

Functional role of histamine in CNS. Histamine is an excitatory neurotransmitter. The exact function of diffuse histaminergic system is not known as yet. It is believed that histamine plays an important role in arousal and sexual behaviour, regulation of secretion of some anterior pituitary hormones, drinking, pain threshold and sensation of itch.

III. AMINO ACID NEUROTRANSMITTERS

Excitatory amino acid neurotransmitters, which cause neuronal depolarization include glutamic acid (glutamate) and aspartic acid (aspartate).

Inhibitory amino acid neurotransmitters, which cause neuronal hyperpolarization include GABA and glycine.

1. Glutamic acid

Glutamic acid (glutamate) is the most prevalent *excitatory neurotransmitter* in the brain and dorsal sensory nerve terminals.

Synthesis, storage and release, and removal. In the CNS neurons, glutamate is mainly derived from either glucose, via the Krebs' cycle, or from the glutamine, which is synthesized by the glial cells and taken up by the neurons. It is stored in the synaptic vesicles and released by calcium-dependent exocytosis. The action of glutamate is mainly terminated by carrier-mediated *reuptake* into the nerve terminals and neighbouring glial cells.

Glutamate receptors present on the post-synaptic neurons are:

- *Ionotropic glutamate receptors* are ligand-gated ion channel, which when stimulated increase the conductance of Na^+ and Ca^{2+} into the cell leading to depolarization. Their further subtypes are:
  - NMDA receptors (N-methyl-D-aspartate)
  - AMPA receptors (α amino-3-hydroxy-5-methylisoxazole-4 propionate)
  - Kainate receptors.

NMDA receptor concentration is high in hippocampus. Blockade of these receptors prevents long-term potentiation.

- *Metabotropic glutamate receptors* are serpentine G-protein coupled receptors. They stimulate the phosphoinositol turnover. They play a role in synaptic plasticity mainly in cerebellum and hippocampus.

Note. Destruction of these receptors leads to severe motor inco-ordination and defects in learning.

2. Aspartic acid

Aspartic acid or aspartate seems to be the chief excitatory transmitter of cortical pyramidal cells.

3. Gamma-aminobutyric acid

Gamma-aminobutyric acid is the major inhibitory neurotransmitter in the whole CNS, i.e. spinal cord, brain stem, cerebral cortex and cerebellum.

Synthesis. It is formed by decarboxylation of glutamic acid by the enzyme glutamate decarboxylase (GAD) pyridoxal phosphate, a vitamin B complex derivative is co-factor for GAD (Fig. 10.7-19).

APPLIED ASPECTS

- *Stiffman syndrome.* An autoimmune disease characterized by progressive muscle rigidity (fluctuating type) due to GABA deficiency, which occurs because of autoimmunity against GAD.
- Type I diabetes mellitus occurs due to ß cell destruction of pancreas because of antibody against GAD.

Removal of GABA from the synaptic cleft occurs chiefly by its reuptake.

GABA receptors. GABA produces pre-synaptic inhibition, i.e. indirect inhibition (see page 782) by two types of receptors:

- *$GABA_A$ receptors* produce inhibition by increasing Cl^- conductance and
- *$GABA_B$ receptors* produce inhibition by increasing K^+ conductance and by decreasing cAMP activity that decreases Ca^{2+} influx (see page 782).
- *$GABA_C$ receptors* are exclusively present in retina.

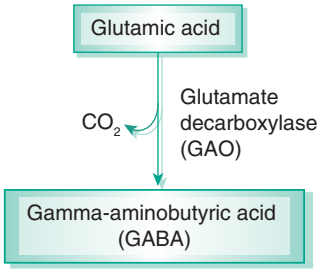


Fig. 10.7-19 Formation of gamma-aminobutyric acid (GABA).

Note. Substances, such as benzodiazepine (diazepam) and barbiturates are used as antianxiety, anticonvulsants, muscle relaxants and sedatives act by binding to $GABA_A$ receptors of brain neurons and facilitates Cl^- conductance.

4. Glycine

- Glycine is an inhibitory neurotransmitter found primarily in the grey matter of the spinal cord and brain stem.
- It produces direct inhibition (post-synaptic inhibition) in the spinal cord and acts by increasing Cl^- conductance by acting on glycine receptor, which functionally resembles GABA receptors.
- Agents, such as strychnine and tetanus toxin, that antagonise the post-synaptic inhibitory action of glycine produce convulsions and muscular hyperactivity.

B. NEUROPEPTIDE TRANSMITTERS

Neuropeptide transmitters are slow acting and have a prolonged effect in contrast to small molecule transmitters, which act rapidly and cause short-lasting acute response. These cannot be synthesized in the cytosol of the axon terminal but are typically synthesized in the soma as integral components of large proteins. These large molecules are cleaved in the cell body and packaged into vesicles in the Golgi apparatus either as an active peptidergic agent or as a precursor of the neuroactive substance. The vesicles are delivered to the axon terminals and the transmitter is released into the synaptic cleft.

Mechanism of action

The peptides can alter ion channel function and modify cell metabolism or gene expression and these actions can be sustained for minutes, hours, days or presumably even longer.

Types of neuropeptides

Many of the peptides function in communication network within the neural, endocrine and immune systems. These are:

1. Neuroactive peptides. These include releasing hormones from hypothalamus, such as TRH, LH releasing hormone and somatostatin.

2. Pituitary peptides. These include ACTH, β-endorphin, vasopressin and oxytocin.

- *Vasopressin (ADH) and oxytocin.* Besides acting as hormones (see page 546), they are also present in the neurons that project to the brain stem and spinal cord. They appear to be involved in control of CVS.

3. Peptides acting on the gut and brain. These include leucine, enkephalin, methionine, substance P, cholecystokinin, VIP, neurotensin, insulin, glucose and opioid polypeptides.

Enkephalins act on three types of receptors which occur in discrete locations in CNS:

- In areas containing pathway that convey pain information,
- In parts of the brain involved in mood and
- In parts of the brain involved in emotions.

Substance P. Substance P is a polypeptide containing 11 amino acids. It is the transmitter released by:

- Primary pain nerve endings in spinal cord,
- Hypothalamus and
- Nigrostriatal system of basal ganglia.

Cholecystokinin and VIP. These are also found in brain, the former in the hypothalamus and the latter in the cerebral cortex. *Opioid polypeptides* have an important role in the inhibition of pain signals in the brain and spinal cord.

4. Neuropeptides from other tissues. These include angiotensin II, bradykinin, bombesin and neuropeptide-α. *Neuropeptide*-Y is closely related to the pancreatic polypeptide and is present in many parts of the brain and autonomic nervous system. It increases the vasoconstrictive effect of norepinephrine. Its level in the circulation from the sympathetic nerves increases during severe exercise.

Chapter 10.8

Somatosensory System

GENERAL SENSORY MECHANISM

INTRODUCTION

SENSATIONS

Sensory division of the human nervous system is concerned with collection of the information about outside world and changes occurring within the body itself. Sensation refers to a conscious perception of sensory information reaching the brain. Sensations may be broadly classified into two groups:

1. Special senses. These include visual sensations, auditory sensations, gustatory (taste sensation) and olfactory (smell) sensations. These have been discussed in detail in different chapters of Section 11.

2. Somaesthetic senses. These, depending upon their point of origin, can be classified into three types:

A. Exteroceptive sensations, also known as cutaneous sensations, arise from the surface of the body.
These include:

- *Tactile sensation,*
- *Pressure sensation,*
- *Pain sensation* and
- *Temperature sensation.*

B. Visceral sensations arise from the viscera, i.e. internal organs and are called visceral sensations.

C. Proprioceptive and kinaesthetic sensations arise from the muscles, tendons and joints. These include:

(i) Proprioceptive sensations: These are concerned with the physical state of the body, i.e. the sense of position, tendon and muscle sensations, deep pressure and sense of equilibrium.

(ii) Kinaesthetic sensations or kinaesthesia: It is the conscious recognition of rate of movement of different parts of the body. Kinaesthetic sensations include both:

- Conscious kinaesthetic sensations and
- Unconscious kinaesthetic sensations.

COMPONENTS OF SENSORY SYSTEM

The sensory division of the human nervous system includes following components:

1. Sensory receptors. These are specialized cells that transduce stimulus energy into neural signals.

2. Afferent neurons. These carry sensory impulses to the sensory cortex and constitute the neural pathway. Sensory neural pathway consists of:

- First-order neurons,
- Second-order neurons and
- Third-order neurons.

3. Sensory cortex. It includes the sensory areas of cerebral cortex. It is formed by fourth-order sensory neurons. The information received in the sensory cortex results in a conscious perception of the stimulus, i.e. a sensation.

RECEPTORS

Sensory receptors are specialized cells that receive stimuli from the external or internal environment and transduce these signals into neural signals. A *stimulus* is a change of environment of sufficient intensity to evoke a response in an organism. The external stimuli may be mechanical, chemical, thermal, auditory or visual.

CLASSIFICATION OF RECEPTORS

Receptors can be variously classified:

A. Depending on the source of stimulus (Sherrington's classification)

1. Exteroceptors, i.e. the receptors which receive stimuli from immediate surrounding outside the body, e.g.

- *Cutaneous receptors* for pain, touch and temperature.

2. Enteroceptors, i.e. the receptors which receive stimuli from within the body, e.g. *chemoreceptors, baroreceptors, proprioceptors, osmoreceptors and glucoreceptors*

3. Telereceptors, i.e. the receptors that receive stimuli from the distance, e.g. visual receptors, cochlear receptors and olfactory receptors.

B. Depending on type of stimulus energy

1. Mechanoreceptors, i.e. those receptors which respond to mechanical stimuli. These include:

(i) *Cutaneous receptors (in epidermis and dermis) for cutaneous tactile sensibility.*
(ii) *Cutaneous receptors for deep tissue sensibility.*
(iii) *Muscle and joint receptors*
(iv) *Hair cells, e.g.* Hair cells in organ of Corti (cochlear) or auditory receptors, and Hair cells in vestibular apparatus or vestibuloreceptors for equilibrium.
(v) *Baroreceptors* of carotid sinus and aortic arch for detecting level of arterial blood pressure.

2. Thermoreceptors, which detect environmental temperature, e.g. cold receptors and warm receptors.

3. Photoreceptors or electromagnetic receptors, i.e. rods and cones of the retina, which respond to light stimuli.

4. Chemoreceptors, which detect change in the chemical composition of the environment in which they are located, e.g.

- Taste receptors,
- Olfactory receptors,
- Osmoreceptors in supraoptic nuclei of hypothalamus,
- Aortic and carotid bodies receptors, which detect level of arterial pO_2, pCO_2 and pH,
- Glucoreceptors,
- Chemoreceptors on the surface of medulla for detecting level of blood pCO_2 and
- Chemoreceptors in hypothalamus detecting levels of blood glucose, fatty acids and amino acids.

5. Nociceptors, i.e. the receptors which respond to extremes of mechanical, thermal and chemical stimuli producing pain.

C. Clinical or anatomical classification of receptors

1. Superficial receptors, i.e. those present in skin and mucous membrane.

2. Deep receptors, i.e. those present in muscles, tendons, joints and subcutaneous tissue.

3. Visceral receptors, which are present in the visceral organs.

SENSORY TRANSDUCTION

Sensory transduction refers to the phenomenon of transduction of environmental signals into neural signals by the receptors. Steps of sensory transduction are:

- Arrival of stimulus to receptor,
- Production of generator or receptor potential and
- Production of action potential in the sensory nerve.

Arrival of stimulus to receptor

The stimulus arriving at the given sensory receptor may be in the form of:

- *Mechanical force* causing depression of the skin, which stimulates mechanoreceptors,
- *Light or electromagnetic wave,* which stimulates photoreceptors of the retina,
- *Chemical*, e.g. a molecule of NaCl on the tongue which stimulates chemoreceptors,
- *Cold or warm temperature* stimulating thermoreceptors and
- *Sound energy* stimulating auditory receptors, and so on and so forth.

Production of receptor potential

When a stimulus excites the receptor, it changes the potential across the membrane of the receptors. This change in the potential is called receptor or generator potential.

Mechanism of development of receptor potential (Fig. 10.8-1)

The change in membrane potential in a receptor is caused by a change in the permeability of membrane of the unmyelinated terminals to Na^+. The resultant influx of Na^+ causes development of generator or receptor potential.

Usually, the current is inward which produces *depolarization of the receptor.* The exception is in the photoreceptors, where light causes *hyperpolarization.*

Properties of receptor potential

The receptor potential is not action potential. It is similar to excitatory post-synaptic potential in synapse, end plate potential in a neuromuscular junction and electrotonic potential in a nerve fibre. The important properties of receptor potential are:

1. Graded response. Receptor potential is a graded response, i.e. its amplitude increases with increasing velocity of stimulus application, and increasing strength of stimulus. Thus, unlike action potential it does not obey all or none law.

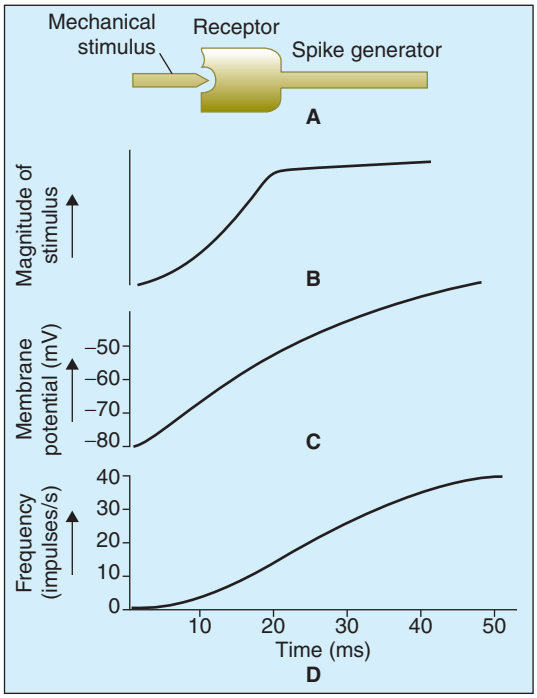


Fig. 10.8-1 Mechanism of development of generator potential and its relationship with intensity of stimulus. A, Stimulus to mechanoreceptor causes its deformation, which opens up channels which are permeable to Na^+ causing membrane depolarization; B, the magnitude of stimulus intensity; C, the receptor potential (generator potential) follows the time course and D, the action potential. *Note.* The magnitude of generator potential and frequency of action potential are proportionate to the magnitude of the stimulus.

2. Summation, i.e. receptor potential from two stimuli can be added if the second stimulus arrives before the receptor potential developed due to first stimulus is over. Thus, receptor potential unlike action potential (which cannot be added) can be added together.

3. Refractory period is not there in the development of receptor potential while the action potential has a refractory period of 1 ms.

4. Local response, i.e. receptor potential cannot be propagated.

5. Duration of receptor potential is greater (approximately 5–10 ms) than action potential (approximately 1–2 ms).

Production of action potential in a sensory nerve

The receptor potential developed in an unmyelinated nerve ending (transducer region) depolarizes the sensory nerve at the first node of Ranvier (spike generator region) by electrotonic depolarization current sink action. When the receptor potential rises above the threshold level (i.e. above 10 mV), it brings the membrane potential of the first node of Ranvier to the firing level causing production of action potential, which is propagated in the nerve fibre (Fig. 10.8-1D). Thus, the first node of Ranvier (spike generator region) converts the graded response of the receptor into action potential. Greater the magnitude of receptor potential, greater is the rate of discharge of action potentials in the nerve fibre.

Recording of receptor potential and action potential

For the purpose of demonstration, the receptor potential can be recorded from pacinian corpuscles because:

- It is a large-sized receptor (Giant receptors),
- It can be easily dissected from the mesentery of experimental animals and
- Its anatomical configuration allows study with microelectrode.

Structure of pacinian corpuscle. A pacinian corpuscle consists of a concentric lamellae of connective tissue surrounding an unmyelinated terminal portion of a nerve fibre. The myelin sheath of the sensory nerve fibre begins inside the corpuscle. Therefore, the first node of Ranvier is located inside the corpuscle but the second node of Ranvier is mostly outside the corpuscle (usually near the point at which the nerve fibre leaves the corpuscle) (Fig. 10.8-2A).

Technique of recording. Recording electrodes (connected to a cathode ray oscilloscope CRO) are placed on the nerve fibre, one on the unmyelinated ending and other on the second node of Ranvier (Fig. 10.8-2A).

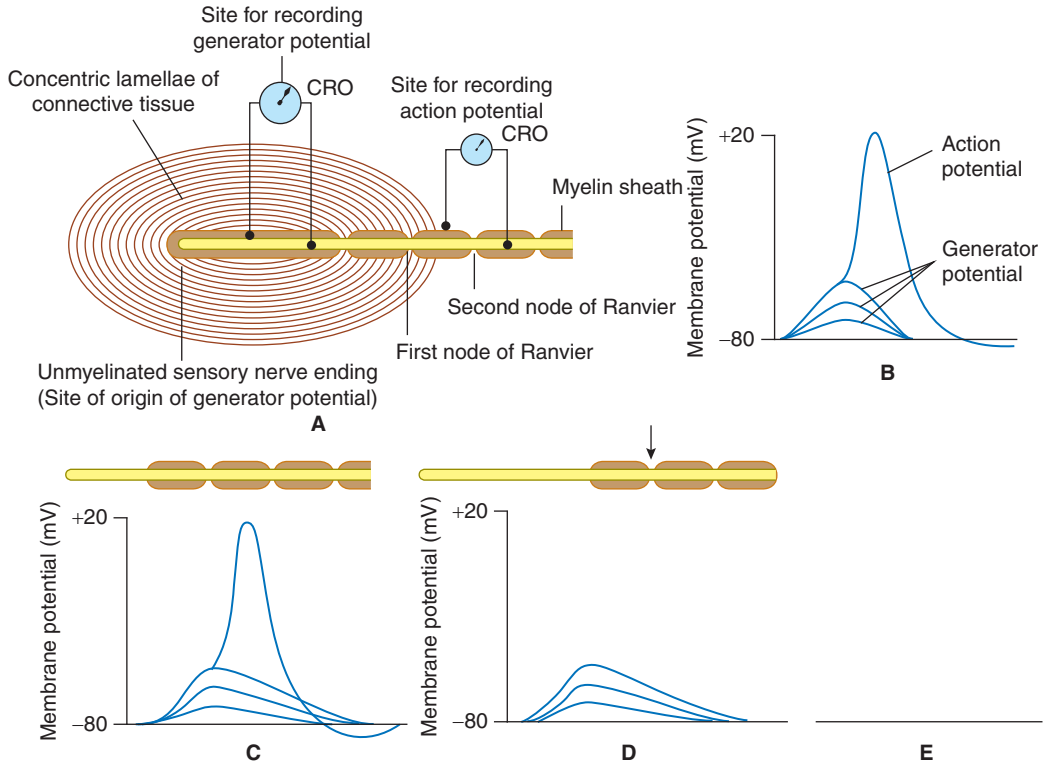


Fig. 10.8-2 Recording of receptor potential from the pacinian corpuscle: A, placement of recording electrodes; B, record of receptor potential and action potential produced by graded pressure to the pacinian corpuscle; C, same response as in B after removal of connective tissue capsule indicates that receptor potential originates from the unmyelinated nerve endings and not the capsule; D, blockage of first node of Ranvier abolishes conduction of receptor potentials produced and E, no response is produced when the sensory nerve is cut.

- When a mild pressure is applied on the corpuscle, a mild non-propagated depolarizing potential, called the generator or receptor potential, can be recorded.
- When the pressure is increased in steps, the magnitude of receptor potential is increased (Fig. 10.8-2B).
- The depolarized segment of the unmyelinated nerve ending produces electrotonic depolarization (current sink action) in the first node of Ranvier.
- When the magnitude of receptor potential is sufficient (above 10 mV), an action potential is generated in first node of Ranvier, which is propagated in the nerve fibre (Fig. 10.8-2B).
- If still greater pressure is applied on the receptor, the frequency of discharge is proportionately increased.

Demonstration of site of receptor potential

The receptor potential originates from the unmyelinated nerve ending and not from the corpuscle or from first node of Ranvier can be demonstrated experimentally as:

- When pressure is applied to the naked unmyelinated nerve ending after removal of the connective tissue of the corpuscle, the receptor potential is still produced but it decays more slowly (Fig. 10.8-2C).
- When pressure is applied to the naked unmyelinated nerve ending after blockage of the first node of Ranvier (by pressure or drug, e.g. narcotics), the receptor potential response persists but action potential cannot be recorded (Fig. 10.8-2D).
- When the sensory nerve is cut and allowed to degenerate, neither the receptor potential nor the action potential can be recorded (Fig. 10.8-2E).

PROPERTIES OF RECEPTORS

1. Specificity of response. Each receptor is easily stimulated (has low threshold) by only one type of appropriate (adequate) specific stimulus. This specificity of response by a particular receptor is also called *law of adequate stimulus*. Although, each receptor is exquisitely sensitive to its adequate stimulus, receptors can respond to other forms of energy if the intensity is high enough. For example:

- *Adequate stimulus* for rod and cones of retina is light. Therefore, retina can detect the presence of a single photon of light. Pressure on the eyeball can also stimulate retinal receptors, but the threshold of these receptors to pressure is much higher than the threshold of the pressure receptors in skin.

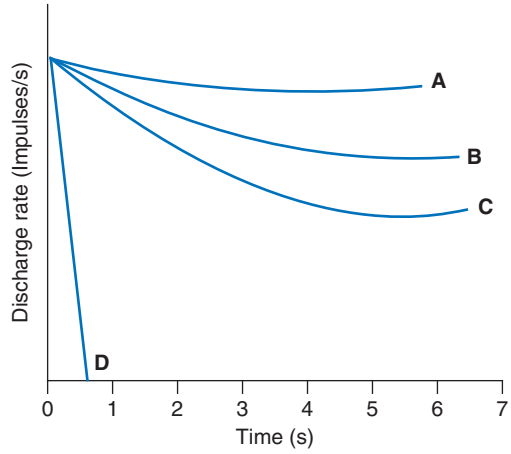


Fig. 10.8-3 Adaptation in sensory receptors with sustained stimulation: A, pain receptors; B, muscle spindles show minimal adaptation; C, thermoreceptors show moderate adaptation and D, touch (Meissner's corpuscles) and pressure (pacinian corpuscle) receptors show most rapid adaptation (phasic receptors).

- *Adequate stimulus for Ruffini's receptors* is warm water at low intensity of stimulus producing a specific response. The warm water at very high stimulus intensity can also stimulate naked nerve ending of pain, but the response produced is not complete.

2. Production of receptor potential on stimulation. As described in detail above (for details see page 756).

3. Adaptation. When a receptor is continuously stimulated with the same strength of stimulus, the receptors respond at a very high impulse rate at first, but the frequency of action potential in its sensory nerve decreases progressively, until finally many of them no longer respond at all in some of the receptors (Fig. 10.8-3). This property is called adaptation. Depending on the rate of adaptation the receptors are of two types:

Tonic receptors. These are slow and incompletely adapting receptors. These receptors keep on firing action potentials continuously during stimulus application. Such receptors are important for life as they keep the brain constantly appraised of status of the body and its relation to its surroundings. Examples of such receptors are:

- *Muscle spindles* are tonic receptors, which continue to discharge as long as muscle is stretched and thus helpful in prolonged postural adjustments.
- *Pain and cold receptors* are tonic receptors, which keep on giving warning to brain about the noxious stimuli till they are present.
- *Baroreceptors and chemoreceptors* are also tonic receptors, which operate continuously in the regulation of blood pressure.

Note. Imagine, if the above receptors would have showed marked adaptation, the life would have not been possible.

Phasic receptors. These are *rapidly adapting* receptors, which fire action potentials at a progressively decreasing rate during stimulus application. These transmit signals only when the stimulus strength is changed. Therefore, number of impulses transmitted is directly proportional to the rate at which the changes take place. Thus, the receptor potential in them is short and decays rapidly. Examples of phasic receptors are:

- Meissner's corpuscles, pacinian corpuscle and olfactory receptors.

Function of adaptation is to decrease the amount of sensory information reaching the brain.

Mechanism of adaptation varies in different receptors, for example:

- *Rods and cones* adapt by changing their chemical composition,
- *Mechanoreceptors* (pacinian corpuscles) adapt due to redistribution of fluid.

Basically, sensory adaptation takes place via two major mechanisms:

- *Failure of transducer mechanism to maintain* a receptor potential despite the continued stimulus application, and
- *Failure of spike generator* to sustain a train of action potentials despite the presence of receptor potential. The decreased excitability of spike generator membrane may be attributed to an increase in the membrane conductance to K^+ or the activity of the electrogenic Na^+–K^+ pump, or the inactivation of Na^+ channels.

4. Effect of strength of stimulus. Receptor potential amplitude depends on the strength of stimulus. During the stimulation of a receptor, if the response given by the receptor is to be doubled, the strength of stimulus must be increased 10 times. This phenomenon is called *Weber Fechner law*.

5. Effect of velocity of stimulus. The magnitude of the receptor potential rises with rate of change of stimulus application. It also applies to removal of stimulus, e.g. off response.

6. Projection. When any part of sensory path is stimulated, conscious sensation referred to the location of receptor is produced. This is called *law of projection* (for details see encoding of sensation).

ENCODING: RECOGNITION OF TYPE OF SENSATION

As discussed above, the sensory receptors transduce all forms of sensory stimuli into a common type of neural signal,

i.e. action potentials which are carried by the peripheral nerves and sensory tracts in the spinal cord and brain stem to the sensory cortex. The question arises how does brain differentiate between the action potentials generated from a touch receptor and a pain receptor and interpret the sensation accordingly. It is believed that the sensory receptors themselves act as peripheral analysers. The intensity, location and quality of a stimulus are encoded as:

1. Encoding of stimulus intensity

The brain interprets different intensities of sensation (i.e. whether the touch is light or heavy; pain is mild, moderate or severe) by following two mechanisms:

- By frequency of action potentials generated in the sensory fibres and
- By number of recruitment of sensory units.

A. By frequency of action potential generated in sensory nerve fibres

The encoding of intensity of stimulus is related to the rate of impulse discharge in the sensory nerve fibres as explained:

(i) The magnitude of receptor potential is directly proportional to the logarithmic increase in the intensity of stimulus. For example, if the response given by a receptor is to be doubled the strength of stimulus must be increased by 10 times.

(ii) The frequency of action potential produced in a sensory nerve is directly proportional to the magnitude of the receptor potential (Fig. 10.8-4).

(iii) From the above statements (i) and (ii), the frequency of action potential (S) in a sensory nerve is directly proportional to the logarithmic increase in intensity of stimulus (I), i.e.

$S = k \log I + C$, where K and C are constants.

The above equation is called '*Weber—Fechner law*' *which states that the magnitude of sensation felt is directly proportional to the log of intensity of the stimulus.*

B. By number of recruitment of sensory units

A single afferent neuron with all its receptor endings makes up a *sensory unit* (Fig. 10.8-5). When the strength of stimulus is increased, it spreads over a large area activating more and more receptors in the neighbouring area and thus more and more *sensory units are recruited* to convey the impulse to brain. This increase in the recruitment of sensory units is interpreted as an increase in the intensity of the stimulus.

2. Encoding of stimulus location

The stimulus location is recognized accurately due to point-to-point representation of the body in the somatosensory cortex. Therefore, when the sensory fibres are experimentally

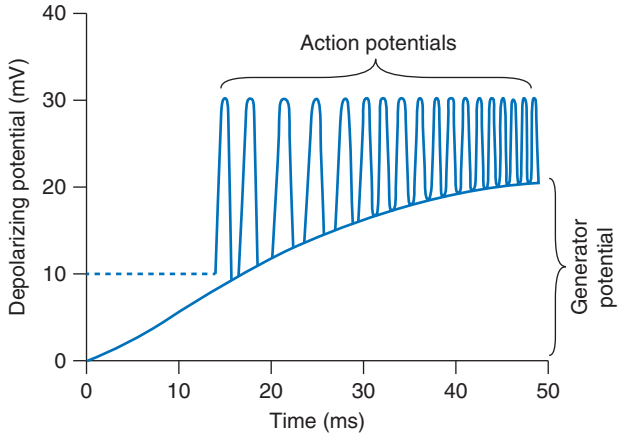


Fig. 10.8-4 Relationship between intensity of stimulus, magnitude of receptor potential and frequency of action potential.

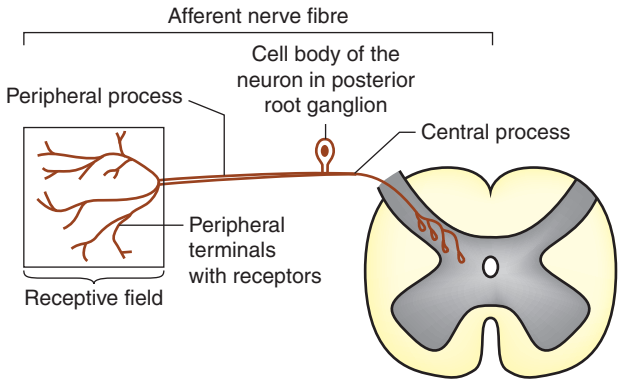


Fig. 10.8-5 Sensory unit and receptive field.

stimulated anywhere in their course to the cortex, the conscious sensation produced is referred to the location of the receptor. This principle is called the *law of projection*. Because of this reason, after amputation of a limb, sometimes patient complains of intense pain in the absent limb (*phantom limb*). These sensations are produced due to irritation of the damaged nociceptive and proprioceptive afferents at the stump of amputated limb. The sensations evoked are projected to the area where receptors are used to be located.

This mechanism of encoding, called *topographic representation*, is also used by visual sensation in addition to the somatosensory system to localize the point of stimulus application. This mechanism of encoding by *topographic representation* is influenced by:

- Receptive field of neurons and
- Phenomenon of lateral inhibition.

Receptive field of neuron

Each sensory neuron receives information from a particular sensory area called its receptive field (Fig. 10.8-5). Generally, the receptive fields of neighbouring neurons overlap and inter-digitate with the areas supplied by others.

The smaller the receptive field, the more precise the encoding of stimulus localization. For example, the ability to distinguish between two adjacent mechanical stimuli to the skin (*two-point discrimination*) is greater on the fingertips and lips where the receptive fields are much smaller and overlap considerably than on the hands and back where the receptive fields are large and widely separated.

Lateral inhibition

Lateral inhibition is a phenomenon by which stronger inputs are enhanced and the weaker inputs of adjacent sensory units are simultaneously inhibited. Stimulus localization can be made more precise by lateral inhibition as explained:

- When two stimuli are applied to the skin, in the absence of lateral inhibition they can be recognized as separate only if they are applied in receptive fields that are separated from each other by a non-stimulated receptive field (Fig. 10.8-6A), otherwise the stimuli will produce equal discharge in all these neurons (Fig. 10.8-6B).
- With lateral inhibition, the neuron with the receptive field in the centre is pre-synaptically inhibited by collaterals from the neurons with receptive fields located laterally. As a result, the receptive field in the centre does not fire, and two stimuli are perceived (Fig. 10.8-6C).

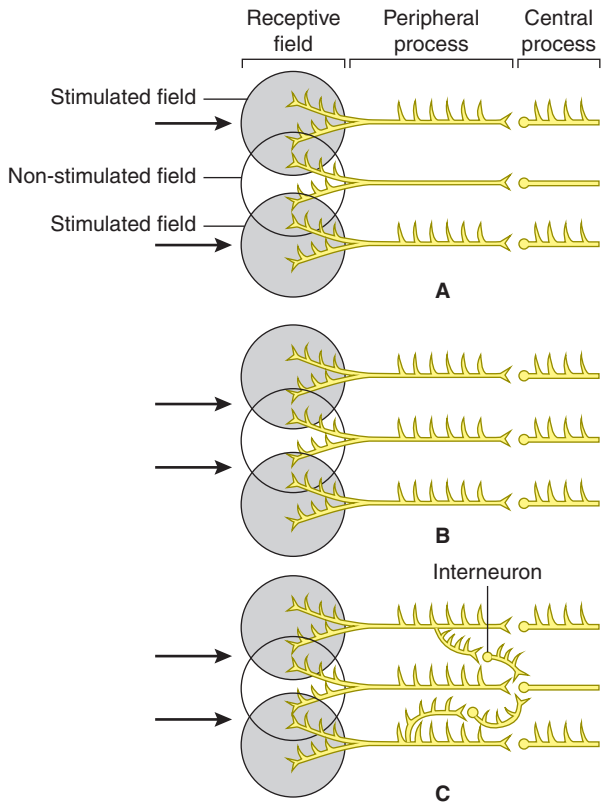


Fig. 10.8-6 Two stimuli can be perceived distinct in the absence of lateral inhibition if they stimulate receptive fields that are separated from each other by a non-stimulated receptive field, A. Otherwise the stimuli produced equal amount of discharge in all the three neurons, B. With lateral inhibition, the neurons with receptive field in the centre are pre-synaptically inhibited by collaterals from the neurons located laterally. As a result, the receptive field in the centre does not fire and two stimuli are perceived, C.

3. Encoding of stimulus quality

In general, action potentials are similar in all nerves, then why stimulation of a touch receptor causes sensation of touch and not of warmth. Similarly, stimulation of photoreceptors causes sensation of light and not of hearing. Stimulus quality is encoded by following mechanisms:

(i) Labelled line mechanism. It is the mechanism in which the stimulus quality is encoded by the particular neural pathway that is stimulated. The basic sensory modalities are encoded by this mechanism. Each fibre or collection of neurons linked by related sensory fibres is referred to as a *labelled line*. For example, action potentials travelling along the fibres and neurons that comprise the anterolateral system (spinothalamic tract) are perceived as pain, whereas action potentials carried over the dorsal column-medial lemniscal system are distinguished as touch or pressure. Further (i) the sensation of touch is elicited whether the receptors on skin are excited by mechanical deformation or by electrical stimulation and (ii) the same type of sensation results, no matter where along the sensory pathway the stimulus is applied.

The specificity of sensory pathway from the receptors to the sensory cortex has been called the *Muller's doctrine of specific nerve energies.*

(ii) Pattern of activity within the neural pathway, that is carrying information to the brain, is used to encode stimulus quality is a more complex mechanism. The two types of pattern coding are:

- In *temporal pattern coding*, the same neuron can carry two different types of sensory information depending upon its pattern of activity. For example, cutaneous cold receptors indicate temperatures below or above 30°C by firing with or without bursts, respectively.
- In *spatial pattern coding*, the activity of several neurons is required to elicit a sensation. For example, three neurons may be required to encode different taste sensations. A sour taste may result if all three neurons are activated, whereas a salty taste may result if only two neurons fire.

(iii) Feature detectors are used in the most sophisticated mechanism of sensory coding. Feature detectors are neurons within the brain that integrate information from a variety of sensory fibres and fire to indicate the presence of a complex stimulus.

- For example, the location of an object in space can be encoded by cortical cells receiving information from a single eye. However, special feature detectors receiving information from both eyes are required to specify the depth of an object in the space.
- Similarly, the location of sound in space requires integration of information from both ears by feature detectors within the brain stem.

SOMATOSENSORY SYSTEM

Somatosensory system can now be discussed in view of the general sensory mechanism under following headings:

- Somatic sensations,
- Pathways in somatosensory system,
- Role of thalamus in somatosensory system and
- Somatosensory cortex.

SOMATIC SENSATIONS

Somatic sensations include sensations of:

- Touch, pressure, two-point discrimination, vibration, stereognosis,
- Temperature,
- Pain and
- Proprioception and kinaesthesia.

TOUCH, PRESSURE AND VIBRATION SENSATIONS

Mechanoreceptors

Mechanoreceptors provide information about touch pressure and vibration stimuli to skin. In general, these receptors consist of an unmyelinated axon surrounded by lamellated connective tissue corpuscles. These include (Fig. 10.8-7):

1. Pacinian corpuscles

Structure. Pacinian corpuscle, about 1–2 µm in diameter, consists of an onion-like concentric lamellae of connective tissue surrounding an unmyelinated terminal portion of a nerve fibre. The myelin sheath of the sensory nerve fibre begins inside the corpuscle. Therefore, first node of Ranvier is located inside the corpuscle but the second node of Ranvier is mostly outside the corpuscle (Figs 10.8-2A and 10.8-7A).

Location. Pacinian corpuscles are located in the skin, subcutaneous tissue, mesentery and in the neighbourhood of tendons and joints.

Function. They *detect tissue vibration* or other extremely rapid changes in the mechanical state of the tissue.

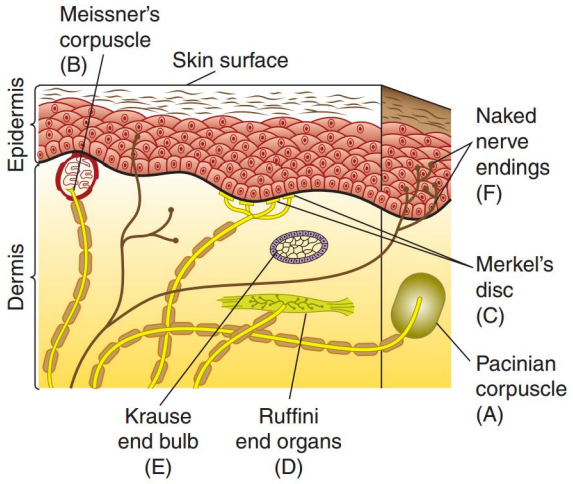


Fig. 10.8-7 Sensory receptors: A, Pacinian corpuscle; B, Meissner's corpuscle; C, Merkel's disc; D, Ruffini end organs; E, Krause end bulb and F, free nerve endings.

Receptive field, stimulus and adaptation. Receptive field of pacinian corpuscles is large. They respond to deformation caused by the mechanical pressure. The pacinian corpuscle is very *rapidly adapting* receptor, so only a few action potentials are generated, regardless of stimulus intensity. The deformation of the nerve terminal is not maintained during continuous stimulus application because the inner lamellae become rearranged. This is called adaptation. That is why we do not feel seat pressure when sitting. A *vibratory stimulus*, however, produces a steady discharge of the pacinian corpuscle. Each time the stimulus is removed and applied, the pacinian corpuscle discharges another action potential.

2. Meissner's corpuscles

Structure and location. They are small encapsulated receptors supplied by A_β type (group II) of myelinated nerve fibres (Fig. 10.8-7B). They are present in non-hairy parts of skin and are abundant at fingertips, lips, nipples and orifices of the body.

Receptive field, stimulus, function and adaptation. They are used to detect rate of stimulus application. They are sensitive to movements of light objects over the surface of skin. The ability to detect the rate of deformation when the skin is moved over an object is especially important to blind individuals using Braille. Meissner's corpuscles rapidly adapt to a maintained or slowly applied stimulus.

3. Merkel's discs

Structure and location. Merkel's discs are unique because the transducer is not on the nerve terminal but on the epithelial cells that make up the disc (Fig. 10.8-7C). The epithelial sensory cells form synaptic connections with

branches of a large single group II afferent myelinated fibre. They are present in areas where Meissner's corpuscles are present, i.e. in abundance at fingertips, lips, nipples and orifices of the body.

Receptive field, stimulus, function and adaptation. Merkel's discs are *slowly adapting* receptors with *small receptive field* that is used to detect the *location of a stimulus*. They along with Meissner's corpuscles play an important role in localizing touch sensations and also in determining the texture of what is felt. Therefore, they are also called *tactile receptors*.

4. Ruffini's end organs

Structure and location. They are multibranched encapsulated endings (Fig. 10.8-7D). The receptor is located on the terminal of a group II axon that is covered by a liquid-filled collagen capsule. Collagen strands within the capsule make contact with the nerve fibres and overlying skin. They are present in the deeper layers of skin and also in the deeper tissues.

Receptive field, stimulus, function and adaptation. Ruffini's end organ is *slowly adapting* receptor with a *large receptive field,* which is used to detect the *magnitude of stimulus*. Since they adapt very little, so continuously signal the state of deformation of the skin and deeper tissues. They are present in joint corpuscles where they detect degree of joint rotation.

5. Hair end organs

Structure and location. Each hair and its basal nerve fibre forms hair end organ.

Stimulus and function. Hair end organ is stimulated by slight movement of hair. These receptors mainly detect the movement of objects on the surface of the body.

6. Krause's end bulbs

Structure and location. They are spherical mechanoreceptors (Fig. 10.8-7E). Their afferent fibres belong to the Aδ group. They are present in the conjunctiva, in the papillae of lips and tongue, in the skin of genitalia and in the sheath of nerves.

Stimulus and function. They detect touch and pressure.

7. Free nerve endings

Structure and location. These are terminal branches of thin myelinated As or unmyelinated C fibres (Fig. 10.8-7F). They are present everywhere in the skin and in many other tissues.

Stimulus and function. As mechanoreceptors, they detect touch and pressure.

Salient features of mechanoreceptors

Salient features of main mechanoreceptors are summarized in Table 10.8-1.

Functions of touch and pressure mechanoreceptors

1. Detection of touch, pressure and vibration sensations by mechanoreceptors

Touch, pressure and vibration are considered to be different form of the same sensation.

- *Pressure* is felt when the force applied on the skin is sufficient to reach the receptors located in the deeper layers of skin.
- *Touch* is felt when the force is insufficient to reach the deeper layers.
- *Vibrations* are felt when there are rhythmic vibrations in the force.
- *Detection of touch, pressure or vibration sensation by a mechanoreceptor* depends, among other factors, on whether they are rapidly adapting or slowly adapting receptors.

Table 10.8-1 Salient features of cutaneous mechanoreceptors

Receptors	Main descriptive feature	Receptive field size	Adaptation sensation	Encoded
Pacinian corpuscles	Onion-like capsule surrounding unmyelinated nerve ending	Large	Very rapid	Vibration, tapping
Meissner's corpuscles	Small encapsulated, present in non-hairy skin	Small	Rapid	Speed of stimulus application
Merkel's disc	Transducer is on epithelial cells	Small	Slow	Location of stimulus
Ruffini's end organs	Multibranched Encapsulated Liquid-filled collagen corpuscle	Large	Slow	Magnitude and duration of stimulus (pressure)
Hair end organs	Hair and its basal nerve	Small	Rapid	Movement of object on the surface of body
Krause's end bulbs		Small	Rapid	Touch and pressure

2. Two-point discrimination

It is the ability to distinguish two touch stimuli separately. It depends upon the interaction of touch sensibility and parietal lobe. The minimum distance by which two touch stimuli can be perceived as separate stimuli varies from 2–3 mm on the lips and fingertips to over 60 mm on the back of the trunk. The difference in the distance between the two points seems to be related to the density of touch receptors in different parts of the body.

3. Stereognosis

Stereognosis refers to the ability to recognize familiar objects, such as a key, coins, pen, pencil, spoons, etc. by merely handling them without looking at them. Touch and pressure receptors are involved in this sensation, but cerebral cortex (somatic sensory association area) plays a major role.

Astereognosis, i.e. loss of stereognosis is an early sign of damage to the parietal lobe when touch–pressure sensation is normal.

Neural transmission

The touch–pressure sensation from the mechanoreceptors is carried to central nervous system (CNS) by:

- A (β and δ) sensory fibres and
- Unmyelinated C fibres also conduct some touch impulses.

Spinal cord tracts which carry touch–pressure sensations are two lemniscus systems (dorsal column and ventral spinothalamic tract).

Dorsal column carries sensations of:

- Fine touch (touch with low threshold excitation),
- Detailed tactile localization,
- Two-point discrimination and
- Stereognosis.

For details of dorsal column see page 701.

Lesions of dorsal column are therefore associated with elevation of touch threshold and loss of above sensations.

Ventral spinothalamic tracts carry sensations concerned with gross tactile sensations of crude touch. For details see page 699.

Lesions of ventral spinothalamic tract are therefore associated with a slight touch deficit. The touch localization remains normal.

PROPRIOCEPTIVE AND KINAESTHETIC SENSATIONS

Proprioceptive and kinaesthetic sensations arise from the muscles, tendons and joints. These include:

1. Proprioceptive sensations. These are concerned with physical state of the body, i.e. sense of position, tendon and muscle sensations, deep pressure and sense of equilibrium.

2. Kinaesthetic sensations. Kinaesthesia is the conscious recognition of rate of movement of different parts of the body. Kinaesthetic sensations include:

- Conscious kinaesthetic sensations and
- Unconscious kinaesthetic sensations.

Receptors concerned

The receptors concerned are called *proprioceptors* and include:

- Muscle spindle or stretch receptors (see page 821),
- Joint receptors located in the joint capsules and ligaments around the joints. Ruffini's end organs are the most important receptors for this function. A few pacinian corpuscles are also involved.
- Golgi tendon organ (see page 828) and
- Vestibular receptors (see page 843).

Neural transmission

Sensations from the above said receptors are carried by the myelinated nerve fibres (group I and II) in the peripheral nerves.

Neural pathway involved is:

- *Conscious sense of position, vibration and deep pressure* is carried by axons from joint receptors and pacinian corpuscles. These enter via posterior root, branch and enter the *dorsal column* of the same side. From spinal cord, ultimately these sensations reach the somatic sensory cortex.
- *Unconscious proprioceptive information* arising from the muscle spindles, Golgi tendon organs and joint receptors travel through group Ia and Ib fibres in peripheral nerves and through spinocerebellar tracts and dorsal column in the spinal cord. From the nuclei gracilis and cuneatus, fibres concerned with unconscious proprioception reach the cerebellum as external arcuate fibres. Spinocerebellar tracts enter the cerebellum through inferior peduncle. For details see page 719.

TEMPERATURE SENSATION

Thermoreceptors

Structure. Thermoreceptors refer to a special type of free nerve endings, which are responsible for detecting temperature sensation. Separate receptors with discrete receptive fields exist for encoding warm and cold sensations and are called warm receptors and cold receptors, respectively. Free nerve endings of unmyelinated C fibres form the *warm receptors* and that of *small myelinated* As fibres form the *cold receptors.*

Location. Thermoreceptors are located in the skin of all parts of the body. However, density of thermoreceptors is

greatest in the lips, moderate in the fingertips and least in the skin of trunk.

Receptive fields. The receptor fields of thermoreceptors (unlike that of mechanoreceptor) do not show any overlap, probably because precise localization of thermal stimulus is rarely important to the body. Because of the lack of overlap, it is possible to delineate distinct *hot spots* (areas having warm receptors) and *cold spots* (areas having cold receptors) on the skin that respond to warmth and cold, respectively. In any area of the body, number of cold spots is about 4–10 times the number of hot spots.

Stimulus. Thermoreceptor responds to the temperature of subcutaneous tissue surrounding them and not to the environmental temperature as such. Because of this reason, cold metal objects feel colder than wooden objects of the same temperature. The metal being good conductor conducts heat away from the skin more rapidly and cools the subcutaneous tissue to a greater degree than the wood. Similarly, the alcohol-induced cutaneous vasodilatation gives a feeling of warmth, even when the person is exposed to extreme cold.

The salient features of response exhibited by warm and cold receptors are:

Warm receptors

They are activated when skin temperature is between 30 and 43°C (Fig. 10.8-8A).

- The steady state firing rate of warm receptors reaches a peak at temperatures of approximately 42°C (Fig. 10.8-8A).
- Warm receptors transiently increase their firing rate when skin temperature increases and decrease their firing rate when skin temperature decreases (Fig. 10.8-8B). This is because the sensation produced by a small change in temperature depends on the current skin temperature. For example, a stimulus of 35°C feels warm if the skin is at 30°C, and cools if the skin is at 40°C.

Cold receptors

- They are activated when the skin temperature is between 10 and 40°C (Fig. 10.8-8A).
- The steady state firing rate of cold receptors reaches a peak at temperatures between 25 and 30°C (Fig. 10.8-8A).
- Cold fibres transiently increase their firing rate when temperature decreases and transiently decrease their firing rate when skin temperature increases (Fig. 10.8-8B).
- Paradoxically, temperatures between 45 and 50°C stimulate cold fibres as well as pain fibres producing mixed sensations of cold and pain (Fig. 10.8-8A).
- Cold temperature below 10°C stimulates only pain receptors (Fig. 10.8-8A).
- Thus, between 30 and 40°C both cold and warm receptors are stimulated, which help the person in fine gradation of temperature. Therefore, between 30 and 40°C (neutral or comfort zone), complete perceptual adaptation occurs (i.e. awareness of temperature disappears).

Adaptation

Thermoreceptors show a moderate degree of adaptation (Fig. 10.8-3C). Therefore:

- On exposure to cold, when skin temperature begins to fall, initially the person feels much colder than at a later

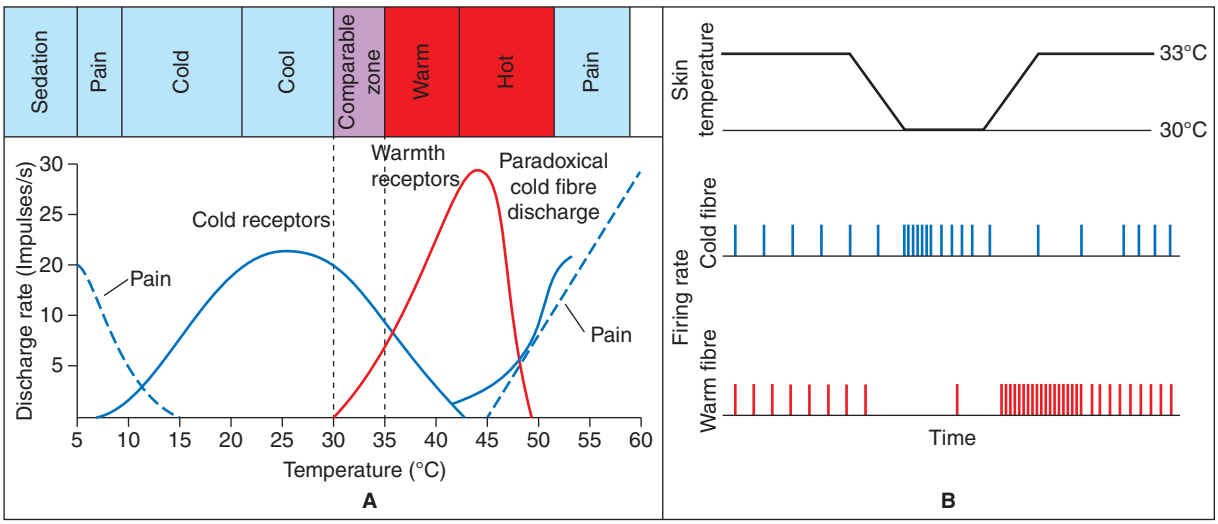


Fig. 10.8-8 A, Impulse discharge rate of cold and warm receptors as a function of temperature and B, spikes train illustrating the dynamic response of warm and cold receptors to a change in temperature. When the temperature decreases, cold fibres increase their firing rate transiently and then adapt to firing rate. Similarly, when temperature increases warm fibres transiently increase their firing rate before adapting to rate indicated by the graph.

stage, even when exposed to same cold environments. This is because when the temperature decreases, cold fibres increase their rate of firing and then adapt to the firing rate (Fig. 10.8-8B).

- Similarly, on sudden exposure to hot environment, the feeling of warmth is more intense in the beginning. This is because, when the temperature increases, warm receptors increase their firing rate before adapting to the rate indicated in the graph (Fig. 10.8-8B).

Neural pathway

The impulses from cold receptors are carried by Aδ myelinated fibres and those from the warm receptors are carried by unmyelinated C fibres in lateral spinothalamic tract (see page 699). In the CNS, the lateral spinothalamic tract and the medial lemniscus carry impulses to thalamus. Ultimately, impulses reach the somatosensory cortex.

PAIN SENSATION

Definition and purpose

Definition. Pain refers to an unpleasant sensory and emotional experience associated with actual or potential tissue damage. The word pain has been derived from a Greek word *Poena* meaning 'penalty or punishment'.

Purpose. Pain sensation is different from other sensations because its purpose is not to inform the brain about the quality of a stimulus, but rather to indicate that the stimulus is physically damaging. Therefore, though the pain sensation is unpleasant, it is useful in the following ways:

- It makes one aware of a harmful agent in close contact with the body and body gives preferential treatment to this information.
- It causes the individual to react to remove the pain stimulus to prevent further damage to the tissues.
- Pain receptors are non-adaptable receptors; therefore, they keep the person apprised of damaging stimulus as long as it persists. Thus pain sensation has a protective function.

Pain receptors, stimuli and chemical mediators of pain

Nociceptors

Nociceptor is the name given to receptors of pain to indicate that they respond to noxious stimuli. The noxious stimuli can be damaging or potentially damaging mechanical, chemical and thermal stimuli.

Structure. Nociceptors refer to special type of free nerve endings of two types of nerve fibres:

- Aδ myelinated nerve fibres and
- C unmyelinated nerve fibres.

Table 10.8-2 Characteristic features of Aδ fibres and C fibre nociceptors

S. No.	Feature		Aδ fibre nociceptors	C fibre nociceptors
1.	Number	:	Less	More
2.	Myelination	:	Myelinated	Unmyelinated
3.	Diameter	:	2–5 μm	0.4–1.2 μm
4.	Conduction velocity	:	12–30 m/s	0.5–2 m/s
5.	Specific stimulus	:	Most sensitive to pressure (Mechanoreceptor)	Most sensitive to chemical agents like: • Local anaesthetics, • Histamine, • Kinins and • prostaglandins
6.	Impulse conduction	:	Conduct impulses only in response to noxious stimuli (Fast component of pain)	Conduct impulses in response to thermal and mechanical stimuli and slow component of pain
7.	Sensitivity to electrical stimulus	:	More	Less

The differences between two types of nociceptors are given in Table 10.8-2.

Location. High density of pain receptors is present in the superficial layers of skin and in many deeper tissues like periosteum, joints, arterial wall and falx and tentorium in the cranium. Parenchyma of liver and alveoli of lungs are insensitive to pain, but liver capsule, bronchi and parietal pleura are very sensitive to pain. Most other deeper tissues have relatively sparse pain nerve endings, but widespread tissue damage always results in pain, even in these areas.

Types. Nociceptors broadly can be grouped as somatic nociceptors and visceral nociceptors.

1. *Somatic nociceptors* are free nerve endings of Aδ and C fibres as mentioned above.
2. *Visceral nociceptors:* There is little evidence for specialized pain receptors in viscera. Visceral pain is often due to excessive tension on the nerve endings in the smooth muscles, i.e. probably stretch receptors produce pain when stimulated to high firing rates by intense stimuli. For example, pain due to uterine contractions during child birth, or pain due to colics of alimentary, biliary or urinary tracts.

Pain stimuli

Pain receptors are activated by three types of noxious stimuli: mechanical, thermal and chemical.

Mechanical and thermal stimuli tend to elicit fast pain. *Fast pain* is felt when a needle is struck into the skin, when the skin is cut with a knife or when the skin is acutely burned. It is also felt when the skin is subjected to electric shock. Fast, sharp pain is not felt in most of the deeper tissues. Mechanical and thermal stimuli, however, can also elicit slow pain.

Chemical stimuli usually tend to elicit slow suffering type of pain that occurs after tissue injury, although this not always is the cause.

As mentioned earlier, pain sensation is associated with actual or potential tissue damage caused by noxious stimuli. Damaged tissue releases certain chemicals, which act on nociceptors and cause pain sensations. *Chemical mediators of pain include:*

- K^+, *ATP* and *ADP* are released following cell death.
- *Bradykinin* is formed by reaction of certain circulating globulins with proteolytic enzymes released by dying cells. It is most powerful in causing tissue damage pain.
- *Leukotrienes* are released from mast cells.
- *Serotonin* is released from platelets.
- *Histamine* is released from mast cells.
- Accumulation of *lactic acid* in tissues due to the anaerobic mechanism during ischaemia also stimulate nociceptors and cause pain.
- *Prostaglandins* are mediators of pain, fever and inflammation. These are synthesized by enzyme cyclooxygenase, which is induced in peripheral tissues by cytokines, growth factors and other inflammatory stimuli. Prostaglandins and substance P enhance the sensitivity of pain endings but do not directly excite them.
- Activation of a nociceptive nerve terminal stimulates the axon reflex and releases *substance P* and *calcitonin gene-related peptide* from other terminals of the same nociceptive nerve fibre.
- Nociceptin.

IMPORTANT NOTE

Recently, vanilloid receptors (VRL) have been isolated. Vanilines are a group of compounds that cause pain, capsaicin is included in this group. VRL receptors respond to capsaicin and temperature above 43°C but VRL-1 responds to temperature (more than 43°C but not to capsaicin).

Qualitative types of pain sensations

Qualitatively pain sensations are of two types: fast pain and slow pain.

1. Fast pain

Fast pain is a sharp, well-localized, pricking sensation that results from the activation of the nociceptors on the Aδ fibres. The fast pain sensations travel faster and thus appear within 0.1 ms after the application of stimulus. It is carried by Aδ fibres, which have a small receptive field and a topographic representation in the cortex.

Accompaniments of fast pain are:

- *Withdrawal reflex*, which causes the individual to move the involved body part away from the source of painful stimulus.
- *Sympathetic response*, i.e. increased blood pressure, tachycardia and mobilization of body energy supply.

2. Slow pain

Slow pain is poorly localized, dull, throbbing, burning sensation that results from activation of nociceptors on the *C fibres*. It appears after 1 s or more following the application of stimulus. It is carried by C type of nerve fibres, which are unmyelinated fibres.

Accompaniments of slow pain are:

- *Emotional perception* in the form of unpleasantness and in long-standing cases irritation, frustration and depression.
- *Autonomic symptoms* in the form of nausea, profuse sweating, vomiting and lowering of blood pressure.
- *Generalized reduction in the skeletal muscle tone.*

Clinical types of pain

In clinical practice, pain sensations can be classified as:

- Somatic pain,
- Visceral pain,
- Referred pain,
- Radiating pain and
- Projected pain.

1. Somatic pain

Somatic pain, as the name indicates, arises from the tissues of the body other than viscera. It is of two types:

Superficial somatic pain arises from the skin and superficial tissues. Its features are usually similar to the *fast pain.*

Deep somatic pain arises from the muscles, joints, bones and fascia. Usually, its features are similar to that of *slow pain.*

2. Visceral pain

Visceral nociceptors see page 805.

Features of visceral pain are:

- *Poorly localized* because pain receptors in viscera are comparatively few.
- *Unpleasant* because of emotional perception.

- *Autonomic symptoms* in the form of nausea, vomiting, profuse sweating and lowering of blood pressure.
- *Reflex contraction of skeletal muscle* of abdominal wall, clinically known as *guarding*, is a common association especially when inflammation of viscera involves peritoneum. It is a protective reflex which helps to protect the underlying inflamed structures from an unintentional injury.
- *Radiates* or is referred to other site (see referred pain).

Common causes of visceral pain are:

1. *Inflammation* of the viscera, e.g. appendicitis, cholecystitis, pancreatitis, etc.
2. *Overdistension of hollow viscer*a, e.g. intestinal distension in intestinal obstruction, urinary bladder distension in urinary obstruction, etc.
3. *Spasm of hollow viscus:* Pain is caused due to mechanical stimulation of pain endings and ischaemia. For example, pain due to uterine contraction during child birth, pain due to colics of alimentary, biliary or urinary tracts.
4. *Chemical stimuli:* Damaging substances may leak from the gastrointestinal tract into the peritoneal cavity, e.g. gastric acid leaking through perforated gastric or duodenal ulcer.
5. *Ischaemia* as occurs in tractions on mesentery. Pain is due to acidic metabolic end products or tissue degenerative products, such as bradykinin and proteolytic enzymes.

Neural pathway. Visceral pain sensation is carried by unmyelinated type C afferent fibres in the sympathetics (from most of the viscera) and in the parasympathetic (from many pelvic viscera) nerves. Their cell bodies are located in the dorsal roots and the homologous cranial nerve ganglion.

In the CNS, visceral pain fibres travel along with somatic pain fibres in the spinothalamic tract and medial lemniscus.

3. Referred pain

Referred pain as the name indicates is that pain which originates due to irritation of a visceral organ and is felt not in the organ but in some other somatic structure (usually skin) supplied by the same neural segment.

Characteristic features of referred pain are:

1. Such a pain is said to be referred to the second structure. For example:
   - In *myocardial ischaemia*, pain is referred to the left shoulder and arm.
   - Pain due to *stone in lower part of ureter* is usually referred to the corresponding testis and inner thigh.
   - *Inflammation of diaphragm* secondary to pleurisy or severe cholecystitis produces pain at the tip of shoulder.
2. Because the skin is topographically mapped and the viscera are not, the pain is identified as originating on the skin and not within the viscera.

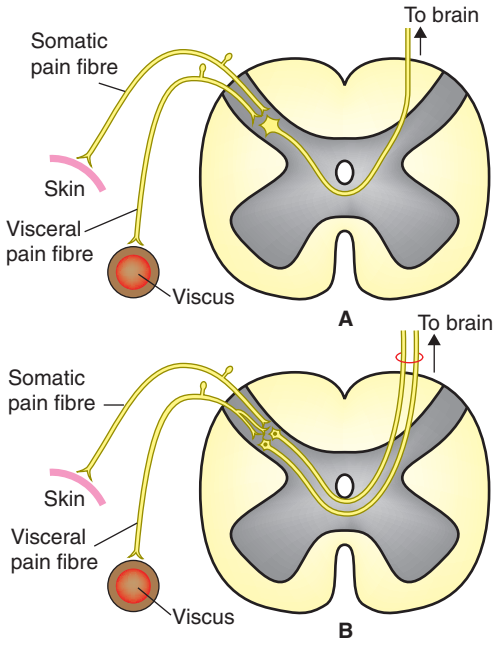


Fig. 10.8-9 Theories of referred pain: A, convergence theory and B, facilitation theory.

3. Pain is usually referred to a structure with common embryonic origin and hence is innervated by a common neural segment. This principle is called the *dermatomal rule*. For example, embryologically the heart and the left arm have the same segmental origin. Similarly, the testes and kidney develop from the same primitive urogenital ridge.

Theories of referred pain are:

1. Convergence theory: According to this theory, when the first-order neurons carrying pain sensation from a somatic area and a visceral organ converge on a common second-order neuron (Fig. 10.8-9A), the brain is unable to identify the source of pain. Since somatic pain is far more common, the brain interprets all pain as somatic pain even when the source is actually visceral.

2. Facilitation theory: According to this theory, the visceral irritation is inadequate for producing pain by itself. However, it facilitates pain fibres from the somatic structures (Fig. 10.8-9B), so that even minor somatic irritation produces perceptible pain.

4. Radiating pain

Sometimes visceral pain is experienced both locally and also at distant point (referred pain). In fact, pain seems to spread from the local area to the distant site. This is called radiating pain. Example of radiating pain is:

- *In appendicitis* pain starts in the right iliac fossa and radiates towards centre of abdomen.

5. Projected pain

When the sensory fibres carrying pain sensations are stimulated anywhere in their course to the sensory cortex, the pain sensations evoked are projected to the area where receptors are located called *projected pain*. Projected pain follows the *law of projection* (see page 799). Examples of projected pain are:

- After amputation of a limb, sometime patient complains of intense pain in the absent limb (phantom limb). The pain sensations are produced due to irritation of nociceptive fibres at the stump, but are projected to the area where receptors used to be located.
- Striking the elbow causes pain to be projected to the hand.

6. Hyperalgesia

Hyperalgesia refers to an enhanced painful response to a stimulus. It is of two types:

(i) Primary hyperalgesia. In it the noxious stimuli produce more severe pain than expected. It occurs over an area of tissue damage. The pain threshold is lowered, so that even non-noxious stimuli (e.g. touch) produce pain (*allodynia*). The movement-related symptoms of osteoarthritis and touch evoked pain of herpetic neuralgia are both examples of mechanical allodynia. Primary hyperalgesia is due to release of algogenic pain-producing substances like histamine, 5-HT, plasma kinin and prostaglandins from the damaged tissues.

(ii) Secondary hyperalgesia refers to the occurrence of far more severe pain than expected in response to noxious stimulus applied to normal healthy skin. In this condition, there is no lowering of pain threshold. Secondary hyperalgesia has been explained to result due to the phenomenon of subliminal fringe. Primary pain afferents from an area of tissue damage not only stimulate the appropriate second-order neurons to threshold level producing pain and primary hyperalgesia, but also excite the second-order neurons belonging to nearby area to subthreshold level. Hence application of noxious stimulus produces more intense pain in this area.

Note. Nociceptin is an opioid-like polypeptide and has no binding affinity for opioid receptors. It causes hyperalgesia when injected intracranially in the experimental animals. It probably has a role in pain transmission.

7. Causalgia

Causalgia is a condition in which spontaneous burning pain sensation occurs after a long time in the area of even trivial injuries. It is also accompanied by *hyperalgesia* and reflex sympathetic dystrophy.

Reflex sympathetic dystrophy means sympathetic discharge reflexly causes pain in the injured skin area. The exact cause is not known, but research in the animals reveals that:

- In the affected area the skin becomes thin, hair growth increases and nerve injury leads to sprouting of the sympathetic nerve fibres. The overgrowth of sympathetic (noradrenergic) endings enters into the dorsal root ganglia of the spinal nerves. Therefore, discharge of these noradrenergic endings stimulates the altered circuitry of nerve fibres in the skin.
- Use of α adrenergic blocker helps in relief of causalgia-type pain.

Neural pathway and perception of pain sensations

Two separate pathways exist for transmission of fast and slow pain to the brain.

Pathway for the fast pain

In peripheral nerves. Fast pain signals are transmitted from Aδ fibres at velocities between 6 and 30 m/s to dorsal root ganglion and then enter the spinal cord at dorsal root of spinal nerve (formed by axons of cells of dorsal root ganglion) (Fig. 10.8-10).

In the spinal cord. Aδ fibres ascend or descend for one or two segments in the tract of Lissauer lying immediately posterior to the dorsal horn and then terminate into neurons of lamina I. These neurons give rise to fibres, which immediately cross to the opposite side of the cord through anterior commissure (Fig. 10.8-10) and then pass upwards to the brain in the anterolateral columns as *neospinothalamic tract.*

In the brain stem, a few fibres from the neospinothalamic tract terminate in the reticular formation, but most of them pass upwards to the thalamus (Fig. 10.8-10).

In the thalamus, most of the fibres project to the ventral posterolateral (VPL) nucleus. From here, thalamic neurons project to the primary sensory cortex (Fig. 10.8-10).

This system is primarily used in the localization of pain stimuli when tactile receptors are also stimulated along with fast pain fibres, localization of fast pain is exact. If only pain receptors are stimulated, localization is poor.

Pathway for slow pain

In peripheral nerves, slow pain impulses are carried by slow conducting unmyelinated fibres at velocities ranging from 0.5 to 2 m/s to the dorsal root ganglion and then enter the spinal cord at dorsal root of spinal nerve (formed by axons of cells of dorsal root ganglion) (Fig. 10.8-10).

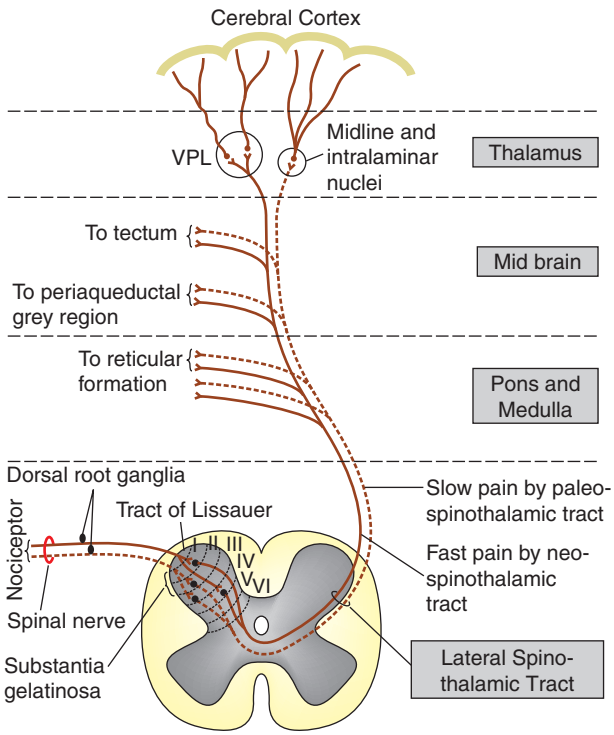


Fig. 10.8-10 Neural pathway of fast and slow pain.

In the spinal cord, the C fibres terminate in the laminae II and III of the dorsal horn. Laminae II and III are together known as substantia gelatinosa.

From here fibres go to lamina V of dorsal horn. Axons of neurons of lamina I of dorsal horn, which receive impulses from the C fibres cross the midline near their level of origin from the *paleospinothalamic tract* which passes upwards to the brain in the anterolateral column along with the fibres of fast pain.

In the brain stem, these fibres terminate very widely mainly in the reticular formation and also in superior colliculus and periaqueductal grey (PAG) region. A system of ascending fibres, mainly from the reticular formation, proceeds rostrally to the intralaminar nuclei and posterior nuclei of thalamus, as well as to the portion of hypothalamus. The intralaminar nuclei of thalamus in turn relay activating signals to all parts of the brain (Fig. 10.8-10).

IMPORTANT NOTE

1. Transmission of pain signals through two routes explains why a single prick with a sharp needle produces almost immediately sharp localized pain, followed about 1 s later by slowly increasing painful sensation, which lasts many seconds and sometimes even minutes.
2. The fact that brain stem reticular areas and the intralaminar thalamic nuclei that receive input from the paleospinothalamic pathway are part of the brain stem activating or alerting systems may explain why individuals with chronic pain syndromes have difficulty in sleeping.

Perception of pain sensations

Perception of pain is the phenomenon by which noxious stimuli reach consciousness. It involves two components:

- Nociceptive component, and
- Affective (cognition and attention) component.

Nociceptive component of pain perception. Pain perception occurs at subcortical levels, i.e. in the thalamus and in the reticular formation of the brain stem. However, somatosensory cortex helps in exact and meaningful interpretation of quality and localization of pain.

Affective (cognitive and attention) component of pain perception is the psychological component. It involves the activity of spinothalamic tracts—limbic system pathway.

Cognitive perceptions are those abilities that recognize, discriminate, memorize or judge afferent information. It involves patient's ability to relate a painful experience to another event, e.g. pain experienced in a pleasant environment elicits a less intense response than an experienced in a setting of depression.

Attention plays a role in the perception of pain on the basis that only a fixed number of afferent stimuli can reach the cortical centres. Therefore, if a patient in pain concentrates on a separate and unrelated image, e.g. getting deeply involved in music or an interesting movie on television, it is possible that he will perceive lesser intensity of pain than otherwise.

IMPORTANT NOTE

The biofeedback and hypnosis, for their positive impact on pain, operate on this principle.

Pain suppression systems in CNS

The degree of reaction to painful stimuli varies from individual to individual, mainly because of existence of pain suppression systems in the CNS. The pain suppression consists of two major components:

- Spinal pain suppression system and
- Supraspinal pain suppression system.

A. Spinal pain suppression system

There exists a pain inhibitory complex in dorsal horn of spinal cord, which blocks the pain signals at the initial entry point to the spinal cord.

Gate control hypothesis has been put forward by Melzack and Wall in 1965 to explain the working of spinal pain suppression system. According to this hypothesis, the dorsal grey horn acts as a gate for transmission of pain sensation and this gate can be partly or completely closed by:

- segmental suppression and
- supraspinal suppression.

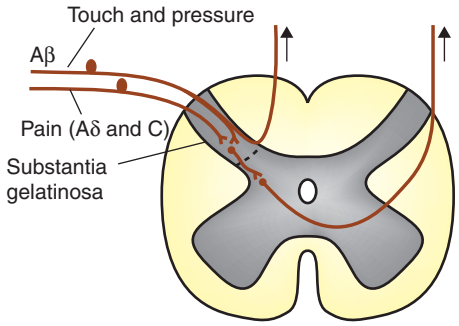


Fig. 10.8-11 Spinal pain suppression system. Note the collateral from Aβ fibres from touch receptors cause presynaptic inhibition of pain afferent Aδ and C fibres.

1. Segmental suppression: It has been observed that the activation of large myelinated touch fibres (Aβ) reduces pain. It is called the gating of pain and occurs because after entering the spinal cord, the Aβ fibres give collaterals, which cause *pre-synaptic inhibition* (primary afferent depolarization) of pain carrying both type C and Aδ fibres, where they synapse in the dorsal horn (Fig. 10.8-11). This is done by blocking calcium channels in the membranes of nerve terminals. Although poorly understood, such circuitry probably explains the relief of pain achieved by following manoeuvres:

- Rubbing or massage or pressure in the vicinity of painful area.
- Local application of warmth or cold.
- Local application of counterirritants, i.e. stimulation of skin.
- Acupuncture and
- Transcutaneous electric nerve stimulation in which pain site or the nerves leading from it are stimulated by electrodes placed on the surface of skin.

2. Supraspinal suppression is caused by the supraspinal suppression system described below.

B. Supraspinal pain suppression system

There exist three different supraspinal descending pain modulation pathways:

- Descending serotonergic and opioid inhibitory system,
- Descending purinergic inhibitory system and
- Descending adrenergic inhibitory system.

1. Descending serotonergic and opioid inhibitory system. It is the most important supraspinal pain inhibitory system. Components of this system (Fig. 10.8-12) are:

(i) Raphe magnus nucleus (RMN). It is a thin midline nucleus located in the lower pons and upper medulla. Its neurons receive innervation from the PAG reticular formation, hypothalamus and frontal cortex. The *serotonergic neurons* of the RMN project down the dorsolateral column to influence the neurons in dorsal horn of spinal cord, which are

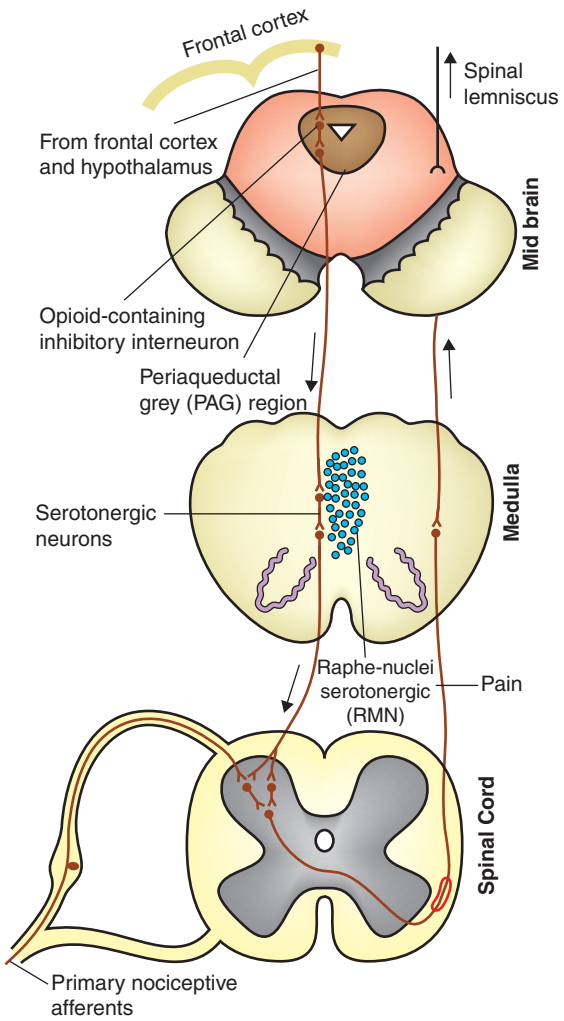


Fig. 10.8-12 Supraspinal serotonergic and opioid pain inhibitory system.

excited by primary nociceptive afferents. The serotonergic fibres exert their effect by *post-synaptic inhibition*.

(ii) Periaqueductal grey area in the mid brain. It inhibits pain by stimulating the RMN (Fig. 10.8-13). Neurons of PAG have opioid receptors on their surface membranes. When *opioid receptors* are stimulated by exogenously administered opioid compounds (analgesics) or by endogenous opioid neurotransmitters (endorphins and enkephalins) found in the brain, the pain suppression circuitry is activated and this leads to reduced pain perception.

Note. It has been observed that electrical stimulation in PAG area produces profound analgesia. Analgesia produced by electrical stimulation is reversed by *naloxone*, implicating endogenous opioid peptides as mediators.

- *Opioid receptors* are also present on the membrane surface of the terminals of primary afferent pain carrying fibres, which terminate in substantia gelatinosa of dorsal horn. These neurons secrete *substance* P as neurotransmitter. The opioid peptides (endorphin and enkephalin)

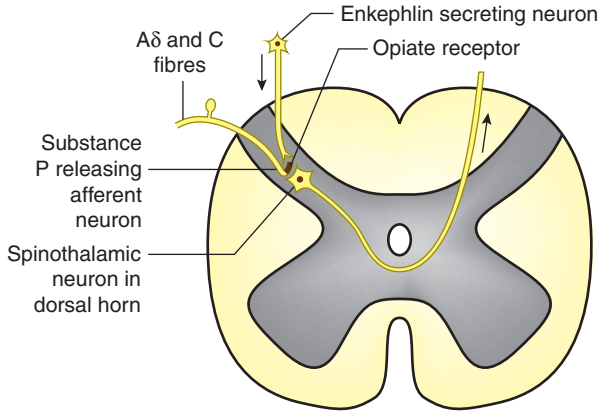


Fig. 10.8-13 Location of opioid receptors on terminals of primary pain afferent neurons, and their relationship with enkephalin-secreting neuron in dorsal horn (mechanism of pre-synaptic inhibition of pain fibres by opioid peptides).

released on stimulation of descending pain inhibitory pathway bind with opioid receptors and decrease the release of substance P from the primary afferent neurons terminating in the dorsal horn (Fig. 10.8-13) and thus inhibit the pain by pre-synaptic inhibition.

Physiological significance. Morphine relieves pain by two mechanisms:

- *At spinal level* by binding to the opioid receptors and thereby decreasing release of substance P and
- *At supraspinal level* by binding to opioid receptors in PAG and thus activating descending inhibitory pathway that produces inhibition of primary afferent transmission in the dorsal horn.
  - *Cannabinoid receptors.* There are mainly two types of cannabinoid receptors: CB_1 present on the central neurons and CB_2 on the peripheral neurons. Some non-neural cells also possess these receptors. An endogenous ligand analogous to these receptors (anandamide) exerts its analgesic effect by binding to these receptors.

(iii) Hypothalamus and frontal cortex also play a role in pain suppression. Neurons descending from the hypothalamus and frontal cortex stimulate both the above described brain stem centres of pain inhibition, i.e. PAG as well as RMN (Fig. 10.8-12).

Conditions under which descending serotonergic and opioid pain inhibitory system are stimulated: The descending pain inhibitory system is stimulated in the following conditions:

(i) When limbic system is stimulated. Limbic system is the seat of emotions. Fibres from the limbic system supply the PAG. This explains why a soldier wounded in the battlefield may feel no pain during the heat of battle.

(ii) Autofeedback. When the spinothalamic tract (STT) is stimulated, the collaterals from STT stimulate the descending inhibitory pathway (Fig. 10.8-13).

2. Descending purinergic inhibitory system, comprising specifically of adenosine, has been recognized. Adenosine exhibits both pre- and post-synaptic actions and produces antinociceptin by indirect interaction with excitatory amino acid release. Role of adenosine on pain suppression system is corroborated by following two observations:

- Significant decrease in circulating blood and cerebrospinal fluid adenosine levels in patients with neuropathic pain and
- Effective attenuation of neuropathic pain following low dose infusion of adenosine.

3. Descending noradrenergic inhibitory system. Fibres of this system originate from the *locus coeruleus* and medullary reticular formation and descend in dorsolateral fasciculus. Environmental factors, such as stress may activate this descending inhibitory mechanism. Norepinephrine depleting agents including reserpine and α2-antagonists, and lesions within the noradrenergic system, all interfere with morphine analgesia.

C. Acetylcholine

Epibatidine, a cholinergic agonist, is a strong non-opioid analgesic agent. Its effect is blocked by cholinergic blocking agents. This suggests that nicotinic cholinergic mechanism is involved in regulation of pain but its exact role is not yet cleared.

OTHER SENSATIONS

This group includes other sensations except somatic sensations (touch, pressure, pain and temperature) like:

- Itch,
- Tickle and
- Synthetic senses.

Itch. It is an irritative skin condition which occurs due to mild stimulation (especially when something moves across the skin).

Characteristic features are:

1. It occurs only in the skin, eyes and certain mucous membranes but not in the deep tissue and viscera.
2. It originates due to stimulation of *itch receptors,* which are naked nerve endings of unmyelinated C fibres. The receptors are stimulated by two ways:
   - By repeated local mechanical stimulation of skin and
   - By certain chemical agents, e.g.
     - Bile salts (raised plasma concentration of bile salts during pregnancy).

  - Histamine (in urticaria severe itching results due to the release of large quantity of histamine from antigen–antibody complex and
  - Kinins.

3. The pathway for itch sensation like pain is carried by fibres into the spinal cord and then conducted by lateral spinothalamic tract.
4. Scratching relieves the itching. The mechanism is same as gate control hypothesis in pain sensation, i.e. scratching stimulates large, fast conducting afferents, which cause pre-synaptic inhibition of fibres in the dorsal horn cells.

Tickle. Tickle is another variable of touch sensation. It is regarded as a pleasurable feeling as compared to itching (which give annoying feeling) and pain (is an unpleasant feeling).

Synthetic sense. The combinations of various cutaneous sensations produce different experiences, which are entirely different from primary sensation. Therefore, the new experience is regarded as synthetic sense.

PATHWAYS IN SOMATOSENSORY SYSTEM (TRANSMISSION OF SENSATIONS)

NEURONS OF SENSORY PATHWAY

Pathways in somatosensory system are formed by a chain of three neurons, which ultimately reach the sensory cortex:

First-order neurons

These are the primary afferent neurons that receive the transduced signals from the sensory receptors and carry them to the spinal cord or brain stem. The cell bodies of the primary afferent neurons are located in the *dorsal root ganglia*. These are T-shaped unipolar neurons with peripheral and central processes. The peripheral processes reach the sensory receptors and form the sensory part of the *spinal nerves* (which are mixed nerves). Central processes constitute the *dorsal nerve root* of the spinal nerve (and also see page 812).

Type of sensory fibres

The fibres of first-order neurons in the spinal nerves and dorsal nerve root comprise Aα, Aβ, Aδ and C type fibres and are often referred to group I, II, III and IV, respectively by the sensory physiologists (Table 10.8-3). Aγ and B fibres are not present in the sensory pathways. For further details about types of fibres, see page 57.

Second-order neurons

The second-order neurons are located in the spinal cord or brain stem. They receive information from one or more primary afferent neurons and transmit it to the thalamus. *In spinal cord*, the neurons involved in sensory functions are present in the dorsal horn of spinal grey matter. Axons of the second-order neurons form the *ascending sensory tracts* described below.

Table 10.8-3 Type of sensory fibres

Sensory group	Fibre type	Origin
Ia	Aα	Annulospinal endings on intrafusal muscle fibres
Ib	Aα	Golgi tendon organs
II	Aβ	Flower-spray endings on intrafusal muscle fibres. Touch and pressure receptors
III	Aδ	Receptors for pain (fast), cold and crude touch receptors
IV	C	Pain (slow) and temperature receptors

Third-order neurons

Third-order neurons of the sensory pathway are located in the specific nuclei of thalamus. From here the encoded sensory information ascends to sensory cortex through the thalamic radiations.

SENSORY NERVES AND DERMATOMES

Sensory nerves

All sensory fibres reach the CNS through their cranial equivalents.

Dorsal nerve roots in spinal cord (Fig. 10.1-4). The different types of sensory fibres forming sensory part of the spinal nerve carry different type of sensations (Table 10.8-3). Each dorsal nerve root is attached to the spinal cord through various rootlets. Each rootlet just before entering the spinal cord divides into medial and lateral divisions.

Medial divisions of each rootlet consists of myelinated group I and II fibres, which include:

- Proprioceptive fibres from muscles and
- Sensory fibres conveying touch, pressure and vibratory sensations.

Lateral division of each rootlet comprises:

- Thinly myelinated group III (Aδ) fibres, which carry fast and discriminative pain and temperature sensations, and
- Unmyelinated group IV (type C) fibres, which carry slow pain and visceral sensations.

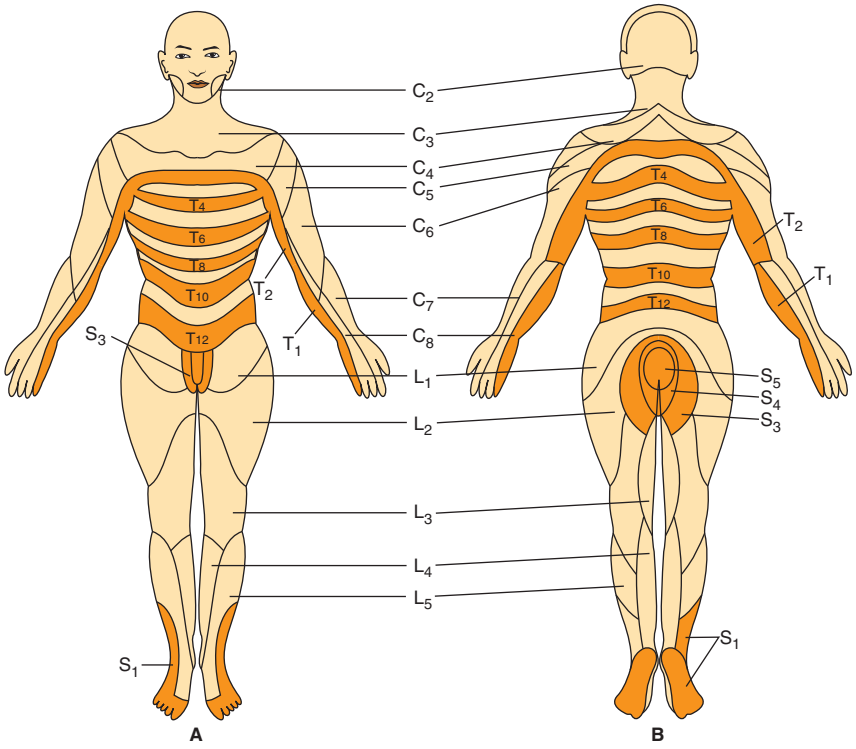


Fig. 10.8-14 Dermatomes as seen from front (A) and back (B).

Dermatomes

Dermatome refers to the area of skin supplied by one dorsal root (spinal cord segment). It is important to note that dermatomes are quite different from the peripheral nerve fields because fibres from one dermatome may be present in the different peripheral nerves.

During embryo stage, body is divided into orderly metameres. In the post-natal life owing to excessive growth of limbs, the metamere arrangement, excepting in the trunk, is no longer present. Therefore, dermatomes are remnants of orderly metameric arrangement, which has survived only in the trunk, where the dermatomes consist of a series of 12 narrow overlapping bands running from the vertebral column to the midventral line (Fig. 10.8-14). The bands slope down as they pass around the body. Apparently, the dermatomes are not arranged in an orderly way, as the L5 dermatome in the leg is at a more distal site than the S4 dermatome, which is near the anus (Fig. 10.8-15). However, this apparent complexity of the dermatomes in man is simplified if the man is visualized as a quadruped animal (like monkey, the ancestor of man) (Fig. 10.8-15). The knowledge of dermatomes is utilized to know the level of spinal cord injury or the level of spinal tumour or other lesions by mapping the area of altered sensation produced.

Plasticity of dermatomes

The dermatomes were originally marked by Sherrington in the later years of 19th century. This is a landmark discovery for which Sherrington may be considered father of neurology. Much later Kirk and Denny Brown reported that under some conditions, dermatomes may alter their area of supply

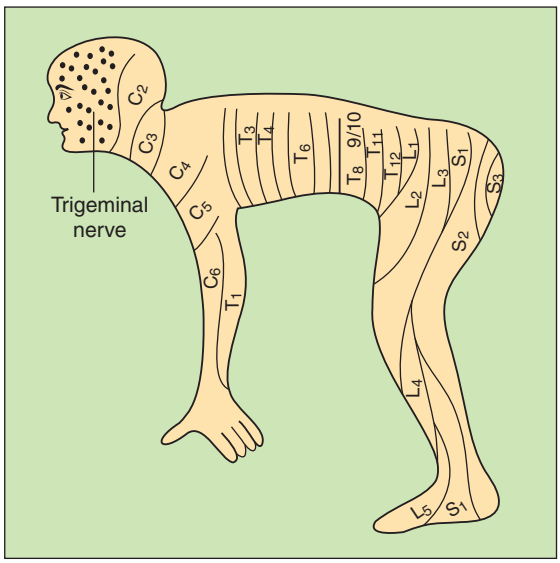


Fig. 10.8-15 Dermatomes in man, visualised as quadruped animal (like ancestors), to clarify the apparent complexity and to memorize the various dermatomes.

to a slight extent. This phenomenon is called plasticity of dermatomes.

ASCENDING SENSORY TRACTS

The major ascending sensory tracts in the spinal cord have been grouped as:

- Dorsal column sensory pathway,
- Anterolateral sensory pathway and
- Dorsolateral column sensory pathway.

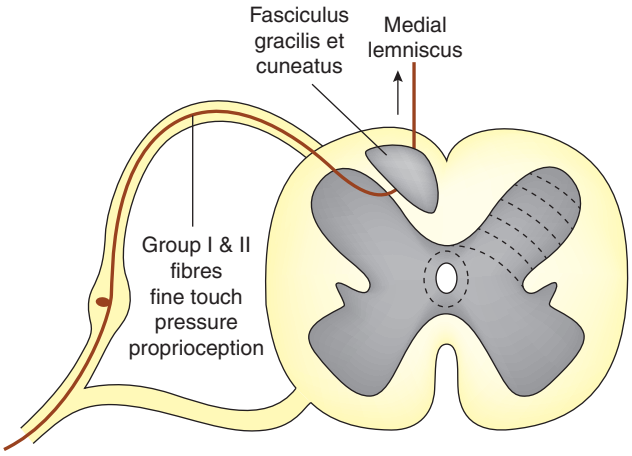


Fig. 10.8-16 Dorsal column sensory pathways. Fibres carrying touch, pressure and proprioceptive sensations are arranged as fasciculus gracilis and fasciculus cuneatus in dorsal column of spinal cord.

The ascending sensory tracts are summarized below (for details see page 697).

Dorsal column sensory pathway

Dorsal column sensory pathway (Fig. 10.1-10) in man is well developed and wholly myelinated.

These carry sensations of fine touch, tactile localization, two-point discrimination, vibration, pressure with intensity discrimination and sense of position and proprioception (Fig. 10.8-16). For detail see page 698.

Functions of dorsal column pathway

(See page 698).

Effects of damage to dorsal column pathway

Sensations of fine touch, tactile discrimination, vibration sense, joint and position sense are carried by dorsal column pathway. Therefore, damage to this pathway will produce following effects:

Sensory ataxia, i.e. imbalance due to the damage to the sensory pathway. In it, it is difficult for the person to detect the position without the help of visual apparatus in erect position. If this person is asked to stand and close the eyes, body cannot maintain balance properly and tends to fall in one direction (Romberg's sign).

Loss of sensations of fine touch, tactile discrimination and vibration sense on the affected side.

Anterolateral pathways

Anterolateral pathways are formed by Aδ (III) and (IV) fibres, which enter the spinal cord as lateral division of the dorsal nerve root. These carry sensations of pain, temperature and crude touch (Fig. 10.8-10). For details see page 699.

Anterolateral pathways terminate in two areas:

- Throughout *reticular nuclei* of brain stem.
- Spinal and medial lemnisci terminate in the *ventrobasal complex* and *intralaminar* nuclei of thalamus. Generally, tactile signals and temperature signals terminate in the ventrobasal complex. Pain signals only partly project to ventrobasal complex of thalamus. Instead most of them enter the reticular nuclei of brain stem and then to intralaminar nuclei of thalamus.

From the ventrobasal complex of thalamus the tactile signals are carried to the somatic sensory area of the cortex along with the fibres of the dorsal column.

Dorsolateral column pathway

Dorsolateral column pathways carry proprioceptive impulses arising from the muscles and joint receptors of the lower part of the body to the cerebellum.

First-order neurons are located in the posterior root ganglia. Their peripheral processes receive impulses from the muscle spindles, Golgi tendon organs and other proprioceptive receptors. Some fibres are related to end organs concerned with the exteroceptive sensations (touch and pressure).

Second-order neurons are located in the junctional area between the ventral and dorsal grey column (laminae V, VI and VII) in the lumbar and sacral segments of spinal cord. Their axons form the:

- Ventral spinocerebellar tract and
- Dorsal spinocerebellar tract.

Note. For details see page 701.

Pathway of sensations from face and oral cavity

The sensations of touch, pain and temperature from the face and oral cavity including teeth, and proprioceptive information from the jaw muscles are carried by the trigeminal nerve.

First-order neurons are located in the *trigeminal ganglion*, which is equivalent to the dorsal nerve root ganglia in the spinal cord. The peripheral processes of these neurons from three divisions of the trigeminal nerve: ophthalmic, maxillary and mandibular which innervate different areas of the facial skin (Fig. 10.8-17). The central processes of these neurons of trigeminal ganglia terminate in different components of trigeminal sensory nucleus as (Fig. 10.8-18): *Principal sensory trigeminal nucleus*, located in the pons, receives fibres carrying *tactile* sensations.

Spinal nucleus is elongated and extends down to the upper spinal cord. It receives fibres carrying *pain* and *temperature* sensations.

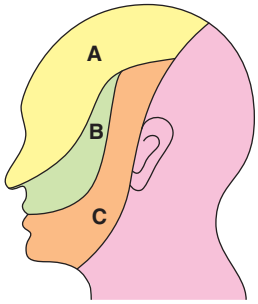


Fig. 10.8-17 The areas of face innervated by three divisions of the trigeminal nerve: A, ophthalmic; B, maxillary and C, mandibular.

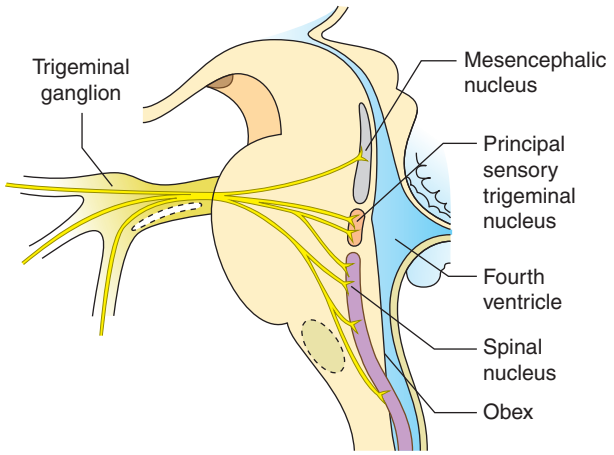


Fig. 10.8-18 Termination of central processes of trigeminal ganglion in three components of sensory nucleus of trigeminal nerve.

Mesencephalic nucleus, which extends from the pons into mid brain, receives fibres carrying *proprioceptive* information.

Second-order neurons are located in the above described three components of the sensory trigeminal nucleus. Axons of these neurons cross to the opposite side and ascend as *trigeminal lemniscus* to the ventroposterior medial (VPM) nucleus of thalamus.

Third-order neurons are located in the VPM nucleus of thalamus. All the sensations reaching this nucleus are carried primarily to sensory area of cerebral cortex by fibres passing through the posterior limb of internal capsule *(superior thalamic radiations*).

ROLE OF THALAMUS IN SOMATOSENSORY SYSTEM

Ventral posterior nucleus of thalamus is concerned with the somatosensory system. It has two divisions:

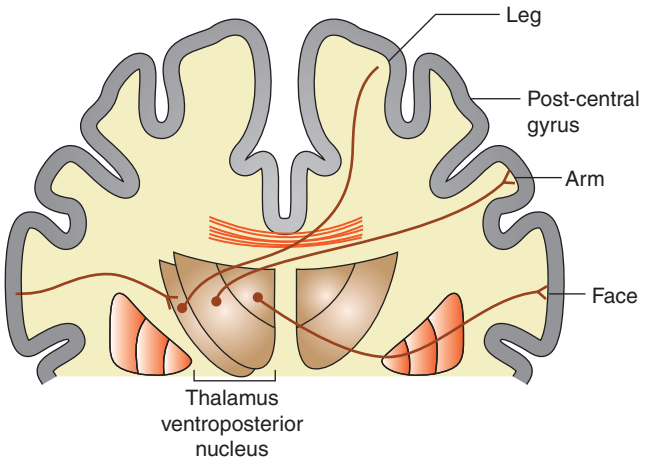


Fig. 10.8-19 Topographic representation of the body in ventral posterior nucleus of thalamus and thalamic projections to sensory cortex.

- Ventral posterior lateral nucleus, and
- Ventral posterior medial nucleus.

For details see page 735.

Topographic representation of the body can be demonstrated in the ventral posterior thalamic nucleus as (Fig. 10.8-19):

- *Face region:* Fibres carrying sensations from the face terminate in the most medial part of the nucleus.
- *Arm region* is represented in the middle part of nucleus and
- *Leg region* in the lateral most part of the nucleus.

Somatosensory functions of thalamus. Thalamus acts as:

- Sensory relay centre,
- Centre for integration of sensory impulses and
- Crude centre for perception of sensations.

In short, pain sensations are perceived in the thalamus itself. All other sensations are transmitted to the cerebral cortex by third order neurons arising from the thalamus (for details see page 738).

SOMATIC SENSORY CORTEX

Somatic sensory cortex is described under following headings:

- Areas,
- Topographical organization of the body in somatic sensory cortex,
- Connections and
- Functions.

For details see page 756.

Chapter 10.9

Somatic Motor System

INTRODUCTION

Effector organ. The motor activity, be it in the form of walking, physical labour, skilled work like typing or even expression of thoughts and feelings through gesture of speech, is a result of highly co-ordinated movements produced by the skeletal muscles. The skeletal muscles thus form the *effector organ* of the somatic motor system.

Lower motor neurons and final common pathway. The somatic motor activity depends ultimately upon the pattern and rate of discharge from the α-motor neurons situated in the ventral (grey) horn of spinal cord and its homologous neurons in the motor nuclei of the cranial nerves present in the brain stem. The *α-motor neurons* are also known as lower motor neurons. The lower motor neurons form the only pathway through which the signals from other parts of the nervous system reach the muscles.

Therefore, the *lower motor neurons* constitute the so-called *final common pathway* of motor system.

Somatic motor activity, in general, comprises voluntary movements, reflex responses, rhythmic motor activities and control of posture and equilibrium.

1. Voluntary movements, like typing, playing musical instruments, writing, drawing, painting, etc. represent the most complex motor activity. Such movements are characterized by being purposeful and initiated at will.

2. Reflex responses are rapid, stereotyped and involuntary activities. They are purposeful but not under voluntary

control. They are produced in response to specific stimuli, e.g. withdrawal reflex in response to a nociceptive stimulus.

3. Rhythmic motor activities like walking, running and chewing combine features of the voluntary as well as reflex responses. These movements are initiated and terminated voluntarily.

4. Control of posture and equilibrium. The maintenance of upright posture is a prerequisite for any goal- and direction-oriented phasic movement. A series of postural reflexes that not only maintain the body in an upright, balanced position but also provide the constant adjustments necessary to maintain a stable postural background for voluntary activity.

Medial versus lateral motor system. The skeletal muscles, which are the effector organs for motor activity, have been organized into two groups: medial proximal and lateral distal. The *medial proximal group* comprises the axial and girdle muscles, whose actions involve the axis and proximal limbs. Their activity determines posture, progression and equilibrium. The *lateral distal group* of muscles are the muscles of digits and distal segments of limbs. They are responsible for skilled voluntary movements. There exists a topographical organization of these groups in central nervous system as medial and lateral motor nervous systems (see page 820).

Control of somatic motor activity. The execution, planning, co-ordination and adjustments of the movements of the body are under the influence of different parts of the nervous system, which together constitute the somatic motor system, which is organized as three-tier system consisting of: highest level of motor control, middle level of motor control and lowest level of motor control.

1. Highest level of motor control involves activities of various areas of cerebral cortex. It is mainly concerned with generation of the idea of voluntary movements (motor plan) and issuing the motor commands for their execution.

2. Middle level of motor control involves activities of various subcortical centres such as basal ganglia, some brain stem nuclei and cerebellum. The middle level of motor control is concerned with developing and perfecting each motor programme and subprogramme for bringing out a motor act. It also supervises the implementation of motor programme.

3. Lowest level of motor control is exerted by cranial nerve nuclei in brain stem and spinal cord. The spinal cord contains the final common pathway through which a movement is executed.

Role of sensory receptors in motor control activity. Feedback signals to central nervous system (CNS) from the proprioceptors in muscles, joints, skin and other sensory receptors are used to adjust the motor commands during the somatic motor activity.

Plan of study of somatic motor control system. In view of the above background, the somatic motor control system is discussed in detail under following headings:

- Components of somatic motor control system,
- Skeletal muscles: The effector organ of somatic motor system,
- Reflexes and
- Regulation of posture and equilibrium.

COMPONENTS OF SOMATIC MOTOR CONTROL SYSTEM

I. HIGHEST LEVEL OF MOTOR CONTROL

CEREBRAL CORTEX

The highest level of motor control is exerted through motor cortex and two major descending pathways emerging from the motor areas (Fig. 10.9-1).

Motor cortex

Areas of motor cortex include (Fig. 10.9-2):

- Primary motor cortex (Brodmann's area 4). It is organized in terms of movements rather than the individual muscles, e.g. stimulation reveals discrete isolated movements on opposite half of the body.
- Premotor cortex. It is located immediately anterior to the lateral portion of the primary cortex. It includes Brodmann's area 6, 8, 44 and 45.
- Supplementary motor cortex is located in the medial surface of frontal lobe rostral to the primary motor area (Fig. 10.4-5).

Note. For details of motor cortex see page 751.

Functional role of motor cortex in control of voluntary movements is summarized:

Supplementary motor cortex is responsible for generating the idea for a movement. There it plans the movements. Lateral cerebellum and basal ganglia are also involved in the planning and programming of movements.

Basal ganglia play their cognitive role through the *caudate loop* (motor association → cortex → caudate nucleus → thalamus → cortex) (see page 728 and Fig. 10.2-13).

Primary motor cortex is responsible for the execution of movement. Programmed patterns of motor neurons are activated in the motor cortex.

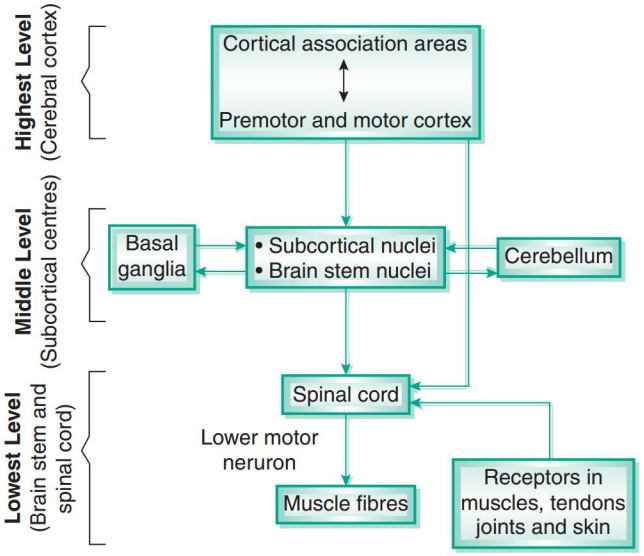


Fig. 10.9-1 Integration of highest, middle and lowest level of somatic motor control system.

Premotor cortex co-ordinates the voluntary activity:

- *Area 6* co-ordinates the proximal and axial muscles during a motor task. It ensures that the skilled movements are accurate and smooth.
- *Area 8,* also called frontal eye field, is involved in the co-ordination of eye movements.
- *Areas 44 and 45*, also called Broca's motor speech areas (especially in dominant hemisphere, e.g. left hemisphere in right-handed person), are engaged in co-ordination of the activity of musculature involved in speech, e.g. activation of vocal cords simultaneously with movements of mouth and tongue during speech.
- *A head rotation area* associated with the frontal eye field is functionally linked to area 8 and serves to enable movements of the head correlated with the eye movements.
- *An area related to the control of fine movements of the hand* is located within the premotor cortex, just anterior to the hand region of area 4. When this area is damaged, the muscles of the hand are not paralysed, but certain hand movements are lost; this is called *motor apraxia.*

Plasticity property of motor cortex. The motor system 'learns by doing' and performance improves with repetition. This involves synaptic plasticity. The motor cortex shows plasticity. This has been confirmed by PET (positron-emission tomography) and functional MRI (fMRI, functional magnetic resonance imaging) in intact experimental animals and humans.

Descending motor pathways from motor cortex

A. Pyramidal tracts include the tracts, which are constituted by the axons that transmit motor signals directly from the cortex to the spinal cord (*corticospinal tracts*) and cranial nerve nucleus (*corticobulbar tracts*). *Corticospinal tracts* include 30% fibres arising from the primary cortex (area 4), 30% from premotor area (area 8) and supplementary cortex and 40% arising from the somatic sensory cortex (areas 3, 1, 2). Corticospinal fibres are divided into two tracts (Fig. 10.1-14, see page 702):

Lateral corticospinal tract is constituted by the 80% of fibres, which cross the midline in the medullary pyramids. These fibres innervate the distal limb muscles, which are responsible for making skilled precision movements (e.g. the muscles that move the fingers and hands and the muscles that produce speech) (*lateral motor system*).

Anterior corticospinal tract is formed by 20% uncrossed fibres, which descend ipsilaterally in the ventral white column of the spinal cord which ultimately cross the midline. This pathway, phylogenetically is oldest that controls the axial (trunk muscles) and proximal limb muscles that are concerned with posture and equilibrium (*medial motor system*).

B. Extrapyramidal tracts. A large number of fibres arising from the motor cortex which do not enter the corticospinal tracts but relay in the various basal ganglia, red nucleus, reticular formation of brain stem and vestibular nuclei. Thus, in contrast to pyramidal tracts, the extrapyramidal tracts constitute multisynaptic pathway affecting the contralateral side of spinal cord (Fig. 10.9-2).

Functions. Extrapyramidal pathways are chiefly concerned with regulation of muscle tone and posture and equilibrium.

Note. For details of corticospinal tract see page 702 and for details of extrapyramidal tracts see page 703.

II. MIDDLE LEVEL OF MOTOR CONTROL

The middle level of motor control is concerned with developing and perfecting each motor programme and subprogramme for bringing out a motor act. It also supervises the implementation of motor programmes.

The middle level of motor control involves activities of basal ganglia, cerebellum and brain stem.

A. BASAL GANGLIA

Physiologically, basal ganglia include corpus striatum (caudate nucleus and lenticular nucleus having two parts, e.g. putamen and globus pallidus), subthalamic nucleus and substantia nigra (see page 726).

Role of basal ganglia in somatic motor activity

1. Control of voluntary motor activity by basal ganglia.

Cognitive control of motor activity is executed by the basal ganglia through the 'caudate loop' (a part of afferent feedback loops).

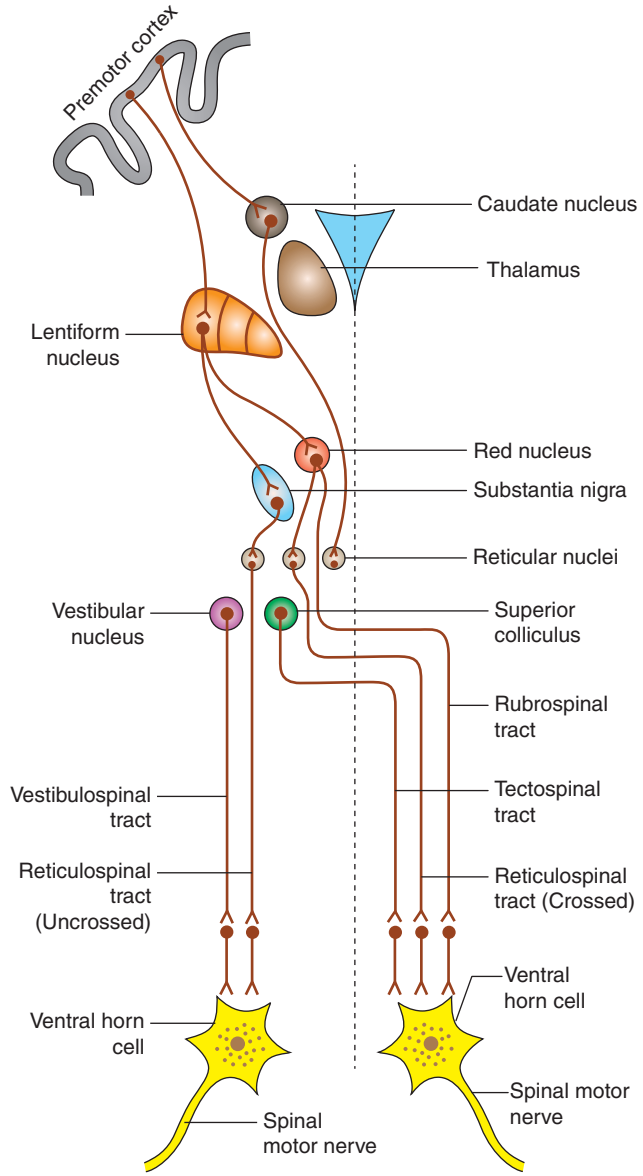


Fig. 10.9-2 The extrapyramidal tracts.

Timing and scaling of the intensity. The basal ganglia play an important role in timing of the movement (i.e. how rapidly the movement should be performed) and scaling of intensity of movement (i.e. how large the movement should be).

Subconscious execution of some movements is done by the basal ganglia during the performance of trained motor activities. Putamen feedback circuit is involved in it. Examples of such movements are:

- Swinging of arms while walking and
- Movements of limbs while swimming.

2. Control of reflex muscular activity. The basal ganglia exert inhibitory effect on the spinal reflexes and regulate activity of muscles which maintain posture.

3. Control of muscle tone. Muscle spindle and γ-motor neurons of spinal cord (which are responsible for maintaining muscle tone) are controlled by the basal ganglia, especially substantia nigra. In lesions of basal ganglia, the muscle tone increases.

Note. For details of basal ganglia, see page 726.

B. CEREBELLUM

Functionally, the cerebellum has been divided into three divisions: vestibulocerebellum, spinocerebellum and corticocerebellum, which play important role in different motor activities.

1. Control of voluntary movements. Corticocerebellum, also called cerebral cerebellum, is intimately associated with control of timing, rate, range (extent), duration, direction and strength of a movement.

The cerebellum controls the voluntary movements by following actions:

(i) Comparator function. The cerebellum receives inputs from the command neurons about the sequential intended plan of movements for the next fraction of second. It also gets feedback (afferents) from the proprioceptive endings of muscles, tendons and joints about what actual movements result. All these information are integrated and the corrective signals are sent to the motor cortex. This happens through the closed feedback loop (Fig. 10.2-9) (see page 724).

(ii) Damping action. Corticocerebellum sends impulses to the cerebral cortex to discharge appropriate signals to the muscles so that any extra or exaggeration of muscular activity does not occur and thus overshooting is prevented. This action of corticocerebellum is called damping action.

(iii) Timing and programming of the skilled movements is done by corticocerebellum through open feedback loop (Fig. 10.2-8), which modulates the motor command of pyramidal tracts through two-way communication (see page 724).

(iv) Servomechanism. Cerebellum lets the cerebral cortex to discharge the signals which are already programmed and stored at the sensory motor cortex and does not influence much. However, if there is any disturbance or interference, the corticocerebellum immediately influences the cerebral cortex and corrects the movements. This action of corticocerebellum is known as servomechanism (see page 724).

2. Control of body posture and equilibrium is done by vestibulocerebellum (see page 722).

3. Control of muscle tone and stretch reflex is the function of spinocerebellum (see page 722).

Note. For details about cerebellum, see page 713.

C. BRAIN STEM

Reticular formation and vestibular nuclei are important components of the motor control system present in the brain stem.

Reticular formation. The *motor control centres* within the reticular formation are a relay station for all descending motor commands, except those requiring the greatest precision, which are transferred directly from the cortex to spinal cord.

- The motor control centres receive and modify the motor commands to the proximal and axial muscles of the body responsible for maintaining normal posture tone.
- These neurons are prevented from firing too rapidly by inhibitory input derived from the cerebral and cerebellar components of the motor control system.

Vestibular nuclei. Vestibular nuclei, located within the brain stem and cerebellum, receive information from *vestibular receptors* via vestibular nerve fibres (8th cranial nerve). Vestibular system reflexes:

- Maintain tone in antigravity muscles,
- Co-ordinate the adjustments made by the limbs and trunk to maintain balance and
- Adjust the position of the eyes to maintain visual fixation when the position of head changes.

III. LOWEST LEVEL OF MOTOR CONTROL

The lowest level of motor control is exerted by motor nuclei of cranial nerves and spinal cord. The spinal cord contains the *final common pathway* through which a movement is executed. By selecting the proper motor neurons for a particular task and by reflexly adjusting the amount of motor neuron activity, the spinal cord contributes to the proper performance of a motor task. The spinal cord activity ranges from a simple withdrawal reflex to co-ordinated movement of all four extremities.

SPINAL CORD

MOTOR NEURONS

Motor neurons of spinal cord present in the ventral horn are:
1. α-motor neurons. These are the largest neurons. The axons of a neurons innervate the extrafusal fibres of the skeletal muscles. These are responsible for the contraction of muscles in the upper limbs, trunk and lower part of the body.

2. γ-motor neurons. These are much smaller and give rise to smaller axons. These neurons innervate the intrafusal fibres of the muscle spindles and are responsible for maintenance of muscle tone.

3. Interneurons. These are highly excitable, and may have a spontaneous firing rates as high as 1500 per second. The interneurons actually receive the bulk of synaptic input that reaches the spinal cord, either as incoming sensory information or signals descending from the higher centres in the brain.

4. Renshaw cells are particular variety of interneurons that receive input from the collateral branches of the axons of α-motor neurons. Their axons carry the impulses back to the cell bodies of the same α-motor neurons. These are inhibitory neurons, which play an important role in synaptic inhibition at the spinal cord.

Arrangement of motor neurons in ventral horn

The motor neurons responsible for the contraction of skeletal muscles are arranged topographically in the ventral grey horn of the spinal cord in three mediolateral column groups:

1. Medial group. It extends along most of the length of spinal cord. The neurons situated in the medial part of ventral grey horn innervate the muscles near the midline of the body called axial muscles and the muscles in the proximal portions of limbs, which are involved in the adjustment of posture and gross movements.

2. Lateral group. This group of neurons is confined to the cervical and lumbosacral enlargements and supplies the muscles in distal portions of the limbs called distal muscles which are involved in the well co-ordinated skilled voluntary movements.

3. Central group. This group of neurons is represented by the phrenic and accessory nuclei (in the cervical region) and by the lumbosacral nucleus (in the lumbosacral region).

Motor functions

Motor functions served by the spinal cord are:

- Control of movement of muscles and joints,
- Control of tone and power of muscles,
- Control of deep (tendon) reflexes and
- Control of superficial reflexes.

SKELETAL MUSCLES: THE EFFECTOR ORGAN OF SOMATIC MOTOR SYSTEM

The skeletal muscles form the effector organ of the somatic motor system. The physiology of skeletal muscle has been discussed in Chapter 2.3. However, certain aspects which need elaboration of skeletal muscle as effector organ

and are relevant to complete the study of somatic motor system are:

- Motor unit,
- Muscle sensors (proprioceptors) and
- Muscle tone.

MOTOR UNIT

The motor unit is the functional module used by the motor control system to carry out a movement. The movement produced by a skeletal muscle basically depends upon the pattern and ratio of discharge of motor neurons supplying the muscle. A *motor unit consists* of single motor neuron and the muscle fibres that it innervates.

MUSCLE SENSORS

Muscle sensors refer to the proprioceptors present in the muscles, tendons of muscles, joints, ligaments and fasciae. Proprioceptors are the receptors which give information about change in position of different parts of the body in space, especially joints or tension of muscles at any given moment. The muscle sensors are:

- Muscle spindle,
- Golgi tendon organ,
- Pacinian corpuscle and
- Free nerve endings.

In addition to the above proprioceptors, the labyrinth also contains proprioceptors (see page 843).

1. Muscle spindle

Muscle spindles are *stretch receptors* present in the skeletal muscles. These are meant for proprioceptive mechanism and so are a type of proprioceptors. Each skeletal muscle contains muscle spindles of variable number depending upon the task performed. Muscles involved in precision movements contain many more spindles than the muscles used to maintain posture. For example, hand muscles have approximately 80 spindles, which is 20% of the number of spindles contained in back muscles weighing 100 times as much.

Structure (Fig. 10.9-3)

Each muscle spindle consists of 3–10 small muscle fibres (called intrafusal muscle fibres), encapsulated in a thin connective tissue capsule containing fluid. The muscle spindles are present in between and parallel to the extrafusal fibres (large force-generating muscle fibres). Either end of the muscle spindle is attached to the endomysium of the extrafusal muscle fibres.

Intrafusal muscle fibres consist of a central non-contractile portion, which does not contain actin and myosin filaments and is thus devoid of striations. Portions on either side of the central part are contractile (as they contain actin and myosin filaments) and are called striated poles. The central part of each intrafusal fibre is sensory portion. Intrafusal fibres are of two types:

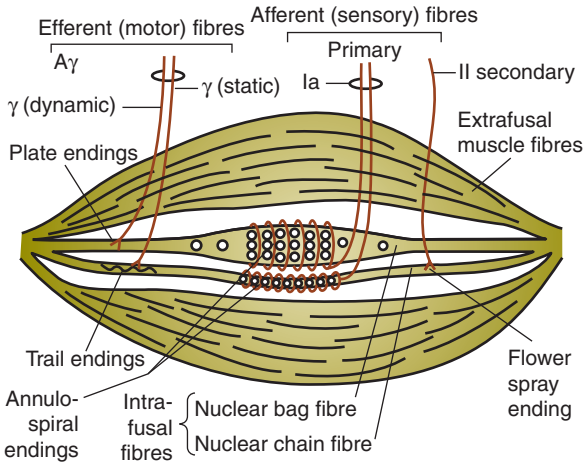


Fig. 10.9-3 Structure of a muscle spindle.

- *Nuclear bag fibres:* Each spindle contains about 2–5 nuclear bag fibres, which are about 30 µm in diameter and 7 mm in length. In these fibres, many nuclei are congregated into an expanded bag in the central portion, hence the name.
- *Nuclear chain fibres* are 15 µm in diameter and 4 mm in length. In these fibres nuclei are arranged in a single file in the central part in the form of a chain, hence the name. Approximately, 6–10 nuclear chain fibres exist in each typical spindle.

Nerve supply of the muscle spindle

The muscle spindle is innervated by both sensory and motor nerve fibres. It is the only receptor in the body which has got motor nerve supply also.

(i) Sensory nerve supply. Central non-contractile portion of each intrafusal fibres is the receptor portion. Sensory fibres supply this area. There are two types of sensory fibres:

Group Ia fibres, also known as primary sensory endings, supplying central receptor portions of both nuclear bag as well as nuclear chain fibres. Since these fibres spirally wind round the intrafusal fibres, these are also called *annulospiral endings.* They have diameter of about 17 µm and carry impulses at the rate of 70–120 m/s.

The primary endings supplying both the nuclear bag as well as the nuclear chain intrafusal fibres are stimulated when the muscle spindle is stretched. But the pattern of response is different:

- *Dynamic response* is shown by nerve endings supplying the nuclear bag fibres and

- *Static response* is shown by the nerve endings supplying the nuclear chain fibres (see below).

Type II fibres, also known as secondary sensory endings, innervate the receptor portion of mainly nuclear chain fibres on one side of the primary endings. They are also known as *flower spray endings.* They have a diameter of about 8 μm. These nerve endings respond mainly to sustained stretch, therefore measure the *muscle length*.

(ii) Motor supply. The efferent fibres to the muscle spindle are called *γ-fibres* because their axons belong to the Aγ group of fibres. There are two types of γ-fibres:

- *Dynamic γ-fibres* primarily innervate the striated poles of nuclear bag fibres, where they end as motor end plate, hence also called *plate endings.* These fibres increase the sensitivity of the Ia afferent fibres to stretch.
- *Static γ-fibres* primarily innervate the striated poles of nuclear chain fibres where they end as a network of branches called *trail endings*. They increase the tonic activity in the Ia afferent fibres at any given muscle length.

Functions of muscle spindle

(i) Role in stretch reflex. Muscle spindle forms the receptor organ of stretch reflex and thus plays *a key role in stretch* reflex (for details see page 826).

(ii) Role in maintaining muscle tone. *Muscle spindle plays an important role* in maintaining the muscle tone by controlling the discharge from the γ-motor neurons (for details see page 828).

(iii) Role in maintaining skeletal muscle at a certain physiological length. Most important function of muscle spindle is to act as a comparator of the extrafusal fibre length. The muscle spindles, through the stretch reflex, act as a feedback device to maintain the skeletal muscle at a certain physiologically useful length. This action of muscle spindles (particularly in the antigravity muscles) is of fundamental importance in the maintenance of standing posture (see page 841).

(iv) Role as proprioceptor. Muscle spindle plays the role of proprioceptor in:

- *Unconscious proprioceptive sensations* and
- *Conscious kinaesthetic sensations.*

2. Golgi tendon organ

The Golgi tendon organs are high threshold stretch receptors present in the tendons. They are supplied by group Ib afferent fibres and detect muscle tension (for details see page 829).

3. Pacinian corpuscle

Pacinian corpuscles are pressure receptors situated in fasciae throughout the muscles, tendons, joints and periosteum. They are supplied by *group II afferent fibres* and detect *vibration.*

4. Free nerve endings

Free nerve endings are basically pain receptors situated in the muscles, tendons, fasciae and joints. They are supplied by *group III and IV afferent fibres* and detect noxious stimuli.

MUSCLE TONE

Definition. Muscle tone is defined as a resistance offered to active or passive stretch. In other words, muscle tone refers to a sustained partial state of contraction of the muscle under resting condition, i.e. a state of partial tetanus. The muscle tone is present in all the muscles, but is well pronounced in the extensor muscles, i.e. antigravity muscles.

Basis of muscle tone. The muscle tone is purely a function of myotactic (stretch reflex), occurring due to low frequency and asynchronous discharge of γ motor neurons. The discharge is out of phase with each other, which ultimately merges to produce smooth muscle contraction.

Anomalies of muscle tone. Anomalies of muscle tone are hypotonia and hypertonia.

1. Hypotonia refers to a decrease in the muscle tone. The hypotonic, or also called flaccid muscle, offers little or no resistance to stretching. The muscles are generally hypotonic when the rate of γ efferent discharge is low, i.e. when stretch reflex becomes hypoactive.

2. Hypertonia refers to an increase in the muscle tone. The hypertonic or spastic muscle offers high resistance to stretch. The muscles are generally hypertonic when the rate of γ-efferent discharge is high, i.e. when stretch reflex becomes hyperactive.

Types of hypertonia: Hypertonia is of two types:

- *Spasticity* refers to the hypertonia, which is confined to only one group of muscles. For example, lesions of internal capsule and upper motor neuron lesions produce spasticity.
- *Rigidity* refers to the hypertonia, which involves both groups of muscles, i.e. extensor as well as flexors equally. For example, lesions of basal ganglia produce rigidity.

REFLEX ACTIVITY

GENERAL CONSIDERATIONS

A reflex is an involuntary response to a peripheral nervous stimulation. In other words, it is a mechanism by which sensory impulse is automatically converted into a motor effect through the involvement of CNS. It is a type of protective mechanism which tries to protect the body from irreparable damage. For example, when the hand is placed inadvertently on a hot object, it is immediately withdrawn reflexly. Thus, the hand is protected from getting burnt.

ANATOMICAL ASPECTS

REFLEX ARC

The pathway for a reflex activity is called reflex arc. It consists of (Fig. 10.9-4):

- Afferent limb,
- Centre and
- Efferent limb.

1. Afferent limb of each reflex arc consists of a receptor and an afferent or sensory nerve.

Afferent neuron carries sensory input from the receptor to the centre. The afferent neurons enter the CNS via the dorsal roots or cranial nerves and have their cell bodies in the dorsal root ganglia or in the homologous ganglia on the cranial nerves.

2. Centre. This is the part of CNS (spinal cord or brain) where afferent limb ends and either synapses directly with the efferent motor neuron or establishes connection with the efferent neuron via interneurons (internuncial neurons). Thus, the number of synapses (connection between afferent and efferent neurons) may vary from one (in the simplest form of reflex) to many hundred.

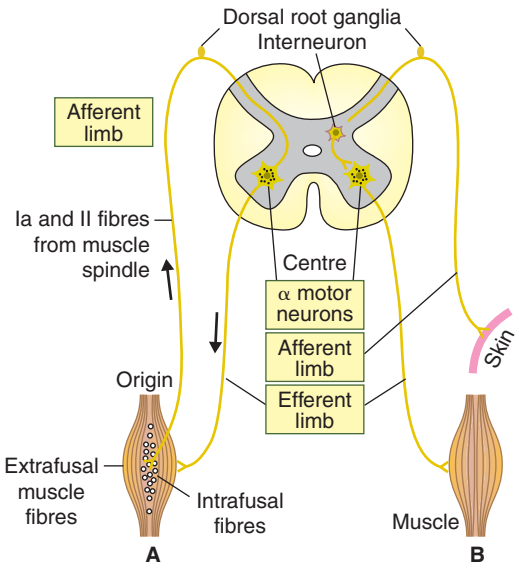


Fig. 10.9-4 Components of reflex arc in a monosynaptic (A) and disynaptic reflex (B).

3. Efferent limb of a reflex arc consists of an efferent or motor nerve and an effector organ.

- Efferent nerve transmits motor impulses from the centre to the effector organ. Since the connection between afferent and efferent neurons is usually present in the CNS; therefore, activity in the reflex arc is modified by the multiple inputs converging on the efferent neuron.
- Effector organ may be in the form of a muscle or a gland, which shows the response to the stimulus.

CLASSIFICATION OF REFLEXES

Reflexes can be classified in different ways:

I. Depending upon the number of synapses (Fig. 10.9-4)

1. Monosynaptic reflexes are those which contain only one synapse, e.g. stretch reflexes (biceps, triceps or knee jerk).

2. Disynaptic reflexes have two synapses, i.e. one interneuron is placed between afferent and efferent neurons of the reflex arc, e.g. inverse stretch reflex.

3. Polysynaptic reflexes are characterized by more than one interneuron placed between afferent and efferent neurons of the reflex arc, e.g. withdrawal reflex, cross flexor reflex and cross extensor reflex.

II. Anatomical classification

Depending upon the location of reflex arc centre, the reflexes can be classified as:

1. *Cortical reflexes*
2. *Cerebellar reflexes* have the centre of reflex arc in cerebellum
3. *Mid brain reflexes*
4. *Bulbar or medullary reflexes* and
5. *Spinal reflexes*

III. Physiological classification

1. Flexor reflexes. These reflexes occur in response to nociceptive (pain) stimuli and are characterized by flexion of the joints, e.g. thorn prick to the sole is immediately followed by reflex flexion of the knee and hip joints. These reflexes are also called *withdrawal reflexes.*

2. Extensor reflexes. Stretch reflexes are extensor reflexes. These are the basis of muscle tone and posture of the body. These are also called *antigravity reflexes.*

IV. Inborn versus acquired reflexes

1. Inborn or unconditional reflexes are present since birth and do not require any previous learning or training, e.g. reflex salivation when any object is kept in mouth.

2. Acquired or conditional reflexes develop after birth. Such reflexes are acquired after conditioning, i.e. after previous learning or training, e.g. reflex salivation by the sight, smell, thought or hearing of a known edible substance.

V. Clinical classification

Clinically reflexes are classified into:

- Superficial
- Deep,
- Visceral and
- Pathological reflexes.

ANIMAL PREPARATIONS FOR STUDY OF REFLEXES

The reflexes can be studied in:

- Spinal preparation and
- Decerebrate preparation.

1. Spinal preparation. In it the spinal cord is transected at the cervical region and respiration is maintained by the respiratory pump. When spinal cord is transected in the thoracic region, artificial respiration is not required since diaphragmatic breathing continues. Such a preparation allows study of properties of spinal reflex.

2. Decerebrate preparation. In decerebrate preparation, the transection is taken in the brain stem between superior and inferior colliculi.

PROPERTIES OF REFLEXES

1. Adequate stimulus. Reflex response is obtained only when a precise stimulus for a given reflex activity is applied. The precise stimulus which involves a reflex response is called *adequate stimulus* for that particular reflex. For example, scratch reflex in a dog is initiated only by multiple linear touch stimuli. If multiple stimuli are widely separated, reflex is not initiated.

2. Delay. All reflex activity is associated with delay. Delay refers to the time interval between the application of stimulus and starting of the response. It is attributed to a synaptic delay and to time required for passage of impulse along the nerves. Therefore, delay is minimum in a monosynaptic reflex.

3. One-way conduction. During any reflex activity, the impulses are transmitted in only one direction through the reflex arc as per the Bell–Magendie law. The impulses pass from the receptors to the centre and then from the centre to the effector organ.

4. Summation of stimuli, both temporal and spatial, play an important role in the facilitation of responses during the reflex activity (see page 780).

5. Irradiation. When a sensory stimulus is too strong, impulse spreads to many neighbouring neurons in the centre and produces wider response. It is due to the transmission of impulse through a large number of collaterals of afferents and their interneurons.

6. Final common pathway. Efferent pathway of the reflex arc is formed by α-motor neurons that supply the extrafusal muscle fibres. All neuronal influences (excitatory and inhibitory) affecting muscular contraction ultimately funnel through the motor neurons; therefore, they are called common final pathway. Numerous inputs converge on them and determine the activity in the final common path (Fig. 10.9-5). If an α-motor neuron is stimulated, skeletal muscle fibres contract; if the α-motor neuron is not stimulated, the skeletal muscle fibres relax. Thus, the α-motor neuron forming the final common pathway serves both as an integrating centre and an efferent pathway.

7. Facilitation. When a reflex is elicited repeatedly at proper intervals the response becomes progressively higher for first few occasions, i.e. each subsequent stimulus exerts a better effect than the previous one. This is due to the facilitation occurring at the synapse.

8. Inhibition. During a reflex activity, impulses through sensory fibres from protagonist muscles inhibit the action of antagonist muscles. For example, when flexor muscles of a joint are stimulated, the extensor muscles are inhibited. The inhibitory activity exerted by the interneurons is responsible for such a reciprocal inhibitory effect.

9. After discharge. When a reflex action is elicited continuously for some time, and then the stimulation is stopped, the reflex response (contraction) may continue for some time even after cessation of the stimulus. This is called after discharge. This is mainly because of the internuncial neurons, which continue to transmit impulses to the centre even after cessation of stimulus.

10. Fatigue or habituation. When a particular reflex is elicited repeatedly at frequent intervals, the response is reduced progressively and then disappears all together. This is called fatigue or habituation. The first site of fatigue is synapse, then the motor endings and lastly the muscle.

11. Rebound phenomenon. The reflex activity can be inhibited for some time by some method. However, once the inhibitory effect is over, the reflex activity reappears and becomes more powerful. This is called rebound phenomenon. Its cause is still not known.

12. Fractionation. The force of a muscle contraction is much higher when it is stimulated directly through motor

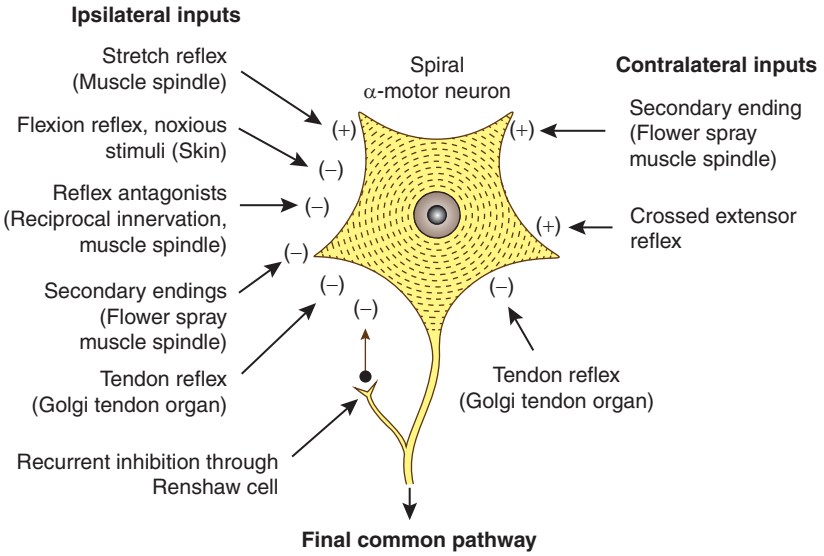


Fig. 10.9-5 The inputs converging on the body of alpha (α) motor neuron (final common pathway).

nerve as compared to when it is stimulated reflexly through a sensory nerve. This is due to phenomenon of occlusion of the motor neurons when sensory nerve is stimulated. Because of occlusion, number of motor neurons stimulated is lesser.

13. Sensitization. When an injurious stimulus is repeatedly applied, there occurs intensification of response. This is known as sensitization. Sensitization, in fact, is the presynaptic facilitation of an impulse.

SPINAL CORD REFLEXES

According to the receptors from which they originate the spinal cord reflexes can be categorized into muscle reflexes and cutaneous reflexes:

Muscle reflexes. Two important reflexes, which originate in the muscles, are:

- Stretch reflex and
- Lengthening reaction or Golgi tendon reflex.

Cutaneous reflexes. The most important of the cutaneous reflexes is:

- Withdrawal (flexor, pain) reflex.

1. STRETCH REFLEX

Stretch reflex, also known as *myotactic*, refers to the reflex contraction of a muscle that is stretched.

- *Type.* It is the best known monosynaptic reflex in the body.
- *Stimulus* that evokes the reflex response is 'stretch' to the muscle.
- *Reaction time,* i.e. the time between the application of the stimulus and the response for a stretch reflex is 19–24 ms. Stretch reflex is the quickest of all the reflexes.
- *Central delay,* i.e. the time taken for the reflex activity to traverse the spinal cord in a stretch reflex (being monosynaptic) is only 0.6–0.9 ms.
- *Stretch reflex is well developed* in antigravity muscles, such as extensor group of muscles of legs and flexor groups of muscles of arm.
- *Examples of stretch reflexes* are knee jerk, ankle jerk, biceps jerk and triceps jerk (see page 832).

Reflex arc of stretch reflex (Fig. 10.9-4)

1. Afferent limb consists of receptor and afferent nerve.

- *Receptor* for a stretch reflex is muscle spindle. As a sensory receptor the muscle spindle detects the degree and rate of muscle stretch. For detailed structure of muscle spindle see page 821.
- *Afferent nerve.* As described earlier (nerve supply of muscle spindle), two types of nerve fibres, group Ia fibre and group II fibres supply the muscle spindle (see page 821). The afferent nerve fibres emerging from the muscle spindle travel along the spinal nerve and enter the spinal cord through the dorsal root and send branches to every α-motor neuron that goes to the muscle from which the Ia originated.

2. Centre. Centre for a stretch reflex is the ventral grey horn area where the afferent nerve ends and synapses directly with the α-motor neuron. Thus, α-motor neuron is the final common pathway, serving as both integrating centre and efferent pathway.

3. Efferent limb consists of the efferent nerve and an effector organ.

- *Efferent nerve.* The axons of α-motor neurons (with which the afferent fibres synapse directly) form the efferent nerve fibres which leave the spinal cord through the ventral root and supply the skeletal muscle fibres.
- *Effector organs.* Both extensor and flexor muscles exhibit stretch reflexes and thus form the effector organs.

Reciprocal innervation in a stretch reflex

The stretch reflex is characterized by the reciprocal innervation, i.e. excitation of one group of muscles is associated with inhibition of the antagonistic group of muscles on the same side, allowing the agonistic muscles to contract without interference. Reciprocal innervation is one of the important features of both flexor and extensor reflexes.

Pathway of reciprocal innervation is biphasic. A collateral from each Ia fibre passes in the spinal cord to an inhibitory interneuron (Golgi bottle neuron) that synapses directly on one of the motor neurons supplying the antagonist muscles. Thus, this is an example of post-synaptic inhibition (Fig. 10.7-15, see page 786).

Significance of reciprocal innervation. Reciprocal innervation is very important in the spinal reflexes, which are involved in locomotion. It helps in the forward movement of one limb while causing the backward movement of other limb.

Dynamic stretch reflex versus static stretch reflex

Dynamic stretch reflex. When the muscle is stretched suddenly, the length of spindle receptor also increases suddenly (as the intrafusal fibres forming muscle spindle are attached in parallel with extrafusal fibres of the muscle). A sudden increase in length of spindle receptor stimulates the primary nerve ending powerfully. The primary nerve endings supplying the nuclear bag fibres show a dynamic response, i.e. they discharge most rapidly while the muscle is being stretched (Figs 10.9-6 and 10.9-7B) and transmit strong signals to the spinal cord, and it causes instantaneous, very strong reflex contraction of the same muscle from which the signals are originated. This is called dynamic stretch reflex, the function of which is to oppose a sudden change in length, e.g. knee jerk, ankle jerk. Dynamic stretch reflex is over within fraction of a second because primary nerve endings supplying nuclear bag fibres are stimulated actively only when there is a rapid change of length, i.e. they are stimulated only when the length is actually increasing. As soon as the length stops increasing, the rate of impulse discharge through those endings returns back to normal. Further, when the muscle contracts reflexly, the spindle receptors shorten and the discharge through primary ending even decreases momentarily (Fig. 10.9-7C).

Static stretch reflex. When the muscle is stretched slowly and kept stretched, signals are continuously sent through

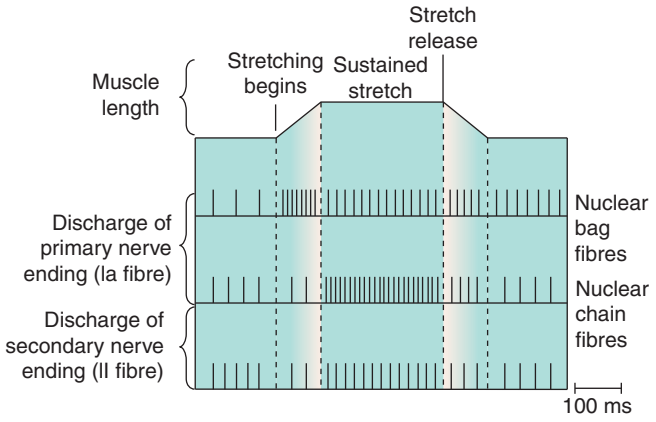


Fig. 10.9-6 Response of primary (Ia) and secondary (II) nerve endings of muscle spindles to muscle stretch.

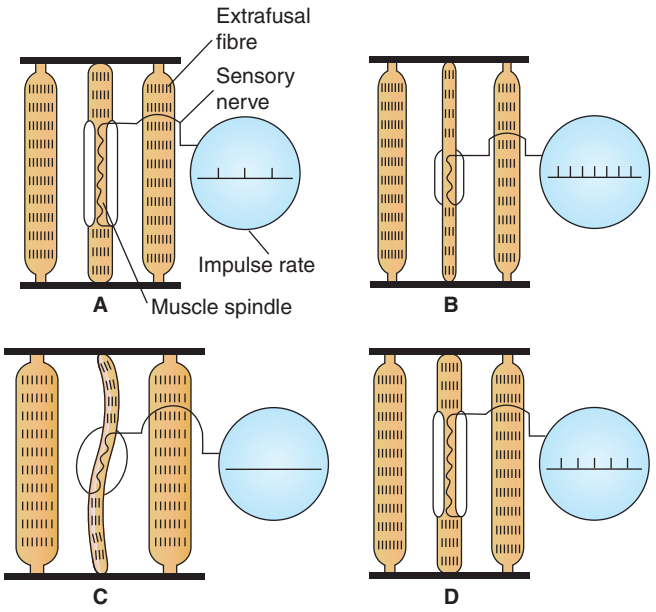


Fig. 10.9-7 Firing rate of primary nerve endings (Ia) under different conditions: A, muscle at rest; B, muscle stretched; C, muscle contracted and D, muscle contracted with increased gamma (γ) efferent discharge.

primary and secondary nerve endings supplying the nuclear chain fibres only and cause reflex contraction of the muscle.

This is because the nerves from the primary ending on the nuclear chain fibres show a static response, i.e. they discharge at an increased rate throughout the period when a muscle is stretched (Fig. 10.9-6). This is called static stretch reflex. This static reflex therefore causes muscle contraction as long as the muscle is maintained of excessive length. The static stretch reflex plays an important role in control of posture, e.g. when the person is standing, gravity causes continuous stretch on the antigravity muscle making them to remain in a contracted state as long as gravity is causing the stretch.

IMPORTANT NOTE

From the above, it is clear that primary nerve ending responds to both changes in length (static stretch reflex) as well as changes in the rate of stretch (dynamic stretch reflex). The response of primary endings to the phasic as well as static events in the muscle is important because the prompt, marked phasic response helps to dampen oscillations caused by conduction delays in the feedback loop regulating the muscle length.

Role of γ-motor neurons

1. Role of γ-efferent discharge in adjusting the spindle sensitivity by preventing unloading. As discussed above, the firing rate of primary nerve endings (Ia fibres) increases when the muscle is stretched (Fig. 10.9-7B) and causes reflex contraction of the muscle by increased α-motor neuron activity. Contraction of the extrafusal muscle fibres makes the muscle spindle slack and decrease the firing rate of Ia fibres (Fig. 10.9-7C). The decreased rate of Ia afferent discharge that occurs during muscle contraction is called *unloading of muscle spindle* and is functionally disadvantageous because the CNS stops receiving information about the rate and extent of muscle shortening. However, by the activity of γ-motor neurons this unloading is prevented (Fig. 10.9-7D). The γ-motor neurons cause the striated poles of intrafusal fibres of muscle spindle to shorten along with shortening of extrafusal fibres during muscle contraction. As a result of contraction of the striated polar regions of intrafusal fibres, the central receptor region of the intrafusal fibres remains stretched during muscle contraction and unloading does not occur. In this way, the γ-motor neuron activity adjusts the sensitivity of the muscle spindle so that it will respond appropriately during muscle contraction as well. Further, the γ-motor neurons control both dynamic as well as static activity of the muscle spindle as described.

- *Dynamic γ-motor neurons* primarily innervate the striated poles of nuclear bag fibres (Fig. 10.9-3). Thus, when they are fired, only nuclear bag fibres shorten. Because the nuclear bag fibres are responsible for the phasic (i.e. velocity sensitive) portion of Ia afferent response to stretch, stimulation of the dynamic γ-fibres increases phasic activity without affecting static activity.
- *Static γ-fibres* primarily innervate the striated poles of nuclear chain fibres (Fig. 10.9-3). When they are fired, only the nuclear chain fibres shorten. Because the nuclear chain fibres are responsible for the static (i.e. length sensitive) component of Ia afferent response to stretch, stimulation of the static γ-fibres increases static activity without affecting phasic activity.

Note. The above described γ-motor neuron-mediated change in length of intrafusal fibres forms the so-called *length servomechanism,* which is a system of negative feedback device that operates to maintain muscle length during body movements and thus helps in regulation of posture (see page 828).

2. Role of co-activation of α- and γ-motor neurons. During a normal voluntary movement (e.g. lifting a weight), the active shortening of the extrafusal fibres would relieve tension on the muscle spindles (i.e. unload the spindle) and hence tend to decrease Ia discharge. However, during voluntary contraction, the motor control system causes α–γ co-activation, preventing the unloading of muscle spindle that would occur during muscle contraction. Thus increased γ-discharge along with the increased α-discharge during voluntary movement maintains constant Ia discharge. The constant level of Ia input to the CNS during a voluntary movement indicates that motor command is being carried out.

Note. The α–γ co-activation also forms the so-called *follow-up servomechanism* during voluntary movements.

3. Role of γ-loop. It is theoretically possible that the CNS is capable of initiating movements directly by stimulating only γ-motor neurons, using a pathway called the γ-loop (Fig. 10.9-8). The loop begins with γ-motor neuron, which discharges to cause intrafusal muscle fibre contraction. This leads to an increase in Ia afferent fibre activity, which in turn causes increased γ-motor neuron discharge via a monosynaptic reflex causing muscle contraction.

Although the γ-loop can elicit movement on its own, it normally does not do so. However, because of co-activation, the γ-loop is activated during all movements and thus contributes to the excitability and firing rate of the α-motor neurons.

Higher control of stretch reflex

Though the stretch reflex is a spinal reflex, the activity in the reflex arc can be modified (inhibited or facilitated) by higher centres through their influence on the nerve fibres involved in the stretch reflex.

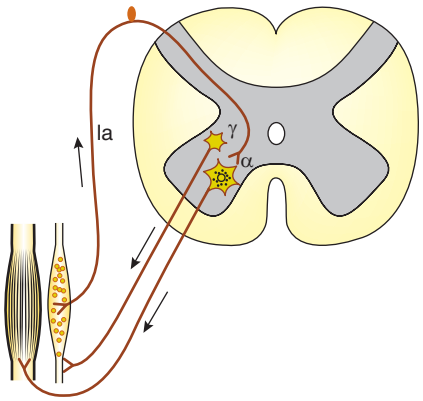


Fig. 10.9-8 The γ-loop system of initiating muscle contraction directly through stimulation of γ-motor neurons.

Some of the important brain areas that facilitate or inhibit the stretch reflex are (Fig. 10.9-9):

- *Facilitatory reticular formation* is a large area in the brain stem which discharges spontaneously in response to afferent input. This increases discharge of γ-motor neurons and stretch reflex becomes hyperactive.
- *Inhibitory reticular formation* is a small area which does not discharge spontaneously. It acts by inhibiting γ-efferent neuron discharge, thereby decreasing the spindle sensitivity.
- *Cerebral motor cortex* and cerebellum reflexly inhibits the stretch reflex by stimulating the inhibitory reticular formation.

Other factors which influence γ-efferent discharge:

- *Anxiety* causes an increased discharge, a fact that probably explains the hyperactive tendon reflexes sometimes seen in the anxious patients.
- *Stimulation of skin,* especially by noxious agents, increases γ-efferent discharge to ipsilateral flexor muscle spindles and decreases that to extensors and produces the opposite pattern in opposite limb. This fact is sometimes used as a reinforcement to elicit deep tendon reflexes (such as knee jerk), which are not being elicited otherwise. For it, the individual is asked to pull the hands apart when the flexed fingers are hooked together; this facilitates the knee jerk (*Jendrassik's* manoeuvre). It is contributed to increased γ-efferent discharge initiated by afferent impulses from the hands.

Functions of stretch reflex

(i) Role in maintaining muscle tone. The muscle tone is the function of stretch reflex, which is under the influence of discharge from the γ-motor neuron. In the brain stem, there are two areas—facilitatory area is the pons and inhibitory area is the lower part of medulla. These areas send the impulses to γ-motor neurons. Facilitatory area is intrinsically active, so it continues to discharge facilitatory impulses causing constant activation of γ-motor neurons. This causes stretching of the muscle spindle fibres resulting into reflex slight contraction of the extrafusal fibres of muscle under resting state (producing muscle tone). Inhibitory area in the medulla becomes active only if it receives impulses from the cerebellum or cerebral cortex.

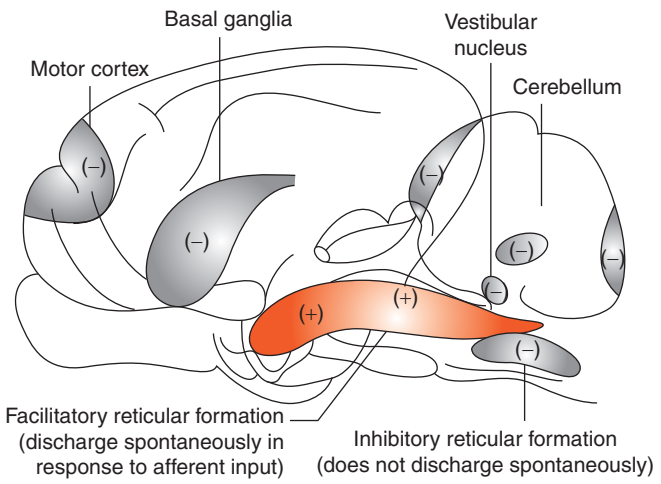


Fig. 10.9-9 Brain areas that have facilitatory (+) and inhibitory (–) effect on stretch reflex.

(ii) Role in maintaining posture. Static component of stretch reflex, the fundamental posture control mechanism, is especially prominent in the medial extensor muscles and antigravity muscles. For example, when a person is standing upright, gravity tends to stretch the quadriceps muscle. This stretching elicits stretch reflex resulting into sustained contraction of quadriceps as long as stretch is there. This maintains the extension around the knee joint and upright posture.

(iii) Role in control of voluntary movement. Stretch reflex helps the motor command system in performing voluntary movements. During activity generated by motor command system, the group Ia fibres from the muscle spindle inform the motor control system about the changes in muscle length. The constant level of Ia input to the CNS during a movement indicates that the motor command is being carried out. An increase in activity of Ia indicates that motor command is not being carried out. The CNS uses this information to readjust its command to the spinal cord. In addition, the Ia activity is also used at the spinal cord level to adjust the α-motor neuron activity as per need. Thus, the Ia activity provides the α-motor neuron with a source of excitatory input in addition to that coming from the higher centres.

2. GOLGI TENDON REFLEX (DISYNAPTIC REFLEX)

The Golgi tendon reflex, also called 'inverse stretch reflex', is a disynaptic reflex. The receptors involved are the Golgi tendon organs.

Golgi tendon organs

Golgi tendon organs (Fig. 10.9-10) are high threshold stretch receptors located in the tendons and musculoaponeurotic junction. They are placed in series between the muscle fibres and the tendon (in contrast to muscle spindles which are located in parallel to muscle fibres) and are thus stretched whenever the muscle contracts. Usually, 10–15 muscle fibres are connected in series with one Golgi tendon organ. Each Golgi tendon organ basically consists of a group of nerve endings covered by a capsule of connective tissue. In a given muscle, the Golgi tendon organs are less numerous than the muscle spindles.

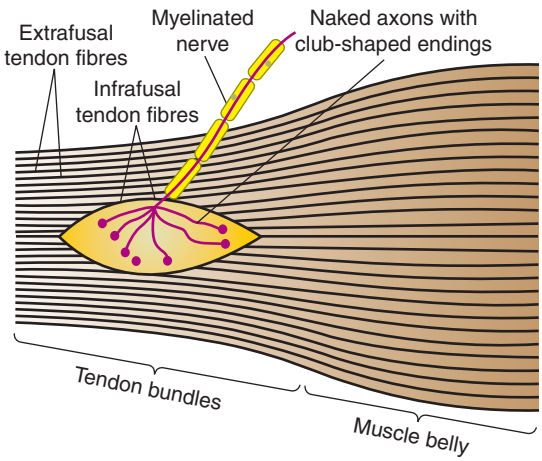


Fig. 10.9-10 Structure of Golgi tendon organ.

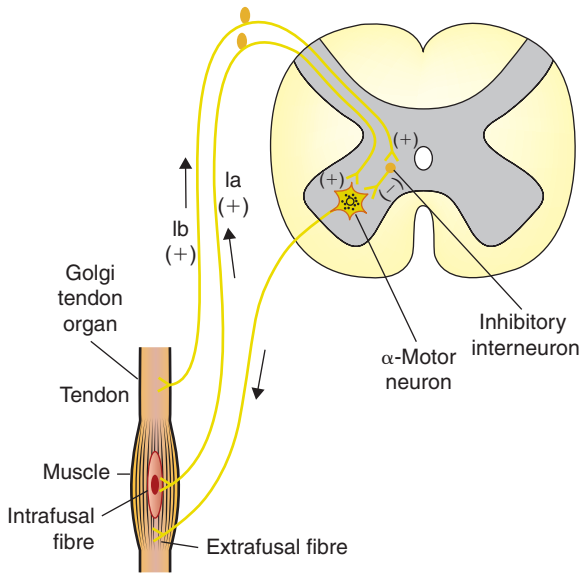


Fig. 10.9-11 Pathway of stretch and inverse stretch reflex.

- The Golgi tendon organs are supplied by *Ib-type sensory nerve fibres.* The nerve fibres supplying the Golgi tendon organ ramifies into many branches. Each branch ends in the form of a knob.
- The Golgi tendon organs have neither muscle fibres nor an efferent innervation.

Pathway and activity of reflex (Fig. 10.9-11)

- When a muscle contracts, the muscle tension increases. The Golgi tendon organ detects the muscle tension and sends impulses through afferent (group Ib) fibres, which enter the spinal cord through dorsal root.
- In the spinal cord, the group Ib afferents stimulate the inhibitory interneurons.
- The inhibitory interneurons in turn release inhibitory mediator glycine, which inhibits α-motor neurons and cause relaxation of the muscle that was originally contracted.
- At the same time, due to reciprocal innervation, the antagonistic muscles are excited.
- The Golgi tendon reflex, thus, displays reciprocal innervation but lacks after discharge and irradiation.

Physiological role or the functions of Golgi tendon reflex are:

- *Protective function.* Historically, this reflex has been described as a protective reflex in which a strong and potentially damaging muscle force reflexively inhibits the muscle, causing the muscle to lengthen instead of trying to maintain the force and risking damage.
- *Regulation of tension during normal muscle activity* is a more important role of this reflex. This reflex has been described as an *autogenic inhibition,* which indicates that the force generated when the muscle contracts is the stimulus for its own relaxation.

Clasp-knife reflex refers to an exaggerated form of the Golgi tendon reflex, which can occur with the disease of the corticospinal tracts (hypertonicity or spasticity). For example, when the arm is hypertonic, the increased sensitivity of the muscle spindles in the extensor muscles (triceps) causes resistance to flexion of the arm. Eventually, tension in the triceps increases to the point at which it activates Golgi tendon reflex, causing triceps to relax and the arm to flex closed like a jack knife, hence the name clasp-knife reflex. The physiological name for it is *lengthening reaction,* because it is the response of a spastic muscle to lengthening.

3. WITHDRAWAL REFLEX (POLYSYNAPTIC REFLEX)

Definition and receptors

Definition. Withdrawal reflex, also known as flexor reflex, is a cutaneous reflex which occurs in response to nociceptive (pain) stimuli and is characterized by the removal of a body part from painful stimulus.

Receptors for withdrawal reflex are *nociceptors* located in free nerve endings of Aδ and C fibres.

Pathway (reflex arc) of withdrawal reflex. Withdrawal reflex is a polysynaptic reflex consisting of following pathways (Fig. 10.9-12):

- The pain fibres carrying impulses, upon entering the spinal cord, synapse on many interneurons. Some of these also convey information to CNS. Others form several reflex pathways.
- A branch from some of the axons of interneuron in the reflex pathway feeds back on themselves forming the *reverberating circuits*, which are responsible for after discharge (Fig. 10.9-13).

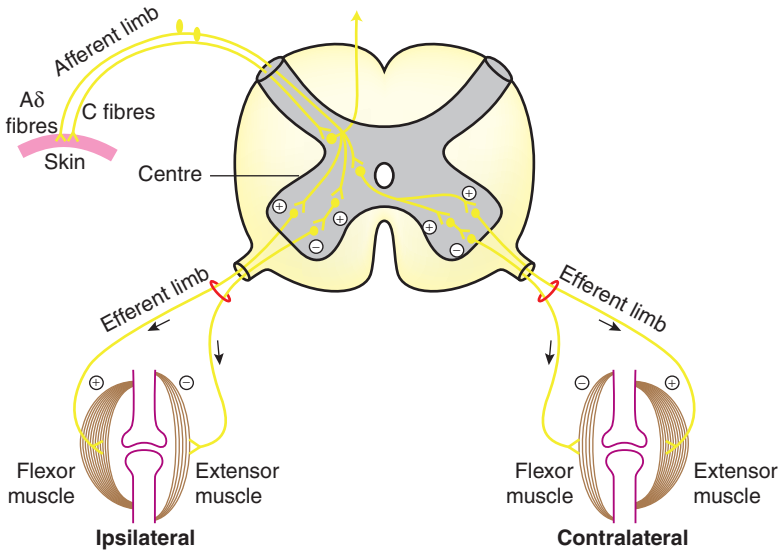


Fig. 10.9-12 Reflex arc of a polysynaptic reflex (withdrawal reflex or crossed extensor reflex).

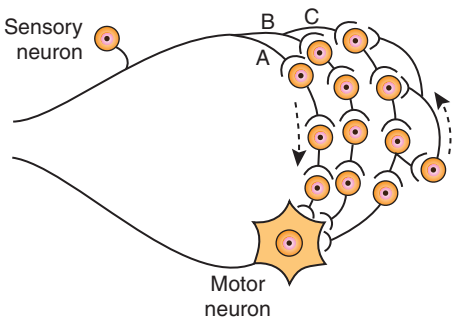


Fig. 10.9-13 Schematic depiction of connections between afferent and efferent neurons in the spinal cord. The dorsal root fibre has been shown to activate pathway A with three interneurons, pathway B with four interneurons and C with four interneurons. Note that one of the interneurons in the pathway C connects to a neuron that feedback on to previously excited neuron forming reverberating circuits.

- The interneurons form several pathways of different lengths to ultimately end on α-motor neurons as follows (Fig. 10.9-13):
  - Some of the interneurons project onto α-motor neurons on the ipsilateral side and stimulate the flexors which withdraw the limbs.
  - Some of the interneurons form inhibitory pathway and terminate on α-motor neurons supplying the extensor muscles on the ipsilateral side producing their relaxation. This is called reciprocal innervation, which ensures that the flexion movement is not impeded by contraction of the extensors.
  - Some of the interneurons cross to the opposite side of spinal cord and end on the α-motor neurons supplying the extensors on the contralateral side. In case of need, this pathway produces extension of the opposite limbs (crossed extensor reflex).

Effector organs

The effector organs of the withdrawal reflex are the skeletal muscles that cause withdrawal of the limb.

Response in withdrawal reflex

The reflex response to a painful stimulus varies from just withdrawal of the affected part to withdrawal of the whole body depending upon the strength of painful stimulus and location of the stimulus. The different types of responses observed in a withdrawal reflex are as follows:

Local sign refers to the ability of the reflex to confine to the portion of body affected by the noxious stimulus. Therefore, if an individual accidently touches a hot stove, it is likely that he or she will jerk only the hand away from the stove (one-limb response).

Flexor response. When a noxious stimulus is applied to a limb, the typical response is in the form of contraction of flexors and inhibition of extensors leading to flexion of the stimulated limb and its withdrawal from the irritating stimulus.

Crossed extensor reflex response (two-limb response). When a strong stimulus is applied to a limb, the response includes not only flexion and withdrawal of the limb but also extension of the opposite limb. This crossed extensor response is produced by the interneuronal pathway, which crosses to the opposite side of spinal cord. In lower limbs, crossed extensor reflex allows one limb to support the body, while other is raised off the ground.

Shifting reaction (four-limb reflex response). It is difficult to demonstrate this response in normal animals but is easily demonstrated in spinal animals (produced by a transverse

section in the lower region of spinal cord) in which the modulating effects of stimulus from the brain have been abolished. Application of electric shock to one hind limb of a spinal animal will produce a response in all the four limbs as:

- Flexion of the hind limb to which stimulus is applied,
- Extension of the contralateral hind limb,
- Extension of the ipsilateral forelimb and
- Flexion of the contralateral forelimb.

Widespread withdrawal response is obtained when the noxious stimulus is very strong. For example, if an individual picks up a hot coal, not only will the fingers open and drop it, but the entire arm will withdraw and the individual may even leap away from the fire.

Mechanism of varied grades of withdrawal response

Irradiation of the stimulus and recruitment of motor units are the mechanisms involved in the varied grades of response in withdrawal reflex (see page 824).

Function of withdrawal reflex

Withdrawal reflex is a protective reflex initiated by a potentially harmful (nociceptive) stimulus. The flexor response takes the limb away from the source of irritation. Withdrawal reflex is associated with a crossed extensor reflex, which helps to support the body and is of physiological significance in the context of regulation of posture. Withdrawal reflex is prepotent, i.e. it pre-empts all other reflex activities taking place at that time in the involved spinal cord segment.

CLINICAL REFLEXES

Clinically, the reflexes can be grouped as:

I. Physiological reflexes

1. Superficial reflexes. These reflexes are initiated in response to stimulation of receptors on skin (cutaneous reflexes, e.g. plantar, abdominal, cremasteric, bulbocavernous) or mucous membranes (mucous membrane reflexes), e.g. corneal, conjunctival and palatal reflex. The superficial reflexes are summarized in Table 10.9-1.

2. Deep reflexes. These reflexes are basically stretch reflexes and are elicited on stroking a tendon, so are called tendon reflexes (e.g. knee jerk, ankle jerk), the stretch reflex has been described in detail on page 825 however, the various clinically known stretch reflexes are summarized in Table 10.9-1.

3. Visceral reflexes are elicited from the visceral organ or at least one part of the reflex arc is formed by autonomic nerve, e.g. carotid sinus reflex (see page 256) micturition reflex (see page 455), oculocardiac reflex. A few of clinically known visceral reflexes are summarized in Table 10.9-1.

II. Pathological reflexes

The pathological reflexes are abnormal reflexes which are not found normally. They are elicited in pathological conditions, e.g.

- Babinski sign,
- Mass reflex,
- Clonus and
- Pendular movements.

1. Babinski sign. It is the abnormal plantar reflex, i.e. instead of plantar flexion of great toe there occurs dorsiflexion of great toe and abduction (fanning out) of small toes and also accompanied with flexion of knee and dorsiflexion at ankle joint. The abnormal plantar response is called *extensor plantar or Babinski sign positive.*

Significance. Babinski sign is present in following conditions:

- Upper motor neuron lesion. It is the most important sign.
- Physiologically, it is present in infants (below age of one year) due to non-myelination of pyramidal tracts and also during deep sleep.

2. Mass reflex. This reflex can be elicited in patients with spinal cord lesions. When the skin (on any portion in the midline) is stimulated by gentle pin pricks, there occurs evacuation of bowel or bladder, flexion of lower limb and sweating of skin below the level of lesion.

Significance. The patients suffering from spinal cord injuries are particularly trained to elicit mass reflex to evacuate bowel and bladder.

3. Clonus. Clonus means a series of rapid and jerky movements which occur due to involuntary contraction of the muscle in response to sudden rapid and constant stretch. Clonus signifies hyperflexia and hypertonia associated with an increased γ-efferent activity. Clonus is seen in calf muscles (producing ankle clonus) and quadriceps (patellar clonus).

Ankle clonus. To elicit ankle clonus support the slightly bent knee on one hand and hold the foot and suddenly dorsiflex the ankle and maintain the stretch for some time. This causes series of rhythmic plantar flexion at ankle joint.

Patellar clonus. The patient's leg is extended, then patella is suddenly pushed downwards towards the foot. Repeated contraction of quadriceps results in rhythmic movements of leg.

4. Pendular movements. In patients of cerebellar dysfunctions, while eliciting tendon jerk slows oscillatory movements develop instead of brisk movement. Such movements are called pendular movement and are manifestation of hypotonia and lack of restrictive effect.

Table 10.9-1 Characteristic features of clinical reflexes

Reflex	Method to elicit	Response	Spinal segment/cranial nerve and centre involved
I. Superficial reflexes			
(a) Cutaneous reflexes			
1. Plantar reflex	Strike the outer aspect of the sole of the foot with a blunt object (e.g. key) and move towards the ball of small toes.	Plantar flexion of the foot and toes	$L_5 - S_1$
2. Abdominal reflex	Lightly stimulate the wall of abdomen by stroking with key or some blunt object from out to inside (parallel to costal margin) in upper quadrants and (parallel to inguinal ligament) in lower quadrants of the abdomen.	Contractions of the underlying abdominal muscle. ***Note.*** The reflex is difficult to elicit in: Elderly individuals, obese persons and in multipara.	$T_7 - T_{12}$
3. Cremasteric reflex	Stimulate the skin of upper and inner part of the thigh	Pulling upwards of scrotum and testicles due to contraction of cremasteric muscles. This reflex may not be elicited in elderly individuals.	L_1 and L_2
4. Scapular reflex	Stroke the skin of interscapular region.	Contraction of supra- and infraspinatus muscles	C_5–T_1
5. Anal reflex	Gently stimulate the skin of perianal region.	Contraction of external and internal anal sphincters.	S_2–S_4
6. Bulbocavernous reflex	Gently pinch the dorsum of glans penis.	Contraction of bulbocavernous muscle.	S_3–S_4
(b) Mucous membrane reflexes			
7. Corneal reflex	Touch the cornea with wisp of cotton from lateral aspect.	Closure of eye of same and of opposite side.	*Afferents*: Via ophthalmic division of Vth (trigeminal) cranial nerve *Centre*: In the pons. *Efferents*: Via facial nerve to orbicularis oculi muscle
8. Conjunctival reflex	Touch the conjunctiva with wisp of cotton.	Closure of the eyes.	Pathway is same as for corneal reflex.
9. Palate reflex	Touch on the either side of posterior pharyngeal wall with a swab stick.	The contraction of the palate.	*Afferents:* Ninth cranial nerve. *Centre* Nucleus ambiguous *Efferents:* Tenth cranial (vagus) nerve.
II. Deep reflexes (Tendon reflexes)			
1. Knee jerk	The subject is in lying or in sitting position. Place the left hand under the knee (to be tested). Tap the tendon of the quadriceps midway between its origin and insertion with knee hammer.	Observe the extension of knee due to contraction of quadriceps femoris muscle. Sometimes if unable to elicit then apply reinforcement. (Jendrassik's manoeuvre, page 828.	Femoral nerve (L_2–L_4)
2. Ankle jerk	With foot slightly everted and dorsiflexed strike on the tendo-Achilles.	Plantar flexion of the foot occurs due to contraction of calf muscle.	S_1 and S_2
3. Triceps jerk	Keep the forearm of the subject to rest across his chest. Then tap the triceps tendon with broader side of the patellar hammer.	Contraction of triceps with extension of the elbow.	C_6 and C_7

4. Biceps jerk	Keep the position of elbow at right angle with forearm and the forearm is semipronated. Examiner then places his thumb or index finger on the tendon of the biceps muscle and then strikes on the finger (kept on biceps tendon) with pateller hammer.	Contraction of biceps with flexion of elbow.	C_5 and C_6
5. Supinator	Tap the lower end of the radius at styloid process. Keeping the position of elbow same as for biceps jerk.	Supination of forearm and flexion of elbow.	C_5 and C_6
6. Jaw jerk	Ask the subject to open the mouth slightly. Then place one finger firmly below the lower lip and tap on the finger in downward direction.	Contraction of masseter muscle causes closure of the jaw.	Afferent and efferents are carried by trigeminal nerve. Centre lies in the pons.

REGULATION OF POSTURE

- Physiologically, posture refers to the subconscious adjustment of tone in different muscles so as to maintain balance during displacement of the body caused by gravity or acceleration.
- The erect posture is a prerequisite to most of the somatic motor activities of man and other higher animals.
- Maintenance of erect posture during movements of the body and more so while performing physical work (*dynamic posture*) is more complicated than maintenance of posture while standing still (*static posture*). This uphill task is accomplished by a very complex and co-ordinated reflex activity occurring in response to afferent input from muscle joints, vestibular and visual receptors.
- For the purpose of understanding, the regulation of posture can be discussed under two main headings: mechanisms involved in maintenance of posture and role of different regions of nervous system in maintenance of posture.

MECHANISMS INVOLVED IN MAINTENANCE OF POSTURE

At any given moment in any position of the body (static or dynamic), the posture is maintained by alteration in the tone of different muscles, which is controlled by the stretch reflex. The stretch reflex is a spinal reflex influenced by supraspinal control. The input to higher centres involved in the control of muscle tone through certain reflexes (called postural reflexes) significantly contributes to the maintenance of tone and hence the posture. Thus, the two main mechanisms involved in maintenance of posture are:

- Muscle tone and
- Postural reflexes.

ROLE OF TONE IN ANTIGRAVITY MUSCLES IN MAINTENANCE OF POSTURE

Largely, the posture is maintained through reflex adjustments of tone in the antigravity muscles. The basic postural reflex involved in the control of muscle tone is stretch reflex described in detail (see page 834).

Posture control is required not only for holding the body in erect position but also for fixation of the body parts over adjoining body segments. The centre of gravity of head passes in front of the centre of gravity of atlanto-occipital joint. Thus head has got always a tendency to roll forwards. To hold the head in an erect position cervico-occipital muscles are to be maintained in a state of constant tension. Similar problem is encountered in maintaining the equilibrium of the body in an erect position.

In the upright position, gravity tends to displace the body downward, stretching quadriceps muscles as the legs flex at the knees. The muscle stretch evokes discharge from the muscle spindles of the quadriceps leading to its reflex contraction. This ensures that the knee joints, i.e. the main weight-bearing joints do not give way under the effect of gravity. This maintains the leg as a pillar of support and thus counteracts the gravitational displacement of the body.

In general, the antigravity muscles of the body are endowed with a somewhat higher muscle tone than the other muscles of the body.

In human beings, the flexors of upper extremity and extensors of lower extremity are the main antigravity muscles. Retractors of neck, the elevators of joint, supraspinatus, the extensors of back, rectal muscles of abdominal wall, extensors of knee and ankle are the muscles which exhibit greatest degree of tone. When these muscles completely relax (as in unconscious person), the body collapses.

Various postural reflexes (described below) influence the medial motor system and the motor neurons of antigravity muscles. The inputs to this system through the postural reflexes significantly contribute to the maintenance of tone.

Thus, tone is the result of activity of various medial system pathways that descend to excite both α- and γ-motor neurons that innervate antigravity muscles and their spindles. The two pathways of medial system that are most important in maintenance of tone are the *lateral vestibulospinal tract* and *pontine reticulospinal tracts*.

MAINTENANCE OF MUSCLE TONE

Stretch reflex, as mentioned earlier, plays the main role in maintenance of muscle tone. Though the stretch reflex is a spinal reflex, supraspinal control modifies the reflex in an intact animal (see page 827).

Mainly, extrapyramidal system is responsible for maintaining tone.

Supraspinal control on muscle tone is (Fig. 10.9-14) exerted by facilitatory and inhibitory areas in the brain stem through γ-motor neurons. For details see page 827.

IMPORTANT NOTE

Normally, the muscle tone is due to tonic discharge of γ-motor neurons to the muscles due to predominant effect of descending fibres of facilitatory reticular formation. Thus, muscle tone is normally not under the tonic control of α-motor neurons (tonic control of motor neurons should not be confused with the stimulation of α-motor neurons through corticospinal fibres during voluntary phasic contraction). It is important to note that tonic control of α-motor neurons is exerted almost entirely through the vestibulospinal pathway. However, the vestibular nucleus (especially the Deiter's nucleus) is constantly inhibited by corticospinal fibres as well as fastigiovestibular fibres from the cerebellum.

Under certain abnormal conditions and under experimental situation when the vestibular nucleus gets disinhibited, there occurs exaggeration of muscle tone that is α-led rather than γ-led.

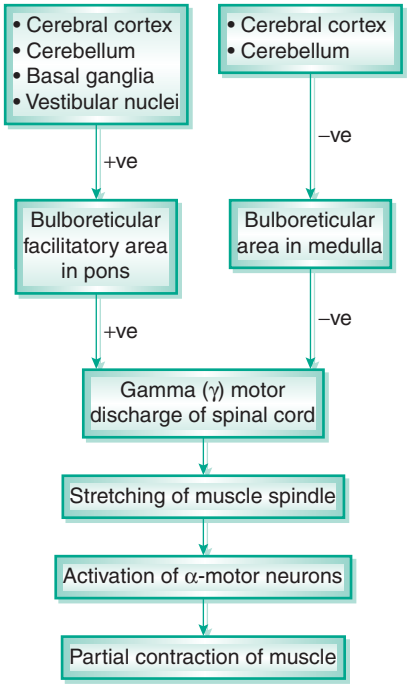


Fig. 10.9-14 Control of muscle tone.

POSTURAL REFLEXES

The postural reflexes help to maintain the body in upright and balanced position. They also provide adjustments necessary to maintain a stable posture during voluntary activity.

Reflex arc of postural reflexes is as follows (Fig. 10.9-15):

- *Afferent pathways* of reflex arc come from the eyes, the vestibular apparatus and the proprioceptors.
- Integrating centres are formed by the neuronal networks in the brain stem and spinal cord.
- *Efferent pathways* consist of α-motor neurons supplying the various skeletal muscles which form the effector organs.

Types of postural reflexes. Broadly, postural reflexes are of two types:

- *Static reflexes:* These are elicited by the gravitational pull and involve sustained contraction of muscles.

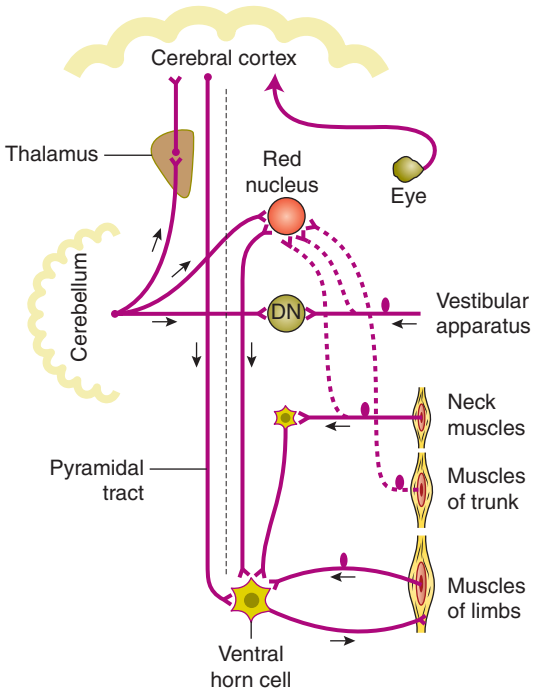


Fig. 10.9-15 Neuronal pathway of postural reflexes. DN = Deiter nucleus.

- *Statokinetic reflexes:* These reflexes, also called phasic reflexes, are elicited by acceleratory displacement of the body. They maintain a stable postural background for voluntary activity.

Both these types of postural reflexes are integrated at various levels in the CNS from the spinal cord to

cerebral cortex and are affected largely by the pyramidal pathways.

A. STATIC REFLEXES

Static reflexes are primarily involved in the adjustments to displacements produced by gravity. These are of three types:

- Local static reflexes,
- Segmental static reflexes and
- General static reflexes.

I. Local static reflexes

As the name indicates, the local static reflexes exert their effect on the same limb from which the stimulus was initiated. Some of the important local static reflexes include:

- Reflex control of antigravity muscle tone,
- Positive supporting reaction and
- Negative supporting reaction.

1. Reflex control of antigravity muscle tone

The most important of the local static reflexes is basic stretch reflexes (which has been described in detail on page 834) controlling tone in those extensor muscles which keep the body upright (*antigravity muscles*).

2. Positive supporting reflexes

Positive supporting reflex or reaction is characterized by simultaneous reflex contractions of both extensors and flexors of a limb (i.e. both the protagonists and antagonists) converting it into a solid rigid pillar. The positive supporting reaction plays an important role of *steading the ankle joint in standing position*. At the ankle joint, both dorsiflexion and plantar flexion are possible, but neither of them is desirable during standing position. The dorsiflexion of the foot would tip the body forward, while the plantar flexion would throw the body backward (Fig. 10.9-16). The stabilisation of ankle joint in intermediate position is possible by simultaneous contraction of extensor and flexors of foot brought about by the positive supporting reaction. Afferent impulses from the stimulated skin of sole (touch-pressure receptors) and the muscles (proprioceptors) cause reflex contraction of both flexor and extensor muscles acting on the ankle joint, converting the leg and ankle joint into one solid pillar.

3. Negative supporting reaction

Negative supporting reaction refers to the disappearance of positive supporting reaction. It is also an active phenomenon initiated by a stretch of the extensor muscles. This helps the limbs to be used for activities other than supporting the body weight.

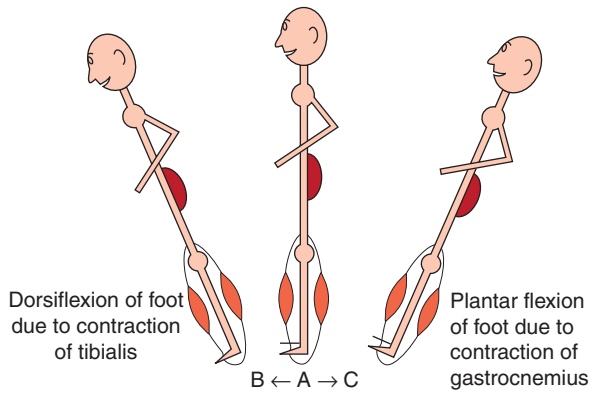


Fig. 10.9-16 Role of positive supporting reaction in stabilizing the ankle joint: A, simultaneous contraction of flexors and extensors of foot to stabilize ankle joint; B, dorsiflexion at foot produces forward fall and C, plantar flexion at foot produces backward fall.

Demonstration of local reflexes. The centres of the local static reflex are located in the spinal cord. These can be demonstrated in a spinal animal (see page 837).

II. Segmental static reflexes

The segmental static reflexes are characterized by a bilateral reflex response when stimulus is applied to one limb. The best example of segmental static reflexes is *crossed extensor reflex response component of withdrawal reflex.* In this reflex, a strong stimulus to one limb produces flexion in the ipsilateral limb and extension in the contralateral limb (see page 833).

Role of crossed extensor reflex in control of posture

- In the lower limb, this reflex allows one limb to support the body while other is raised off the ground. For example, when due to painful stimulus one limb is flexed reflexly, the extensor of the other limb compensates and sees to it that the body is not thrown off balance.
- The crossed extensor reflex also plays an important role during walking. During walking, on one side the flexors are active and the extensors are inhibited, while the reverse is seen on the other side.

Demonstration of static segmental reflex. The centres for these reflexes are situated in the spinal cord. These can be best demonstrated in a *spinal animal* (see page 838).

III. General static reflexes

General static reflexes are characterized by a generalised effect from the many muscle groups in the body in response to a stimulus that arises at one side of the body. For example, numerous postural adjustments occur in response to

changes in the head position. Broadly, general static reflexes can be divided into three groups:

- Attitudinal or statotonic reflexes,
- Long loop stretch reflexes and
- Righting reflexes.

(a) Attitudinal reflexes

Statotonic reflexes, also known as attitudinal reflexes, are initiated when the attitude of the body is changed, i.e. while standing on an inclined plane. These reflexes are of two types:

- Tonic labyrinthine reflexes and
- Tonic neck reflexes.

1. Tonic labyrinthine reflexes. These reflexes are produced in response to *alteration in the position of head relative to the horizontal plane,* e.g. while standing on an inclined plane. These reflexes decrease or increase the tone of the skeletal muscles of the limbs in accordance with the attitude of head.

Stimulus for tonic labyrinthine reflex is gravity.

Receptors for these reflexes are in the otolith organs, present in the labyrinthine apparatus.

Afferents. The afferent impulses generated from the receptors (present in otolith organ) travel along the vestibular nerves.

Centres for these reflexes are in the vestibular and reticular nuclei present in the medulla oblongata.

Efferents. The descending tracts employed are vestibulospinal and reticulospinal tracts which end on α-motor neurons of spinal cord.

Reflex response. The labyrinthine reflexes are particularly effective in the extensor muscles. The impulses from labyrinthine exert the same effect on all the four limbs.
Depending upon the position of head in relation to horizontal plane the reflex response produced is:

- When a quadruped stands on an inclined plane in such a manner that its *head sets tilted to right.* Tilting of the head to the right stimulates the labyrinth (vestibular apparatus) and evokes the tonic labyrinthine reflex. The reflex causes flexion of the left limbs and extension of the right limbs.

2. Tonic neck reflexes. These reflexes are produced in response to alteration in the position of head relative to the body.

Stimulus for tonic neck reflexes is stretch of neck muscles. *Receptors* of tonic neck reflexes are probably pacinian corpuscles in the ligaments of the cervical joints particularly atlantooccipital joint and also *muscle spindles* of neck muscles.

Centre for these reflexes lies in the medulla oblongata. *Efferent* paths are the corticospinal tracts.

Reflex response obtained depending upon the position of the head in relation to the body is:

- *Dorsiflexion (turning up) of head* causes extension of the forelimbs and flexion of the hindlimbs (Fig. 10.9-17A).
- *Ventroflexion (turning down) of head* causes flexion of the forelimbs and extension of the hindlimbs (Fig. 10.9-17B).
- *Turning of head sideways*, i.e. towards right or left produces flexion of the ipsilateral limbs and extension of contralateral limbs (Fig. 10.9-17C&D).

Role of tonic neck and labyrinthine reflexes. The tonic neck and labyrinthine reflexes bring about a redistribution of muscle tone in all the limbs and ensure that the body is not thrown off balance even when standing on an inclined plane.

The tonic labyrinthine reflex is active during the erect posture. This is because in erect posture the vestibular apparatus is thrown about 30° backwards. This results in a slight flexion of the upper limbs and extension of the lower limbs. When the head is tilted 30° forwards, the tonic labyrinthine reflex ceases but the concomitant flexion of the neck triggers the tonic neck reflex, which has the same effect on the limbs.

(b) Long-loop stretch reflexes

The long-loop stretch reflexes, also called *functional stretch reflexes*, are polysynaptic reflexes with their reflex arc centred in the cerebral cortex. These reflexes are continuously

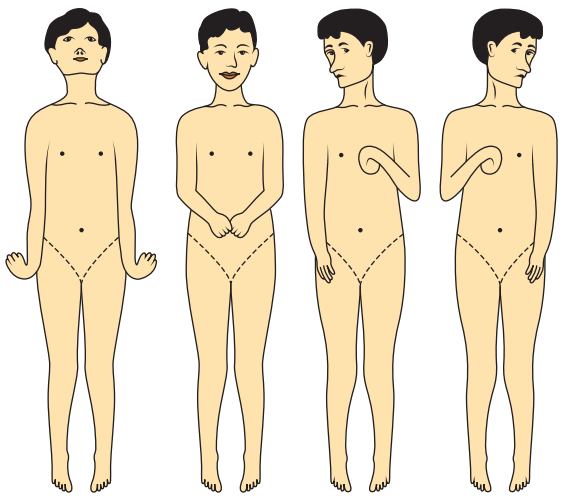


Fig. 10.9-17 A, Decerebrate rigidity (note extension of upper and lower limbs with extension of head), B, C and D decorticate rigidity (note in 'B' patient lying supine with head unturned, in 'C' and 'D' changes in position of hands and arms due to tonic neck reflex produced by turning of head to the right and left).

active in the erect posture and bring about a continuous correction of the sways that occur from moment to moment during standing. For example, when the body sways forwards, there occurs stretching of the gastrocnemius muscle. This initiates monosynaptic stretch reflex as well as long-loop polysynaptic reflex, which bring about reflex contraction in the gastrocnemius muscle resulting in correction of forward sway. In addition, the visual inputs which suggest that the body is swaying, also initiate long-loop postural reflexes.

The two long-loop reflexes (one proprioceptive, and the other visual) ensure that the body is not thrown off balance when tipped over its centre of gravity.

APPLIED ASPECTS

The importance of these two reflexes can be realised in patients with lesions of dorsal column, such as tabes dorsalis. *Sensory ataxia* seen in such patients is accentuated on closing of the eyes (Romberg's sign). The Romberg's sign is pathognomonic of sensory ataxia and helps to differentiate it from the cerebellar ataxia, in which this sign is absent.

(c) Righting reflexes

Righting reflexes help to maintain head and body into erect position under all circumstances. For example, if an animal is laid on its side or back, head at once rights itself, body follows and animal finally resumes the upright posture.

Note. Decerebrate animal though remains in the upright position, it can never actively resume the upright posture as it has no righting reflexes.

The righting reflexes consist of a chain of reactions following one another in an orderly sequence. Each reflex causes the development of the succeeding one. The righting reflexes are summarized in Table 10.9-2.

Centres of righting reflexes. Chief centre for all the righting reflexes, except the optical righting reflexes, is red nucleus lying in the mid brain. Red nucleus controls these reflexes through following tracts:

- *Rubrospinal tract.* It arises from the small number of large nerve cells, which form the *nucleus magnocellularis* part of the red nucleus.
- *Rubroreticular tract:* It arises from the large number of small nerve cells forming the *nucleus parvocellularis* part of the red nucleus.
- *Centre for optical righting reflex* lies in the visual cortex, from where impulses ultimately pass to neck muscles to right the head.

B. STATOKINETIC REFLEXES

Statokinetic reflexes are elicited by angular (rotatory) and linear acceleratory (progressive) stimuli to the labyrinthine receptors of vestibular apparatus.

These are programmed reflexes that depend on the motor cortex. Ultimately, these reflexes are mediated by lateral vestibulospinal tracts. These include:

1. Vestibular placing reaction. This reflex is evoked by linear acceleration through stimulation of receptors in the utricle and saccule. This reflex response is an adaptive reaction that prepares the animal or appropriate support by the limbs on surface contact. Thus, as soon as the foot comes in contact with any firm surface, the foot is reflexly placed on the surface and the leg muscles are adjusted so as to support the body.

2. Visual placing reaction. The placing response as described above can be initiated by visual cues as well and is then labelled as visual placing reaction. Many postural reflexes mediated by the vestibular system can be stimulated by visual stimuli. Thus, the visual system frequently compensates for lesions of the vestibular apparatus or its central pathways.

3. Hopping reaction. Hopping reactions occur in the form of hopping movements that keep the limbs in position to support the body when a standing animal is pushed laterally.

Thus, placing and hopping reactions, like the long-loop stretch reflex, ensure that the body is not thrown off balance when tipped over its centre of gravity.

SUMMARY OF POSTURAL REFLEXES

The various postural reflexes are summarized in Table 10.9-2.

ROLE OF DIFFERENT REGIONS OF NERVOUS SYSTEM IN MAINTENANCE OF POSTURE

The role of different regions of the nervous system in the maintenance of posture can be experimentally investigated (usually in a cat) by producing transection in the neuraxis at various levels.

ROLE OF SPINAL CORD: SPINAL ANIMAL

Spinal animal

The role of spinal cord in the maintenance of posture can be studied in a spinal animal. The spinal animal can be produced by a transection in the spinal cord at cervical region and respiration is maintained artificially by respiratory pump. If spinal cord is transected below the origin of phrenic nerve in the mid-thoracic region then diaphragmatic respiration continues and so the artificial respiration is not required.

Table 10.9-2 Various postural reflexes

Reflex	Stimulus	Response	Receptors	Integrating centre in CNS
A. Static reflexes				
I. Local static reflexes				
• Stretch reflex	Stretch	Contraction of antigravity muscles	Muscle spindles	Spinal cord and Mid brain
• Positive supporting reflex	Contact of skin of the sole of foot with ground	Contraction of flexors and extensors of the limb.	Touch and pressure receptors from skin of sole of foot. Proprioceptors from distal flexors.	Spinal cord
• Negative supporting reaction	Stretch of extensor muscles.	Disappearance of positive supporting reaction	Proprioceptors in extensors	Spinal cord
II. *Segmental static reflexes*				
• Crossed extensor reflex	Painful stimulus	Contraction of flexors of the ipsilateral limb and extensors of contralateral limb to support the body.	Nociceptors	Spinal cord
III. *General static reflexes* Attitudinal reflexes				
• Tonic labyrinthine reflex	Gravity (alteration of position of head relative to horizontal plane)	Extensor rigidity.	Otolith organs	Vestibular and reticular nuclei present in the medulla oblongata.
• Tonic neck reflex	Stretch of neck muscles due to alteration of position of head relative to body.	Flexion of forelimbs and extension of hind limbs on ventroflexion of head (turning down). Extension of fore limbs and flexion of hindlimbs on upward turning of head. Flexion of ipsilateral limbs and extension of contralateral limbs on turning the head side-ways.	Pacinian corpuscles in the ligaments of cervical joint (atlanto-occipital joint), and Muscle spindles of neck muscles.	Medulla
2. Long-loop stretch reflex	Stretch of the muscle due to swaying of body.	Continuous moment to moment corrections of sways which occurs during standing.	Muscle spindles (monosynaptic reflex) Visual receptor (long-loop reflex)	Spinal cord Cerebral cortex
3. Righting reflexes				
• Labyrinthine righting reflex	Gravity	Brings the head in upright level	Otolith organs in saccules of labyrinth.	Mid brain
• Body righting reflex (body on head righting reflex)	Pressure on side of body (differential stimulation of deep structures of the body wall).	Righting of head.	Exteroceptors	Mid brain
• Neck righting reflex (Neck on body righting)	Stretch of neck muscles	Righting of thorax and shoulders and then pelvis	Muscle spindles	Mid brain
• Body on body right ing reflex	Pressure on side of the body	Righting of body even when righting of head is prevented.	Exteroceptors	Mid brain
• Limbs righting reflex	Stretch of limb muscles	Appropriate posture of limbs	Muscle spindles	Mid brain
• Optical righting reflex	Visual cues	Righting of head	Eyes	Cerebral cortex

B. Statokinetic reflexes				
• Vestibular placing reaction	Linear acceleration	Foot placed on supporting surface in position to support body.	Receptors in utricle and saccule.	Cerebral cortex
• Visual placing reaction	Visual cues	Foot places on supporting surface.	Eyes	Cerebral cortex
• Hopping reactions	Lateral displacement while standing	Hops, maintains the limb in position to support the body.	Muscle spindle	Cerebral cortex

Effects of spinal cord transection. As described earlier, the effects produced by complete spinal cord transection occur in three stages:

- Stage of spinal shock,
- Stage of reflex activity and
- Stage of reflex failure.

Note. For details see page 706.

Posture in spinal animal during stage of reflex activity

Except the basic stretch reflex and supporting reflexes which are integrated in spinal cord (Table 10.9-2), all other postural reflexes are absent, as they require the integrity of upper motor neurons coming from various levels of neuraxis.

Postural characteristics of a spinal animal thus are:

- *Stretch reflex* (page 835) and *supporting reaction* (page 838) though present but are very weak and cannot support the weight of the animal. Therefore, the animal cannot stand on its legs.
- *Muscle tone* returns first in the flexor muscles; therefore, flexors become less hypotonic than extensors producing *paraplegia in flexion* (both lower limbs are in state of flexion).

ROLE OF BRAIN STEM: BULBOSPINAL ANIMAL OR DECEREBRATE ANIMAL

Decerebrate animal

Decerebrate animal is one in whom the brain stem is transected at an intercollicular level (between superior and inferior colliculi).

Characteristic features of a decerebrate animal are:

1. Decerebrate rigidity, i.e. spasticity in all the antigravity muscles occurs immediately after decerebration.

2. No spinal shock. Spinal shock does not develop with lesion at this level or any other higher level.

3. Postural reflexes present in decerebrate animal are those which have their integration centre in the spinal cord or medulla or pons. These include:

- *Stretch reflexes.* These are strongly positive. Decerebrate rigidity is basically due to harmoniously operating group of stretch reflexes.
- *Positive supporting reaction.* This can be elicited by application of pressure on the pads of fingers or toes. The afferent impulses from the skin and interossei muscles (which are stretched) cause reflex contraction of both extensors and flexors of the limb, converting limb into a rigid pillar. All the joints are locked. Limbs support the weight of the body and the degree of tone is adequate to maintain the upright posture, but is not sufficient to take up upright position.
- *Negative supporting reaction.* This can be elicited by a passive plantar flexion, which releases the limbs from positive reaction.
- *Crossed extensor reflex.* This can also be demonstrated, i.e. when one forelimb is flexed, the other forelimb is adjusted (page 838).
- *Tonic neck and tonic labyrinthine* reflexes are also present (for details see page 836). Therefore, in decerebrate animals, posture of limbs and trunk can be adjusted accordingly with the help of these reflexes.

4. Righting reflexes are absent, therefore, decerebrate animal can stand on its four legs but slight displacement causes the decerebrate animal to topple over.

Decerebrate rigidity

Decerebrate rigidity refers to a marked increase in the tone (hypertonia) of extensors, i.e. antigravity muscles occurring immediately after decerebration of the animal.

Characteristic features of decerebrate rigidity (Fig. 10.9-18)

- Hyperextension of all the four limbs.
- Dorsiflexion (hyperextension) of tail and head,
- Extreme hyperextension of the spine (opisthotonus) produces concave configuration of the back,

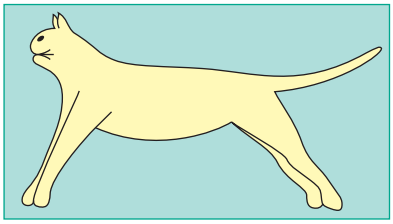


Fig. 10.9-18 Characteristic features of decerebrate rigidity in cat.

- The animal can be made to stand on four limbs but is easily toppled by a slight push and
- Postural reflexes which can be elicited in decerebrate rigidity are described above.

Mechanism of decerebrate rigidity

Depending upon the mode of production, the decerebrate rigidity is of two types:

- Classical decerebrate rigidity and
- Ischaemic decerebrate rigidity.

Mechanism of classical decerebrate rigidity. Classical decerebrate rigidity refers to the decerebrate rigidity, which occurs following transection of brain stem at intercollicular level. It is produced by an exaggerated stretch reflex due to increased activity of γ-motor neurons.

- Transection at the mid-collicular level cuts off all facilitatory and inhibitory corticobulbar extrapyramidal pathways. Hence, following decerebration, the inhibitory reticular formation having no intrinsic activity, becomes less active since none of it is driven by cerebellum only. While facilitatory reticular formation, which is mainly derived by ascending sensory stimuli, remains strongly active.
- Thus, the resulting release of spinal γ-motor neurons from the descending inhibitory reticular formation and continued effect of facilitatory reticular formation, markedly increases muscle spindle sensitivity to stretch resulting in rigidity of muscles.
- This rigidity is lost by deafferentation (cutting of afferents from muscle). This proves that decerebrate rigidity is due to increased activity of γ-motor neurons causing exaggerated stretch reflex.

Mechanism of ischaemic decerebrate rigidity. Ischaemic decerebration is obtained by ligating the common carotid artery and basilar arteries in which cerebral cortex is rendered ischaemic and non-functional. This safer alternative method of decerebration was attempted since classical decerebration was frequently associated with death of experimental animal.

- The rigidity observed after ischaemic decerebration is in fact due to disinhibition of α-motor neurons, i.e. exaggerated α-motor neuron discharge. The increased α-motor neuron drive results in direct stimulation of extrafusal fibres (*α-rigidity*).
- This rigidity is not lost by deafferentation (cutting off afferents from muscles). This proves that ischaemic decerebrate rigidity is not due to increased γ-motor neuron activity but is due to increased α-motor neuron activity.

Table 10.9-3 Differences between classical and ischaemic decerebrate rigidity

Classical decerebrate rigidity	Ischaemic decerebrate rigidity
1. *Produced by:* Transection of brain stem between superior and inferior colliculi.	Ligating both the common carotid arteries and basilar artery at the junction of pons and medulla.
2. *Rigidity observed* is type of spasticity which exhibits clasp-knife effect.	Rigidity produced is due to marked muscle tone which does not exhibit clasp-knife effect.
3. *Rigidity is mainly due to:* increased activity of gamma-motor neurons, hence also called γ-rigidity.	Increased α-motor neuron activity, hence also called α-rigidity.
4. *Deafferentation, i.e. cutting off posterior nerve root:* Abolishes rigidity, proving that it is reflex in origin.	Does not abolish rigidity, indicating that hypertonia is induced directly and not reflexly.
5. *Local injection of procaine into nerve trunk.* Reduced spasticity.	Does not reduce rigidity.
6. *Systemic administration of chlorpromazine* reduces spasticity.	Has no effect on rigidity.
7. Removal of anterior lobe of cerebellum increases rigidity.	Has no effect.

Classical versus ischaemic decerebrate rigidity. Differences between classical and ischaemic decerebrate rigidity are summarized in Table 10.9-3.

ROLE OF MID BRAIN: MESENCEPHALIC ANIMAL OR HIGH DECEREBRATE ANIMAL

Mesencephalic or high decerebrate animal is one in whom the brain stem is transected at the rostral border of mid brain.

Characteristic features of a mesencephalic animal are:

1. *Decerebrate rigidity,* similar to that of bulbospinal animal, is present but it disappears when the limb is performing a reflex activity.
2. *No spinal shock,* similar to the bulbospinal animal.
3. *Animal cannot only stand* but also typical quadrupedal walking movements can be reflexly performed.

4. *Righting reflexes,* having integration centre in the mid brain are present. These include:
   - Labyrinthine righting reflex,
   - Neck righting reflex,
   - Body on head righting reflex and
   - Body on body righting reflex.

   The chief advancement in postural regulation in mesencephalic animal over the bulbospinal animal lies in the presence of righting reflexes. By means of the righting reflexes, the mid-brain animal can bring its head right way up and get the body into the erect position under all circumstances.
5. *Pupillary light reflexes,* having integration centre in the mid brain are present (for details see page 921).
6. *Nystagmus,* the reflex response to rotational acceleration can be elicited (see page 846).

ROLE OF CEREBELLUM

Spinocerebellum regulates the postural reflexes by modifying muscle tone. It facilitates the γ-motor neurons in the spinal cord via cerebello-vestibulo-spinal and cerebello-reticulo-spinal tracts. The γ-motor neurons reflexly modify the activity of α-motor neurons and thus regulate the muscle tone. Thus, cerebellum forms an important site of linkage of α–γ systems responsible for muscle tone (for details see page 722).

ROLE OF BASAL GANGLIA: DECORTICATE ANIMAL

Decorticate animal is one in whom the whole cerebral cortex is removed but the basal ganglia and brain stem are left intact.

Postural characteristics of a decorticate animal

Moderate rigidity is present due to the loss of the cortical area that inhibits spinal γ-motor neurons discharge via reticular formation. It is seen only when the animal is at rest. It commonly occurs on the hemiplegic side after haemorrhage or thrombosis in the internal capsule.

Decorticate animal does not have such intense hypertonia as a decerebrate preparation. This is because the basal ganglia which are intact in decorticate animal activate the descending inhibitory reticular formation and thereby prevent hypertonia.

Typical posture in decorticate man consists of full extension of legs, arms lying across the chest, with semiflexion at elbow, slight pronation of forearm and flexion of wrist and fingers (Fig. 10.9-17A).

Postural reflexes. In decorticate man or animal, following reflexes can be elicited:

- Typical neck reflexes,
- Righting reflexes,
- Postural reflexes, which are seriously disrupted by decortication are:
  - Hopping reactions and
  - Placing reactions.

Note. It is easier to maintain a decorticate animal than a mid brain animal because temperature regulation and integration of visceral homeostatic mechanism is present in the hypothalamus.

MECHANISM OF STANDING IN MAN

As mentioned earlier, the tall human being has to stand over a narrow base of feet; therefore, maintenance of erect posture is more difficult than the quadruped animals. Mechanisms which play an important role in erect standing posture are:

Reflex adjustment in muscle tone of antigravity muscles undoubtedly plays an most important role in making the man stand erect. From this statement, it may be presumed that a continued contraction of most of the trunk and leg muscles keeps the posture upright. However, electromyographic studies have revealed very little muscle activity in a person standing quietly in upright position.

Configuration of hip and knee joints is such that they are kept extended by the gravity itself. However, a little activity of the antigravity muscles is required to maintain the very precarious balance. This explains the little muscle activity revealed by electromyographic studies.

The effect of gravity has to be opposed by reflex contraction of some of the antigravity muscles all the time, otherwise a standing man may fall in any direction (forwards, backwards or sideways). The different antigravity muscles which oppose the fall under various circumstances are:

- *Extensors of the trunk and flexors of the legs* contract sufficiently to restore the balance when the body sways forward.
- *Recti abdominis and leg extensors* contract to restore the balance when the body sways backward.
- *Contralateral external oblique abdominal muscles* maintain the balance when the body leans sideways.
- *Head has a tendency to sway more than the trunk:* Since the centre of gravity of head passes in front of the centre of gravity of atlanto-occipital joint; therefore, head always has got a tendency to roll forwards. To hold the head in erect position the cervico-occipital muscles are to be maintained in a state of constant tension.

Reflex changes in antigravity muscles described above are induced by:

- *Stretch receptors* in the trunk and leg muscles,

- *Visual afferents* also play an important role in reflex maintenance of upright posture in man. This is why, when the eyes are closed, the upright posture is less steady and there occurs more swaying (bending) of the trunk.
- *Vestibular afferents* help in maintaining the erect position of head.

VESTIBULAR APPARATUS AND EQUILIBRIUM

FUNCTIONAL ANATOMY

The internal ear or labyrinth is situated in the petrous part of the temporal bone. It consists of a bony labyrinth and membranous labyrinth (Fig. 10.9-19).

Bony labyrinth consists of three parts: vestibule, semicircular canals and the cochlea.

Membranous labyrinth is lodged within the bony labyrinth. It consists of:

- *Utricle and saccule*, which are lodged in the bony vestibule and are collectively called *otolith organs,*
- *Three semicircular ducts,* which lie within the body of semicircular canals and
- *Duct of cochlea,* which lies within the bony cochlea.

Vestibular apparatus

The semicircular canals and the utricle and saccule collectively form the vestibular apparatus. The vestibular apparatus plays an important role in maintaining posture and equilibrium.

Semicircular canals. The three semicircular canals are arranged at right angles to each other, so that all the three planes are represented as (Fig. 10.9-20):

- *Anterior semicircular canal* is vertical and placed at right angles to the long axis of the petrous bone. Thus, it lies in a plane that points forward and outward at about 45° from the sagittal plane.
- *Posterior semicircular canal* is also vertical but is placed parallel to the long axis of the petrous bone. Thus, it lies in a plane that points backward and outward at about 45° from the sagittal plane.
- *Lateral semicircular canal* is set in a horizontal position making an angle of about 30° with the horizontal plane. It is important to note that the right anterior and left posterior canals lie in the one plane while the left anterior and right posterior canals lie in the other plane.
- One end of each semicircular canal is dilated and is called *ampulla*. The ampulla contains the receptor organ known as *crista ampullaris* (Fig. 10.9-21).
- The semicircular canals open into the utricle by means of five orifices. The ampullary end of each canal and narrow end of horizontal canal open independently, while narrow ends of anterior and posterior canals open jointly by a common orifice.

Otolith organ refers combined to the two vestibular sacs called the utricle and saccule.

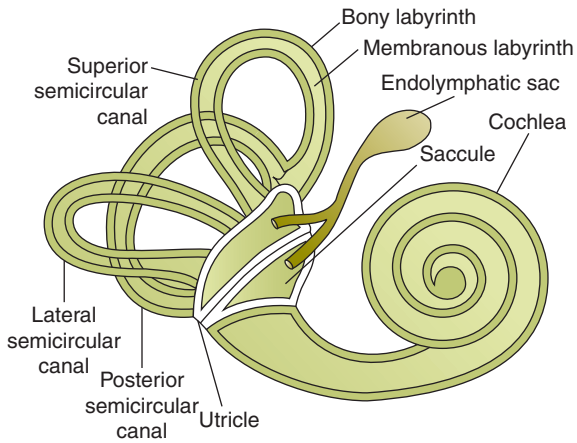


Fig. 10.9-19 Vestibular apparatus: semicircular canals and otolith organs.

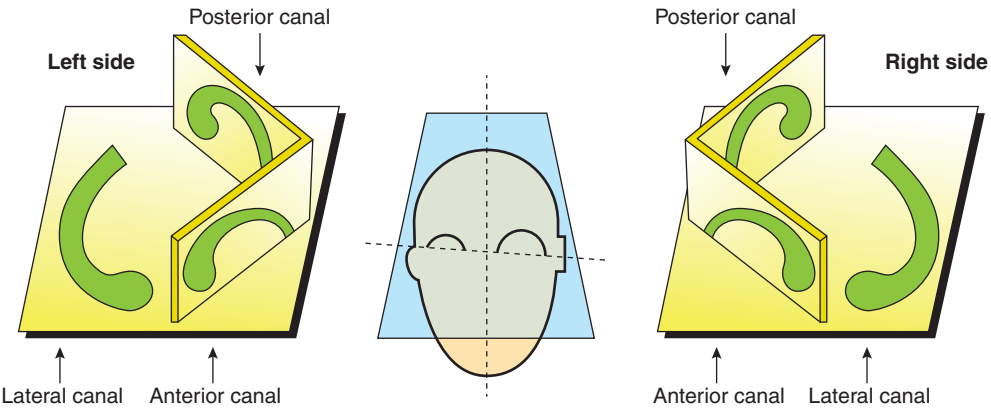


Fig. 10.9-20 Position of semicircular canals when head is tilted forward at 30°.

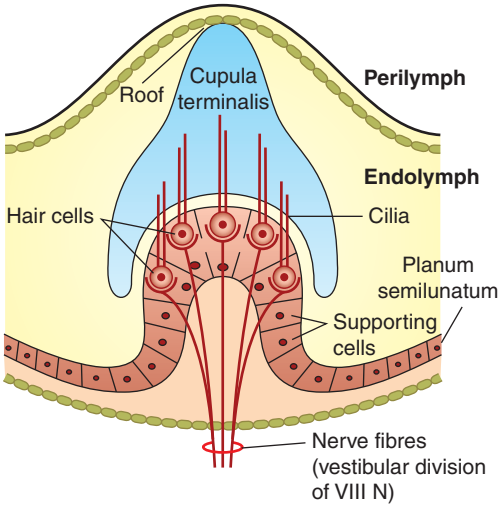


Fig. 10.9-21 Structure of crista ampullaris.

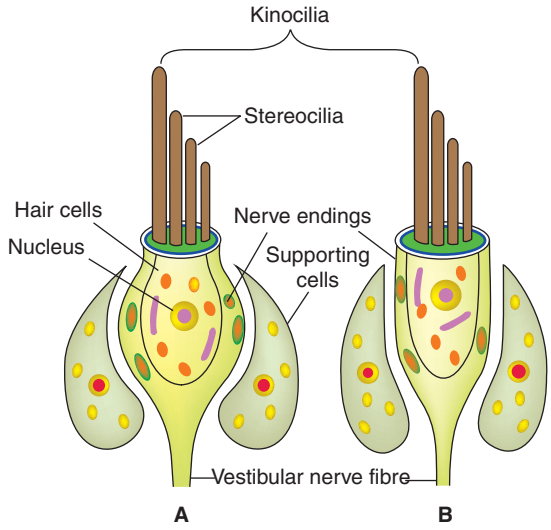


Fig. 10.9-22 Structure of hair cells: A, Type I and B, Type II.

Utricle is the larger of the two vestibular sacs in which open the three semicircular canals. It is indirectly connected to the saccule and ductus endolymphaticus by the *ductus utriculosaccularis.* The *ductus endolymphaticus*, after being joined by the ductus utriculosaccularis passes on to end in a small bag-like structure called *endolymphatic sac* (Fig. 10.9-19).

Saccule is a globular sac which is connected to utricle indirectly through the ductus utriculosaccularis and cochlea via the *ductus reunion* (Fig. 10.9-19).

VESTIBULAR RECEPTORS

The receptor cells of the vestibular system are called hair cells which are slowly adapting mechanoreceptors:

- The hair cells of the semicircular canals are located in a mass of tissue within the ampulla called *crista ampullaris.*
- The hair cells of the utricle and saccule are located in a mass of tissue called the *macula.*

Hair cells (Fig. 10.9-22)

The vestibular hair cells are of two types:

- *Type I hair cells* are flask-shaped. These make synaptic contacts with afferent nerve fibres only.
- *Type II hair cells* are cylindrical in shape and make synaptic contacts both with afferent and efferent nerve fibres.
- *Cilia of hair cells* (Fig. 10.9-22): The apex of each hair cell has a cuticular plate from which arise about 40–60 cilia. These cilia are called *stereocilia* which are motile.

A large non-motile cilium located at one end of the cell is called *kinocilium.*

Activity of hair cells. The hair cells are polarized cells. The membrane potential of hair cells is about –60 mV. When the stereocilia are bent toward the kinocilium the cell depolarizes and membrane potential is decreased to about –50 mV. When the stereocilia are bent away from the kinocilium, the cell hyperpolarizes. The changes in the activity of hair cells are conveyed to central nervous system by the afferent fibres, which form the vestibular part of eighth cranial nerve.

Receptors in semicircular canals

The receptors, i.e. the hair cells of the semicircular canals are located on a raised mass of tissue in the ampulla, called the *crista ampullaris.*

Structure of crista ampullaris. The crista ampullaris is a ridge-like area having following structures (Fig. 10.9-21):

- *Neuroepithelium* is formed by the hair cells (described above), which are innervated by the primary afferent fibres of vestibular nerve.
- *Secretory epithelial cells* surround the hair cells and form the so-called planum semilunatum around them.
- *Cupula* is dome-shaped large mass of gelatinous material in which are embedded the cilia arising from the hair cells. At its free end the cupula is in loose contact with the wall of ampulla. As a result, it forms a compliant seal that closes the lumen of the canal, preventing free circulation of endolymph.

Stimulation of receptors in semicircular canals. The movements produced in the endolymph by the angular movements of head pushes the cupula backwards, causing the cilia of hair cells to bend. Depending upon whether the stereocilia are pushed towards or away from the kinocilium, the hair cell depolarizes or hyperpolarizes.

It is important to note that cupula is unaffected by linear acceleration force, as it has the same specific gravity as the endolymph.

Receptors in otolith organs

The receptors (hair cells) of the otolith organs (utricle and saccule) are located in a raised mass of tissue called macula.

Structure of macula. The macula consists of (Fig. 10.9-23):

- *Neuroepithelium* of macula like that of crista ampullaris is formed by hair cells (both type I and II).
- *Supporting cells* are present around the hair cells.
- *Otolith membrane.* It is a flat gelatinous membrane covering the hair cells. This contains crystals of calcium carbonate called *otoliths or otoconia* (ear dust), which increase its specific gravity as compared to endolymph. The cilia of hair cells project in the gelatinous membrane.

Stimulation of receptors in otolith organs. The movements produced in the otolith membrane by linear acceleration of the head cause the cilia of hair cells to bend. This leads to excitation of vestibular afferents supplying these cells. *Orientation of the macula* is (Fig. 10.9-23):

- *Macula of utricle* is directed horizontally, so its cilia are in a vertical plane, which are stimulated by horizontally directed linear acceleration, e.g. moving in a car.
- *Macula of saccule* is directed vertically, so its cilia are in a horizontal plane and are stimulated by vertically directed linear acceleration, e.g. moving in a lift.

VESTIBULAR PATHWAYS

First-order neurons. The afferent fibres carrying impulses from the hair cells are dendrites of the bipolar cells, having their cell bodies in the *vestibular or Scarpa's ganglion* situated in the internal auditory meatus. These bipolar cells form the first-order neurons of the vestibular pathway. Axons of these cells form the vestibular division of vestibulocochlear (8th cranial) nerve, which enters the medulla ventral to the inferior cerebellar peduncle. These axons divide into ascending and descending branches which end in vestibular nuclei of the same side (Fig. 10.9-24).

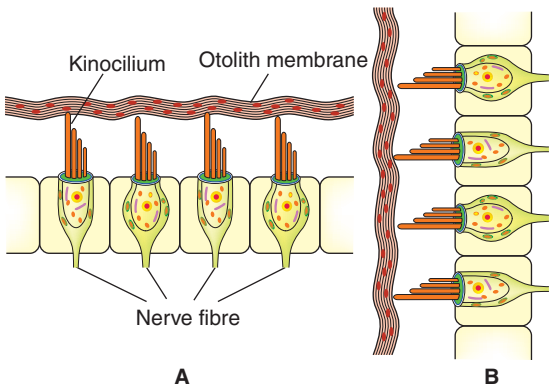


Fig. 10.9-23 Structure of macula: A, utricle (horizontally placed) and B, saccule (vertically placed).

Vestibular nuclei. The vestibular nuclei contain cell bodies of the second-order neurons of the vestibular pathway.

Afferent connections. In addition to the main afferents from vestibular apparatus, the vestibular nuclei also receive inhibitory fibres from the cerebrum and cerebellum (Fig. 10.9-24).

Efferent connections. Efferents from vestibular nuclei are:

- *Vestibulospinal tracts* (anterior and lateral) end directly at ventral horn cells. The inputs from the vestibular nuclei is *excitatory to antigravity* α-motor neurons.
- *Vestibulo-ocular tract.* These are the fibres, which ascend through the *medial longitudinal fasciculus* and terminate in the nuclei of third, fourth and sixth cranial nerves. These fibres are concerned with movements of eyeballs in relation to the position of the head.
- *Vestibulocerebellar fibres* pass through the inferior cerebellar peduncle and terminate in flocculonodular lobe and fastigial nuclei in the cerebellum of both sides.
- *Vestibuloreticular spinal tract.* Some fibres from the vestibular nuclei reach the reticular formation of brain stem, ultimately forming the vestibulo-reticulo-spinal tract.
- *Vestibulo-rubro-spinal tract.* Some fibres from the vestibular nuclei reach the red nucleus forming the vestibulo-rubro-spinal tract.
- *Vestibulo-thalamo-cortical fibres.* Some fibres from the vestibular nuclei pass via medial lemniscus to the opposite thalamus and thence to the opposite temporal lobe.

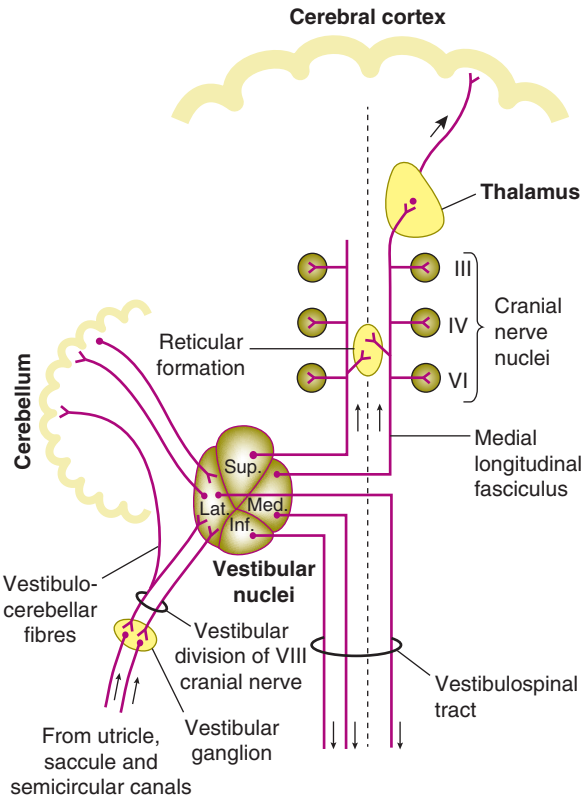


Fig. 10.9-24 Neural pathway from vestibular apparatus.

MECHANISM OF FUNCTIONING OF VESTIBULAR APPARATUS

A. Mechanism of functioning of semicircular canals

Salient features of functioning of semicircular canals

- *Receptors* of semicircular canals are stimulated by rotatory movements or *angular acceleration* of the head.
- *Semicircular canals are oriented* in three different planes, so movement of the head in any direction generates an unique pattern of activity within the semicircular canals. The three axes of the semicircular canals are those activated while:
  - Nodding the head up and down (as in signifying yes). This movement occurs along transverse axis,
  - Shaking the head from side to side (as in signifying no). This movement occurs along the vertical axis and
  - Tilting the head so that ear touches the shoulders. This movement occurs along the anteroposterior axis.
- Receptors of horizontal canals are stimulated during rotation of head in vertical axis while receptors of vertical canals are stimulated during rotation of head in anteroposterior or transverse axis. However, the mechanism of stimulation of receptors is same in all the canals.
- Receptors of semicircular canals are stimulated only at the beginning and at the stoppage of rotatory movements. During continued rotation at a constant speed, these receptors are not stimulated rather they are adapted as explained.

Mechanism of stimulation and adaptation of receptors of semicircular canals

1. At the beginning of movement

Movements to the right (i.e. clockwise rotation along the vertical axis) stimulate the hair cells in the right horizontal canal and inhibit these in the left horizontal canal. As shown in Fig. 10.9-25, when the head begins to move the horizontal canals move in clockwise direction but the endolymph within the semicircular canals lags behind because of inertia. This phenomenon causes relative displacement of endolymph in the direction opposite to that of the rotation of the head. That is endolymph is pushed in anticlockwise direction (Fig. 10.9-25B).

- *In right semicircular canal,* the endolymph is pushed towards the ampulla causing the cupula to move towards ampulla. As a result, the stereocilia are pushed towards the kinocilium, leading to depolarization (stimulation) of hair cells.
- *In left semicircular canal,* the endolymph is pushed away from the ampulla causing the cupula to move away from the ampulla. As a result, the stereocilia are pushed away from the kinocilium leading to hyperpolarization (inhibition) of hair cells. This combination of excitation of one ampulla and inhibition of ampulla from other canal forms the basis of the direction of movement. At the beginning of movement frequency of discharge from excited hair cells may increase to a frequency of 100–500 impulses/min from a resting discharge of 50–100 impulses per minute.

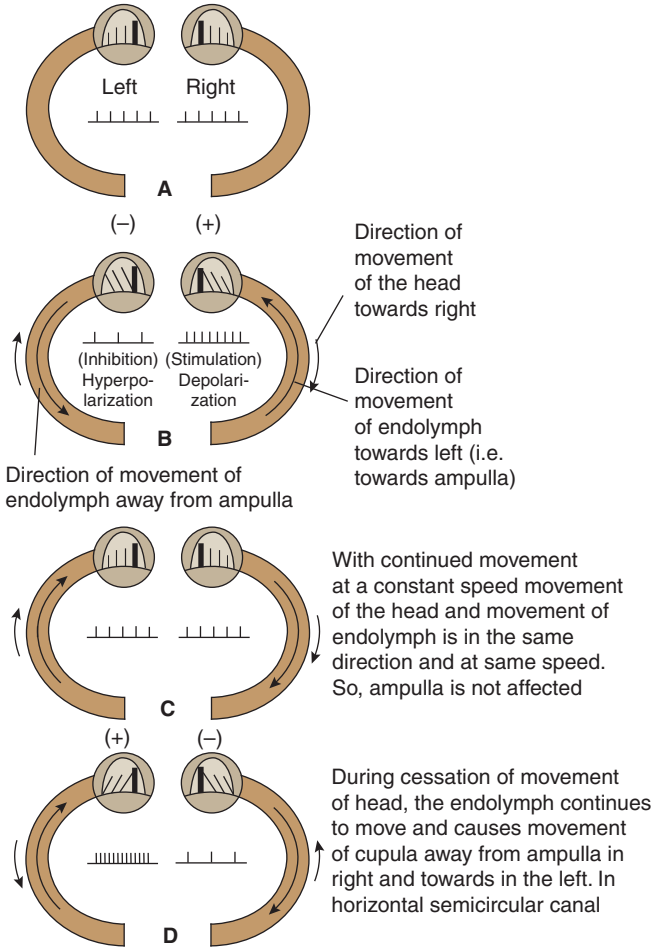


Fig. 10.9-25 Mechanism of stimulation of receptors in horizontal (lateral) semicircular canal during rotation of head towards right: A, resting position; B, when head begins to rotate to right; C, after 15–20 s of continued movement of head at a constant speed and D, when the head stops moving.

Movements to the left (i.e. counterclockwise movements), on the other hand, stimulate the hair cells in the left horizontal semicircular canal and inhibit those in the right horizontal canal by the same mechanism as explained above.

2. After 15–20 s of continuous movement at a constant velocity there occurs adaptation of receptors. After 15–20 s of continued movement of head at a constant velocity, the endolymph also takes up the same rate of movement as its canals and the cupula returns to its original resting position. So, hair cells are no more excited or inhibited and return to their resting membrane potential and resting discharge of about 50–100 impulses/min (Fig. 10.9-25C). Thus, the receptors in semicircular canals show signal

changes in motion (acceleration) but are insensitive to movements at a constant angular velocity. This state of insensitiveness of receptors during a constant angular velocity is referred to as *state of adaptation of receptors.*

3. When the head stops moving. When the head stops moving, i.e. during cessation or deceleration of movement, the endolymph within the canals continues to move. That is endolymph is now pushed in opposite direction (Fig. 10.9-25D):

- *In right semicircular canal,* the endolymph is pushed away from the ampulla causing the cupula to move away from the ampulla. As a result, the stereocilia are pushed away from the kinocilium leading to hyperpolarization (inhibition) of hair cells.
- *In left semicircular canal,* the endolymph is pushed towards the ampulla causing cupula to move towards ampulla. As a result, stereocilia are bent towards the kinocilium leading to depolarization (stimulation) of hair cells.

The information received from the semicircular canals during rotation of the head along three perpendicular axes is used by the CNS to interpret the speed and direction of head movement and to make appropriate adjustments in posture and of eye positions.

B. Mechanism of functioning of utricle and saccule

General features of functioning of utricle and saccule are:

- These provide information about linear acceleration and change in head position relative to the force of gravity.
- *Receptors* (hair cells) present in the maculae of utricle and saccule act as the stretch receptors, the effective stimulus being the pull of gravity on the otolith membrane. These receptors discharge tonically even in the absence of head movement because of pull of gravity on the otolith. So, these receptors show little adaptation (of receptors of semicircular canals).
- During linear acceleration of the head the otolith membrane having more specific gravity lags behind due to inertia. This causes cilia of hair cells embedded in otolith membrane to bend. This leads to excitation of vestibular afferents supplying these cells.

Functioning of utricle. As mentioned earlier, the macula of utricle is directed horizontally and so its cilia are in vertical plane (Fig. 10.9-23). These vertically oriented cilia are stimulated by horizontally directed linear acceleration, e.g. moving in a car. These hair cells are also stimulated during dorsiflexion or ventroflexion of the head, i.e. by nodding the head up and down (as in signifying yes).

Functioning of saccule. As mentioned earlier, macula of saccule is directed vertically and so its cilia are in horizontal plane (Fig. 10.9-23). These horizontally oriented cilia are stimulated by vertically directed linear acceleration, e.g. moving in a lift up or down. These hair cells are also stimulated when the head is tilted sideways, e.g. if the head is tilted laterally to the right the otolith membrane of macula of right saccule hangs downwards and pulls on its macula, which is maximally stimulated; and the otolith membrane of left saccule points upwards and rests on the macula. This being the position of minimal stimulation of the nerve endings.

VESTIBULAR REFLEXES

1. Vestibulo-ocular reflex

The vestibulo-ocular reflex maintains the visual fixation during movements of the head by producing reflex nystagmus and post-rotatory nystagmus as described:

Nystagmus. For example, when the head is rotated to the left, the eyes move slowly toward the right in order to keep the image on the fovea. When the eyes have rotated as far as they can, they are rapidly returned to the centre of the socket. These reflex movements of the eyes are called nystagmus. Thus, nystagmus has two components of the movements:

- *Slow components*, i.e. slow movement of the eyes to maintain visual fixation is initiated by receptors in the semicircular canals. When the head rotates to the left, receptors in the left horizontal canal are stimulated. Their axons activate reflex movements of the eyes toward the right through the impulses reaching the nuclei of third, fourth and sixth cranial nerves.
- *Quick component.* When slow movement of eyeballs is limited, the eyeballs move to a new fixation point in the direction of rotation of head. This movement to a new fixation point occurs with a jerk. So, it is called the quick component. The quick component of nystagmus is due to impulses from the vestibular nuclei to the ocular muscle.

Post-rotatory nystagmus occurs after the body has been rotated and the movement ceases. This is due to movement of cupula in the opposite direction caused by the endolymph when rotation is stopped.

2. Otolith reflexes

Otolith organs initiate a reflex that prevents leg injuries when an individual walks downstairs or jumps from a platform. When making such a descent, the muscles of the leg begin to contract before the feet reach the ground to cushion the force of impact.

- The otolith receptors responsible for this reflex are stimulated by the linear acceleration of the head that occurs during descent.
- Individuals lacking otolith reflexes are prone to leg injuries because of the large contact force that occurs during descent (e.g. stepping off a bus).

FUNCTIONS OF VESTIBULAR APPARATUS

1. Role in maintenance of equilibrium. The otolith organs detect change in position of head and help in maintenance of equilibrium under static condition.

- The otolith organs also detect linear acceleration of the head and help in maintenance of equilibrium during such movements.
- Semicircular canals detect angular acceleration and help in maintaining equilibrium during dynamic phase. They also have a predictive function.
- When the person is in dynamic state, they predict ahead of time that the person is likely to fall off balance and help nervous system to do adjustments to prevent a fall.

2. Role in maintenance of posture. The vestibular apparatus plays an important role in maintenance of posture through vestibular reflexes which include:

- Vestibular placing reaction,
- Righting reflexes,
- Vestibulo-ocular reflex and
- Vestibulo-otolith reflex.

MAINTENANCE OF EQUILIBRIUM

Equilibrium refers to the maintenance of line of gravity constant at rest and during movement by adjusting the tone of different muscles. While the term posture signifies an unconscious adjustment of tone of different muscles so as to maintain balance during rest as well as during movements.

Role of various parts of neural system in maintaining equilibrium

1. Role of vestibular apparatus (as described above)

2. Role of cerebellum

Uvula of cerebellum gets impulses from macula of utricle and saccule and helps in maintaining equilibrium under static conditions.

Flocculonodular lobe of cerebellum gets impulses from the semicircular canals and helps in maintaining equilibrium during rapid changes in direction of motion.

3. Role of brain stem

Main role is played by four pairs of vestibular nuclei present in the brain stem:

Superior and medial vestibular nuclei receive signals from the semicircular canals and send impulses to:

- *Medial longitudinal fasciculus* to cause corrective movements of eyes and
- *Medial vestibular tract* to cause appropriate movements of the neck and head.

Lateral vestibular nuclei receive signals from the otolith organs and in turn send:

- *Through lateral vestibulospinal tract* to spinal cord for controlling body movements.

Inferior vestibular nuclei receive signals from the semicircular canals and utricle and in turn send signals to:

- Cerebellum and
- Reticular formation of brain stem.

APPLIED ASPECTS

The important applied aspects in relation to vestibular apparatus which need special emphasis are:

- Vestibular dysfunctions and
- Experimental stimulation of semicircular canal.

A. Vestibular dysfunctions

1. Motion sickness

Aetiopathogenesis. The motion sickness is a symptom complex occurring due to excessive and repeated stimulation of vestibular apparatus while travelling in automobile, ship, aircraft or spacecraft. The psychological factors like anxiety about the unfamiliar mode of travel may be an additional factor in causation of motion sickness. The disease occurring during travelling by ship is referred to as sea sickness.

Characteristic features of motion sickness are:

- Unpleasant sensation of rotation accompanied by nausea, vomiting, sweating, pallor, salivation, headache, disorientation and even diarrhoea. Most of the symptoms and signs are the effects of vestibular stimulation on the medullary autonomic centres.

Prevention. Motion sickness can be prevented by taking antiemetic drugs, such as Avomine and by avoiding greasy and bulky food before travelling.

2. Meniere's disease

Aetiopathogenesis. It is caused by overdistension of the membranous labyrinth, probably due to oversecretion *(endolymphatic hydrops).*

Characteristic features. Meniere's disease originates in the labyrinth and typically present as a triad consisting of:

- Fluctuating deafness of sensorial type,
- Tinnitus which may be very troublesome and
- Episodic attacks of rotatory vertigo.

The disease is usually unilateral to start with and the common age of onset is 35–50 years and comes in attacks. Patient usually has nausea, vomiting and fullness of ear in addition to the above listed triad.

Treatment in patients with frequent attacks involves the implantation of a small tube or shunt into the abnormally swollen endolymphatic sac.

3. Labyrinthectomy

Bilateral labyrinthectomy, i.e. removal of labyrinthine apparatus on both sides is characterized by:

- *Equilibrium* is maintained by visual sensation. The individual cannot right himself when blindfolded.
- *Postural reflexes* are severely affected.
- *Muscle tone* is decreased but there is no permanent loss.
- *Hearing* loss is also there.

Unilateral labyrinthectomy, i.e. removal of labyrinthine apparatus on one side is characterized by:

- *Oblique deviation of the eyeballs*, i.e. one eyeball is rolled upwards and outwards and the other downwards and inwards,
- *Nystagmus*,
- *Rotation and lateral flexion of the head*, so that occiput is turned to the side of lesion and
- *Flexion of limbs* on the side of lesion and extension of limbs on the opposite side.

B. Experimental stimulation of semicircular canals

The semicircular canals can be stimulated by two methods:

- Rotational movement by Barany chair and
- Caloric stimulation.

1. Stimulation by rotational movement using Barany chair

Method. The subject is made to sit in the chair with head tilted forward at 30°. The chair is rotated at 30 rpm for 20 s.

Effects. *During rotation* with eyes open, *nystagmus* occurs continuously throughout the period of rotation. After rotation in Barany's chair for 20 s at 30 rpm, following effects are noted:

- *Post-rotatory nystagmus* occurs for about 30 s.
- *Dizziness*, i.e. feeling of unsteadiness occurs immediately after stoppage of rotation. It is associated with feeling of rotation in the opposite direction.
- *Vertigo*, i.e. feeling of rotation even after stoppage of rotation.
- *Nausea and vomiting* may occur after rotation for a longer period.

2. Caloric stimulation

The semicircular canals can be stimulated by introducing hot (44°C) or cold (30°C) water into the external auditory meatus.

Mechanism. The transmission of change in temperature into labyrinth alters the specific gravity of the endolymph. As a result the cupula is set into motion and the hair cells are stimulated.

Effects. Caloric stimulation produces the same effects as rotational movement, i.e. there occurs:

- Vertigo,
- Dizziness and
- Nystagmus.

APPLIED ASPECTS

1. Caloric stimulation is used as a clinical test for diagnostic purpose.
2. While irrigating, the ear canal for treatment of ear infections, it must be ensured that fluid used is at the body temperature level, otherwise annoying symptom of caloric stimulation will occur.

Chapter

10.10

Limbic System and Physiology of Emotional, Behavioural and Motivational Mechanisms

LIMBIC SYSTEM
- **Physiological anatomy**
- **Functions**

PHYSIOLOGY OF EMOTIONS
- Components of emotions
- Theories of genesis of emotions
- Emotional behaviour
- Neural substrate of emotions

PHYSIOLOGY OF MOTIVATION
- Neural mechanism
- Concept of reward and punishment
- Role of neurotransmitters

PHYSIOLOGICAL BASIS OF PSYCHOTIC DISORDERS
- Depression
- Mania
- Schizophrenia

LIMBIC SYSTEM

PHYSIOLOGICAL ANATOMY

Components of limbic system

The term limbic has been derived from the word 'limbus' which means a ring. Thus, the term limbic system is applied for those parts of the cortex (limbic cortex or limbic lobe) and subcortical structures that form a ring around the brain stem. Previously, this area was called rhinencephalon because of its relation to olfaction. It is now known to play, apart from olfaction, a role in functions like behavioural activity, emotions, motivational drives, memory and regulation of viscera and so it is also referred to as 'visceral brain.' Components of limbic system are (Fig. 10.10-1):

Limbic cortex, or the so-called limbic lobe, surrounds the subcortical structures of the limbic system. Phylogenetically, limbic cortex is an older part of the cerebral cortex (allocortex) having primitive histological structures. Limbic cortex is composed of (Figs 10.10-1 and 10.10-2):

- Orbitofrontal cortex,
- Subcallosal gyrus,
- Cingulate gyrus,
- Parahippocampal gyrus and
- Uncus.

Subcortical structures included in the limbic system are:

- Hypothalamus,
- Septum,
- Paraolfactory area,
- Anterior nuclei of thalamus,

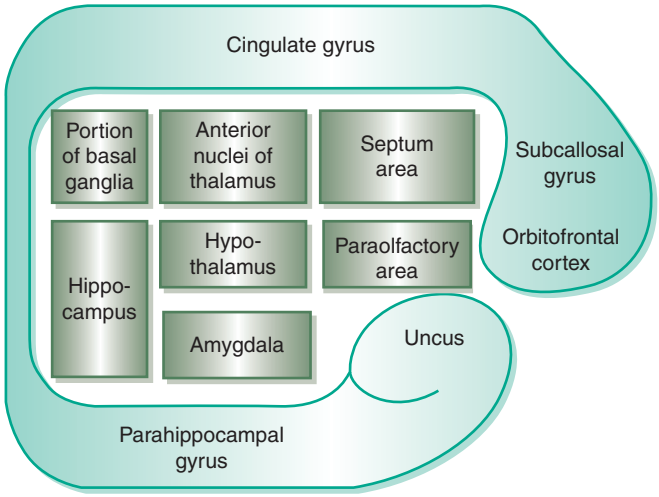


Fig. 10.10-1 Diagrammatic representation of the structures forming limbic system.

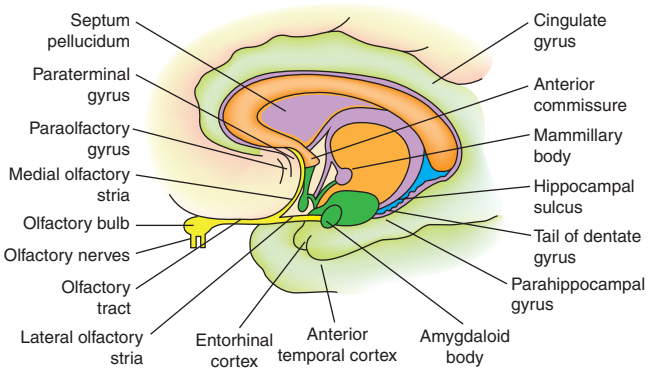


Fig. 10.10-2 Medial surface of right cerebral hemisphere showing limbic cortex (limbic lobe) and other components of limbic system.

- Amygdala,
- Portions of basal ganglia and
- Hippocampus.

Connections

Bundles of axons connecting the various components of limbic system closed circuit are called papez circuit.

Papez circuit refers to a closed circuit formed by connections between the cingulate gyrus (located in the prefrontal lobe), hippocampus, mammillary bodies and anterior nucleus of thalamus (Fig. 10.10-3). This circuit is responsible for resting EEG (page 864) and for those emotions and aspects of behaviour that are related to preservation of the individual and species.

Efferent projection of limbic system is shown in Fig. 10.10-4.

Characteristic features of limbic system connections are:

- *Limbic system has very little connection with the neocortex.* Because of this, emotional and instinctual behaviour is not under voluntary control, especially in lower animals. Thus, from a functional point of view, neocortical activity does modify emotional behaviour but it cannot be turned on and off at will.
- *Prolonged after discharge* is shown by the anatomic closed circuit of the limbic system following a sensory experience.

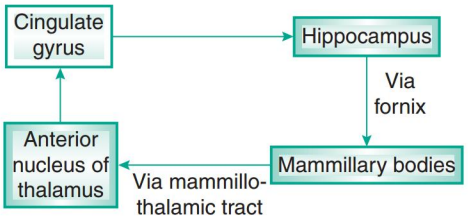


Fig. 10.10-3 The Papez circuit.

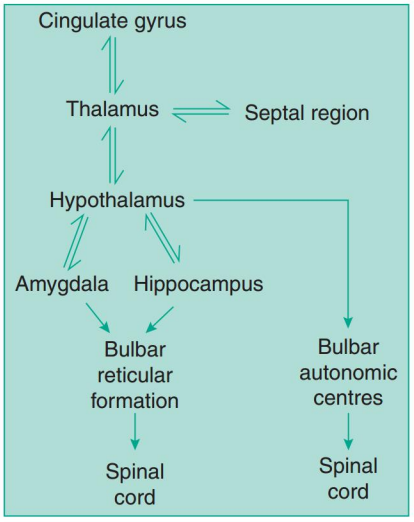


Fig. 10.10-4 Efferent projection of limbic system.

Therefore, the emotional responses are usually prolonged, i.e. continue long after the end of the stimuli that produce them.

FUNCTIONS OF LIMBIC SYSTEM

Most of the functions of limbic system are intimately related to the functions of hypothalamus, which have been described in Chapter 10.3, page 744. These include:

1. Autonomic functions

Stimulation of many parts of the limbic system specially that of amygdala produces autonomic responses, such as changes in cardiovascular, respiratory and gastrointestinal system through hypothalamus. Such changes are also observed during the emotional states. Autonomic functions of hypothalamus are described on page 741.

2. Regulation of feeding behaviour

Limbic system regulates feeding behaviour mainly through hypothalamus and amygdala.

Hypothalamus regulates food intake through the *feeding centre* and *satiety centre* (for details see page 743).

Amygdala. Stimulation of amygdala produces movements associated with eating (chewing, swallowing and licking). On the other hand, lesions of amygdala produce moderate hyperphagia (overeating). There may be indiscriminate ingestion of edible or non-edible materials.

3. Regulation of sexual behaviour and reproduction

The sexual activity comprises two components:

Sexual behaviour. The basic sex drive (urge to copulate) is an instinctual behaviour as food intake. It is the function of limbic system and hypothalamus, which in turn are influenced by gonadal hormones and cerebral cortex.

Sexual behaviour is controlled by neural and endocrinal factors.

(i) Neural control

Neocortex, limbic cortex and hypothalamus play an important role in determining sexual behaviour.

Role of neocortex and limbic cortex

- *Limbic cortex and neocortex*, particularly in the frontal region, stimulate sex behaviour (partner-seeking behaviour). Therefore, its removal produces inhibition of sexual behaviour. In female animals, removal of neocortex and limbic cortex abolishes the excitement reaction during oestrous (heat period) without affecting the other aspects of heat.
- *Piriform cortex overlying amygdala* inhibits sex drive in males. Therefore, its destruction in male animals produces

hypersexuality. However, amygdaloid and periamygdaloid lesions do not produce hypersexuality in female animals. In human males also bilateral lesions of this area produce hypersexuality (*Klüver–Bucy syndrome).*

Role of hypothalamus. Anterior hypothalamus and median forebrain bundle stimulation elicits sex behaviour in males as well as females. A decorticate female animal will have regular oestrous cycle provided the hypothalamus is intact. Lesions of anterior hypothalamus abolish oestrous cycle in female animals and sexual interest in male animals (for further details see page 744).

Role of encephalization. In human beings, sex behaviour is largely encephalized, i.e. the perception that sexual act produces pleasure is a big cause of sex behaviour. Therefore, menopausal women (akin to castrated female animals) continue to have sex behaviour. Further, sex behaviour is strongly influenced, in human beings, by social customs, rules and social taboos.

(ii) Endocrinal control

Role of gonadal hormones

- *In males,* testosterone stimulates sex drive in the males. Castration (removal of testes) is associated with marked decrease in sex drive, which can be restored by injection of testosterone.
- *In females animals,* plasma oestrogen levels are raised during the oestrous period. In human females, sexual activity persists throughout the menstrual cycle, which is slightly increased during the time of ovulation. Castration in female animals (removal of ovaries) causes decline and eventual abolishment of sex drive.

Role of pheromones. Pheromones are chemicals which by their smell act as sex attractants in animals.

The role of pheromones in human sexuality is, however, uncertain.

4. Maternal behaviour

Maternal behaviour is the function of cingulate gyrus and retrosplenial portion of the limbic cortex. In animals, maternal behaviour is primarily neurogenic, i.e. it depends on the olfactory, auditory, visual and thermotactile stimuli arising from the young ones. Prolactin and oxytocin, though absolutely not essential, have been reported to facilitate maternal behaviour. In general, the maternal behaviour is concerned with the nursing (breastfeeding) and protection of the offspring by the mother.

5. Emotional behaviour

Emotional behaviour is one of the most important functions of limbic system. It has been discussed separately in 'physiology of emotions'.

6. Motivational behaviour

Motivational behaviour is also an important function of the limbic system and has been discussed separately as 'physiology of motivation'.

PHYSIOLOGY OF EMOTIONS

COMPONENTS OF EMOTIONS

Emotions refer to an aroused state involving intense feeling, autonomic activation and related behaviour, which accompany many of our conscious experiences. Emotions have two major components: mental and physical. The components of emotions are explained below by considering the example of response of an individual to sudden very loud noise.

I. Mental or sensory component

Mental or sensory component of emotions comprises cognition, affect and conation.

Cognition. It refers to a phenomenon by which one becomes aware (sees) and recognizes a situation. For example, when an individual hears a sudden very loud noise and from his experience recognizes it to be bomb blast. This is called cognition. Thus, mere seeing but not recognizing is not cognition.

Affect. It is a German word which means development of a feeling. In the above example, the person after cognizing the loud sound as bomb blast is frightened; this feeling of frightening is called affect.

Conation. It is the force which directs or urges to take some action. For example, the desire to run away from the site of loud noise after getting frightened is conation.

II. Physical or expressive or peripheral component

Physical or expressive or peripheral component of the emotions is motor side of emotional behaviour. It consists of two subcomponents—the somatic and autonomic.

Somatic part of the physical component of emotions basically comprises changes in the skeletal muscles. The accomplishment of the act of running away from the site of noise in the above example constitutes the somatic part of the physical component.

Autonomic part of the physical component of emotions involves the co-ordinated activity of sympathetic and parasympathetic nervous system. For example, occurrence of tachycardia, raised blood pressure, increased respiration rate, etc. after getting frightened from the sudden loud noise constitutes the autonomic part of the physical component.

- *Sympathetic expression.* Fear (as in the above example) is associated with sympathetic expression, which is characterized by an increase in the heart rate, increase in respiration rate, cutaneous vasoconstriction, sweating (cold sweat), piloerection, pupillary dilatation and dryness of mouth.
- *Parasympathetic expression* is noticed during grief or pleasure.

IMPORTANT NOTE

In many instances, the somatic part of the physical component of emotions may be absent. For example, after getting insulted and provoked one may beat the insulter (somatic part present). While the other individual may feel enraged, develop high blood pressure plus tachycardia but restrains oneself and does not show any somatic side of expression (indeed, this is common in civilized societies).

THEORIES OF GENESIS OF EMOTIONS

Physical changes are secondary to the emotional feelings or vice versa, have been the matter of debate. Following theories have been put forward from time to time in this regard, the genesis of emotions as explained by Arnold is as under:

Arnold theory. According to this theory of genesis emotions (Fig. 10.10-5):

- By *cognition,* one becomes aware and recognizes a situation.
- By *unconscious evaluation,* the situation is judged as to be harmful or beneficial.
- *Affect* is conscious reflection of unconscious appraisal. A feeling is thus generated consciously in response to unconscious evaluation of a situation. Such a feeling may be in the form of fear, joy, grief or rage.

Thus, according to the Arnold's theory, emotions have their own logic and that the peripheral component of emotions results from an unconscious evaluation of situation as potentially harmful or harmless. Therefore, in response to a particular situation the different individuals react differently, e.g. in response to a bomb blast by terrorist attacks:

– Some will be frightened,
– Proterrorist persons will have a feeling of joy,
– Antiterrorists will develop a feeling of rage and so on.

- Further, Arnold pointed out that the autonomic responses are not an essential component of emotions.

EMOTIONAL BEHAVIOUR

Different emotions produce different sets of behaviour. Behaviour is considered an expression of emotions. Some of the emotional behaviours are:

- Rage, fear and placidity (see page 744),
- Sexual behaviour (see page 744) and
- Feeling of reward and punishment (see page 744).

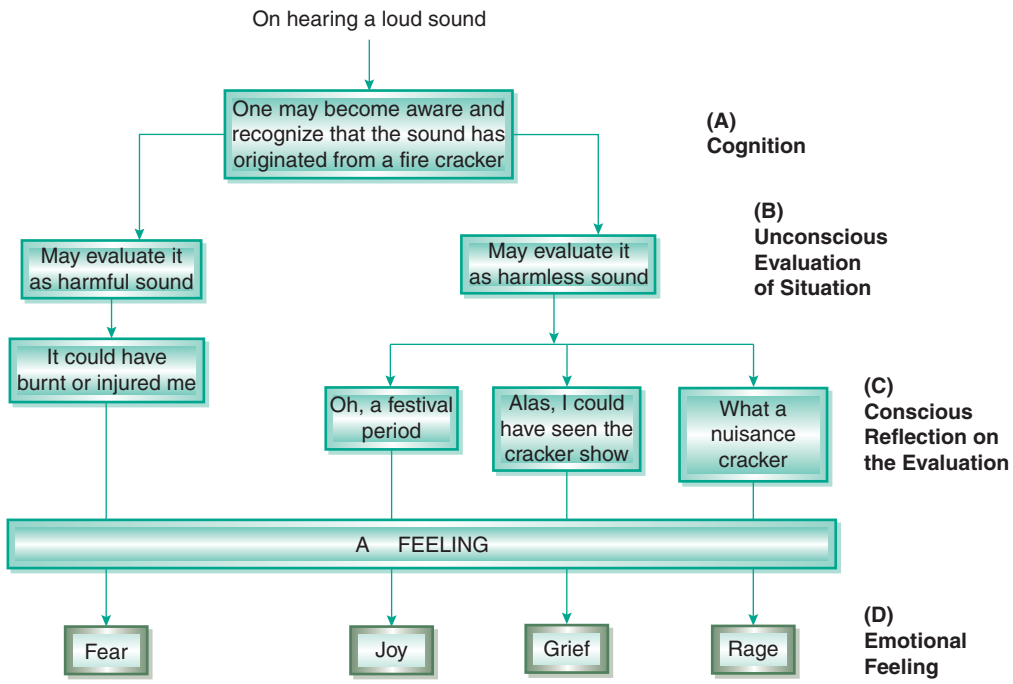


Fig. 10.10-5 Steps of Arnold's theory of emotions: A, cognition; B, unconscious evaluation of situation; C, conscious reflection on the evaluation and D, emotional feeling.

NEURAL SUBSTRATE OF EMOTIONS

1. Role of central nervous system

(i) Role of cerebral cortex

Cerebral cortex, especially the frontal, cingulate and parahippocampal cortices, play an important role in *affective component* of *emotions*:

- Detailed processing of conscious experience of emotional feeling occurs in the cerebral cortex.
- Cortical mechanisms also provide the means by which memory and imagination too can evoke emotional feeling.
- Cortex also provides the mechanisms that direct the motor responses to the external event during emotional behaviour, for example, to approach or avoid a situation.
- Cortical mechanisms also provide the means, which account for the modulation, direction, understanding or even inhibition of emotional behaviour. For example, once we know that an explosive sound came from only a fire cracker, the fear subsides by the cortical suppression of reflex emotional responses.

Limbic cortex acts as an association area for control of behaviour:

- *Anterior temporal cortex* has a gustatory and an olfactory association.
- *Parahippocampal gyrus* has a complex auditory association and a complex thought association derived from the Wernicke's area of the posterior temporal lobe.
- *Posterior cingulate cortex* has a sensory motor association.

Lesions of different parts of limbic cortex produce certain symptoms which suggest their functions, e.g.

- *Bilateral destruction of anterior temporal cortex* leads to the Klüver–Bucy syndrome (see page 854).
- *Bilateral lesions in the posterior orbitofrontal cortex* lead to insomnia and restlessness.
- *Bilateral destruction of anterior cingulate and subcallosal gyri* evokes an extreme rage reaction.

(ii) Role of hypothalamus

Hypothalamus has been considered the main seat of emotions. The hypothalamus along with limbic structures is concerned with affective nature of the sensory impulses. For details see page 809.

Areas of hypothalamus *associated with behavioural control* functions are:

- Increased level of general activity, leading to *rage and aggression.* It occurs when lateral hypothalamus is stimulated.
- *Sexual arousal* occurs when most anterior and posterior portion of hypothalamus is stimulated.
- *Feeling of reward, tranquility and pleasure* are appreciated when reward centre is stimulated.
- *Fearing, feeling of punishment and aversion* are felt when punishment centre is stimulated.

Lesions of hypothalamus are associated with:

- Extreme passivity and loss of drive,
- Excessive eating and drinking and
- Rage and violent behaviour.

(iii) Role of amygdala

Amygdala is a large aggregate of cells located above the inferior horn of the lateral ventricle and is embedded in the uncus. It consists of two subdivisions: a corticomedial nuclear group and a basolateral group of nuclei. In human beings, basolateral nuclei of amygdala are very well developed and they play an important role in behavioural activities, not generally associated with olfactory stimuli.

Afferents to the amygdala come from all portions of limbic cortex as well as from neocortex and, therefore, it is called the 'window' through which the limbic system sees the place of the person in the world.

Efferents from the amygdala are varied and extensive, reaching the cortex, hippocampus, septum, thalamus and hypothalamus.

As far as emotions are concerned amygdala co-ordinates the affective component of emotions (function of cerebral cortex) with the autonomic response to emotions (function of hypothalamus).

Affective component of emotions is influenced by amygdala through the *ventral amygdalofugal pathway* that projects from the central nucleus of amygdala to the brain stem, the dorsal medial nucleus of thalamus, and the association areas of cortex, especially the rostral cingulate gyrus of the cortex and the orbitofrontal cortex.

Peripheral component of emotions is influenced by amygdala through the *stria terminalis* that projects from the central nucleus of amygdala to:

- *Hypothalamus*, which mediates neuroendocrinal response to fearful and stressful stimuli, and
- *Nucleus accumbens* that controls the body language in the emotional states.

Concepts of extended Papez circuit. Recently, a concept of extended Papez circuit has been described in which the focus has been shifted to the main role of amygdala in emotions (Fig. 10.10-6).

Bilateral destruction of temporal pole is associated with destruction of amygdala, which leads to the *Klüver–Bucy syndrome* characterized by:

- Extreme orality, i.e. excessive tendency to examine objects orally,
- Loss of fear,
- Decreased aggressiveness,
- Tameness,
- Changes in eating behaviour,
- Psychic blindness and
- Excessive sexual drive.

(iv) Role of hippocampus

Hippocampus is formed due to the projection of the hippocampal sulcus into the floor of the inferior horn of lateral ventricle.

Connections. Hippocampus has many indirect connections to many portions of the cerebral cortex.

Stimulation of hippocampus can evoke rage, passivity and excessive sexual drive. It is also hyperexcitable and weak stimuli can produce epileptic seizures.

Functions of hippocampus are:

- Like amygdala, it is an additional channel through which incoming signals can lead to appropriate behavioural pattern.
- It is suggested that hippocampus also provides the signal for memory consolidation, e.g. the transformation from short-term to long-term memories of verbal and symbolic type.

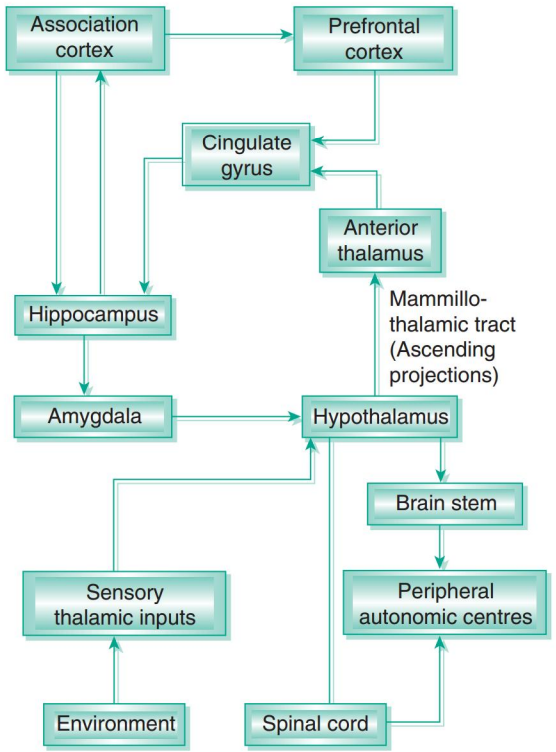


Fig. 10.10-6 The extended Papez circuit of emotions.

Lesions of the hippocampus lead to *antegrade amnesia,* i.e. a profound inability to form new memories based on any type of verbal symbolism (language).

2. Role of peripheral nervous system

Autonomic as well as the somatic motor peripheral nervous system is involved in the peripheral expression of the emotions.

(i) Autonomic nervous system is the chief mediator of the emotional output. Cannon described the emotional response to an emergency or involving diffuse sympathetic activation and outpouring into the blood of excitatory substances, i.e. the catecholamines.

(ii) Somatic motor nervous system is involved in the somatic part of the physical component of emotions, which basically comprises changes in the skeletal muscles.

PHYSIOLOGY OF MOTIVATION

Motivation is that component of behaviour, which is responsible for accomplishing a particular task.

NEURAL MECHANISM OF MOTIVATION

Neural mechanisms involved in the motivation are based on the concept of reward and punishment. It has been elucidated by the effects of self-stimulation of brain.

CONCEPT OF REWARD AND PUNISHMENT

Almost anything that we do is related in some way to reward and punishment. If we do something that is rewarding, we continue to do it. If we do something that is punishing we cease to do it.

Reward and punishment centre. It appears that conditions which result in stimulation of reward centre produce motive to do a job. For details about *reward and punishment* centres see page 744.

Experiment to demonstrate activity of reward and punishment centres. This experiment demonstrates that the animal itself regulates the stimulation of reward and punishment centres (self-stimulation). The *Hess technique* of experiment is: a rat is placed in a cage having a bar (lever) that can be pressed by the rat. The electrical connections are made in such a way that on pressing the bar a stimulus may be applied to the brain of rat through an electrode (Fig. 10.10-7). Following observations are made:

- *When no stimulus is applied* to the brain on pressing the bar, the animal presses it occasionally (at random).

- *When stimulation is applied to the reward centre* (located along the course of medial forebrain bundle, especially in lateral and ventromedial nucleus of hypothalamus) on pressing the bar, the animal presses the bar repeatedly at a rate much above the rate of random pressing. It has been reported that stimulation of reward centre produces a *pleasure sensation* (feeling of complete relaxation).
- Further, it has been observed that even if a painful barricade is placed between the rat and bar (lever), the rat ignores the pain to cross the barrier to press the bar. This means that the rat must have developed a strong motive to derive pleasure sensation.
- Instead of painful barrier, if a complicated maze is made between the rat and bar (lever), the rat learns to cross the maze. This means that development of a motive is a strong factor for learning. From this, it can be concluded that if there is no motive to do a job, people will not do the job nor they will learn to do the job.
- When stimulation is applied to the punishment centre (located in medial hypothalamus, periventricular zone) on pressing the bar, the animal avoids further stimulation of this area. The pressing rate of bar is decreased much below the rate of random pressing. It has been reported that electrical stimulation of punishment centre leads to pain, fear, defence, escape reactions and the other elements of punishment.

ROLE OF NEUROTRANSMITTERS

Neurotransmitters involved in the pathway that stimulate the *reward centre* are:

- Catecholamines (norepinephrine and dopamine),
- Morphine and
- Enkephalin.

Drugs that increase stimulation of reward centre are those which increase synaptic activity in catecholamine pathway, e.g.

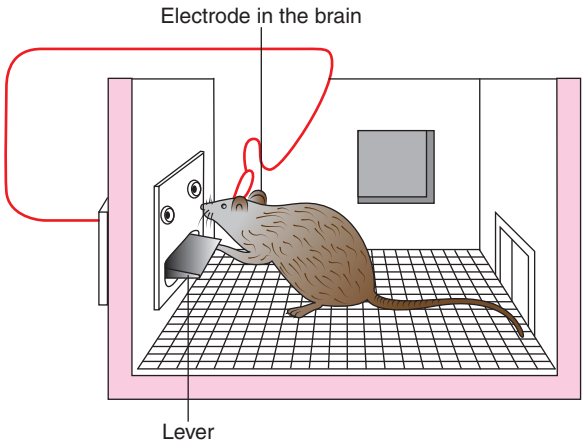


Fig. 10.10-7 Experimental set-up to demonstrate reward and punishment centres by self-stimulation.

- *Amphetamine*, which causes increased release of dopamine,
- *Nicotine and alcohol* increase the amount of dopamine,
- *Cocaine* inhibits the reuptake of dopamine and norepinephrine.

APPLIED ASPECTS

- *Addiction.* Addiction is repeated and compulsive use of a substance. Tobacco, alcohol, cannabis, opiates, LSD, cocaine, and amphetamine are well known to produce addiction. These drugs act by increasing dopaminergic activity in the reward centre, particularly nucleus accumbens. Thus, a strong motive develops to use them again and again.
- *Learning.* Catecholamines and enkephalins are also involved in the pathways responsible for learning. Therefore it seems that reward and punishment constitute the incentives for learning.

Drugs that decrease stimulation of reward centre are those which lower synaptic activity in the catecholamine pathway, e.g. chlorpromazine hydrochloride (Largactil).

PHYSIOLOGICAL BASIS OF PSYCHOTIC DISORDERS

1. DEPRESSION

In a normal person the mood usually swings, i.e. with bad news (e.g. failure in examination) the mood is down and with good news (e.g. distinction in examination) the mood is elated. However, when mood chronically remains down without any specific reason then the condition is called as depression.

Signs and symptoms are:

- Chronic depression of mood,
- Lack of interest,
- Suicidal tendency and
- Excessive sleep and overeating.

Causes. The physiological basis of this disorder is decreased activity of either noradrenergic or serotoninergic fibres. The defect may be at the receptor level or there is deficiency of neurotransmitters (noradrenaline, NA or serotonin).

Treatment. Drugs that increase the excitatory effects of NA are effective in treating depression, these include monoamine oxidase inhibitors, tricyclic antidepressants and drugs that enhance the action of serotonin. Manic-depressive conditions

(bipolar disorder) can be effectively treated by lithium compound that diminish the actions of NA and serotonin.

2. MANIA

In this condition, mood remains chronically elated without any specific reason. It is due to overactivity of noradrenergic fibre activity.

3. SCHIZOPHRENIA

Schizophrenia is another common psychotic disorder in which there is false perception of sensations (hallucinations) though there is no anatomical lesion in the sensory pathway.

Cause. Schizophrenia is thought to be associated with the excessive activity of dopaminergic mesolimbic pathway. Evidence supporting this theory derives from the fact that schizophrenic symptoms are reduced by drugs, such as chlorpromazine and haloperidol that diminish dopamine release at axon terminals.

Characteristic features of schizophrenia are:

- Hallucinations, auditory as well as visual,
- Delusions of grandeur, intense fear, or paranoia and
- Withdrawal from the society, i.e. patient prefers extreme isolation, avoids company with persons and has no interest in the surroundings.

Treatment. As mentioned above, the drugs which decrease the dopamine concentration in the central nervous system are used. But the main drawback of these drugs is that they cause deficiency of dopamine which precipitates parkinsonism.

Chapter 10.11

Reticular Formation, Electrical Activity of the Brain, and Alert Behaviour and Sleep

RETICULAR FORMATION AND RETICULAR ACTIVATING SYSTEM

- **Neuronal aggregates of recticular formation**
  - Reticular nuclei
  - Functional neuronal aggregates
- **Reticular pathways**
  - Cortico-reticulo-spinal pathways
  - Cortico-reticulo-cerebellar connections
  - Visceral control pathways
  - Reticular activating system
- **Functions of reticular formation**

ELECTRICAL ACTIVITY OF THE BRAIN

- **Evoked cortical potentials**
  - Types of evoked potentials
  - Clinical uses of evoked potentials
- **Electroencephalogram**
  - Normal EEG
  - Neurophysiological basis of EEG
  - Abnormal EEG waveforms

WAKEFULNESS AND SLEEP

- **Wakefulness**
  - Neural substrate for wakefulness
  - Chemical mediators of wakefulness
- **Sleep**
  - Sleep–wake cycle and factors affecting sleep
  - Types and stages of sleep
    - Non-REM
    - REM sleep
  - Sleep cycle
  - Genesis of sleep
  - Physiological significance of sleep
  - Sleep disorders

RETICULAR FORMATION AND RETICULAR ACTIVATING SYSTEM

Reticular formation (RF) refers to the complex network of neurons and nerve fibres, which occupy midventral portion of brain stem around the central cavity and is exclusive of the specific nuclei and tracts. The brain stem RF can be considered to comprise medullary RF, pontine RF and mid brain RF. Structurally, brain stem reticular formation consists of:

- Neuronal aggregates,
- Afferent connections and
- Efferent connections.

] Reticular pathway

NEURONAL AGGREGATES OF RETICULAR FORMATION

Reticular nuclei

A number of reticular nuclei have been described. These can be divided into three longitudinal columns (in each half of the brain stem):

Nuclei of median column lie next to middle line and are called *nuclei of raphe*, e.g. raphe nuclei in the mid brain.

Nuclei of medial column lie lateral to nuclei of median column. These are made of large cells and so also called *magnocellular nuclei*, e.g. nucleus gigantomedullaris in the medulla and pontine tegmental nuclei.

Nuclei of lateral column lie lateral to nuclei of medial column. These are made of small neurons and so also called *parvocellular nuclei*. Examples of such nuclei are central nucleus of medulla and central nucleus of pons.

Functional neuronal aggregates

Functional neuronal aggregates, though not anatomical entities, have been described to have fairly well-defined physiological functions. These include:

- Cardiac centres,
- Respiratory centres,
- Vasomotor centres,
- Salivatory centres and
- Chemoreceptor neurons.

RETICULAR PATHWAYS

Connections of RF are:

- Afferent connections

- Efferent connections, which include: Descending projections and Ascending projections.

The afferent and efferent connections of the RF form several pathways. Some of the major pathways are:

- Cortico-reticulo-spinal pathways,
- Cortico-reticulo-cerebellar connections,
- Visceral control pathways and
- Reticular activating system (RAS).

Cortico-reticulo-spinal pathways

Afferents of this pathway to neurons of RF come from the motor and other areas of cerebral cortex.

Areas of reticular formation which receive impulses from the cerebral cortex are:

- Bulboreticular inhibitory area located in the lower part of the medulla (see page 828) and
- Bulboreticular facilitatory area located in the pons (see page 828).

Cortico-reticulo-cerebellar and cortico-reticulo-basal ganglia connections

Some afferents from the cerebral cortex, after relay in reticular formation project to cerebellum and basal ganglia. The influence of cerebellum and basal ganglia on motor function has been described in Chapter 10.9.

Visceral control pathways

Certain centres in the reticular formation regulate respiration, heart rate and blood pressure. These effects are mediated through connections between the reticular formation and autonomic centres in the brain stem and spinal cord, but the pathways concerned are not well defined.

Reticular activating system

Reticular activating system, also known as ascending reticular activating system, is a complex polysynaptic pathway that projects diffusely from the brain stem reticular formation to the cerebral cortex. ***Collaterals*** to RAS funnel from the following sources (Fig. 10.11-1):

- Long ascending sensory pathways, such as spinothalamic tracts, are the important sources of collaterals to RAS. The fibres of the tracts conveying slow pain send the richest collateral connections to the RAS.
- In addition to long ascending sensory tracts, collateral to the RAS also funnel from the trigeminal, auditory, visual and olfactory pathway systems.

Efferent projections from RAS are:

- Majority of RAS fibres end in non-specific thalamic nuclei (intra-laminar and midline nuclei) and from there projected diffusely and non-specifically to the whole neocortex (Fig. 10.11-1).

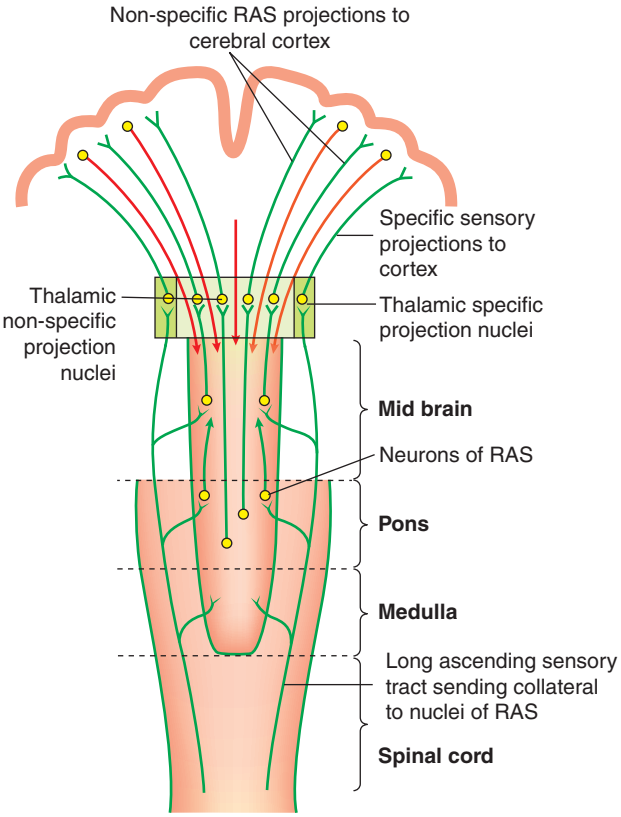


Fig. 10.11-1 Diagrammatic depiction of reticular activating system (RAS) vis-a-vis specific sensory projections.

- Another part of RAS bypasses the thalamus to project diffusely to the cortex.

The RAS fibres occupy the core portion of the brain stem (Fig. 10.11-1). Whereas the specific fibres occupy rather the lateral parts of brain stem.

Stimulation of RAS. The reticular activating system is stimulated by impulses funneled into it through the collateral described above. Thus, the RAS is a *non-specific system,* which can be excited by any sensation. Whereas, the classic sensory pathways are specific in that the fibres in them are activated by only one type of sensory stimulation.

FUNCTIONS OF RETICULAR FORMATION

1. Sleep-wakefulness. The RAS of reticular formation is the neural substrate of the consciousness and sleep waking cycle.

- Reticular activating system sends a strong facilitatory drive to the central neurons, raising their background excitability and increasing their responsiveness to specific stimuli. Thus, when RAS is stimulated, there is wakefulness and alertness of the subject and the subject becomes fully conscious. This alertness is necessary even for proper sense perception. Conversely, when the RAS is inhibited, the subject is asleep.

- *Lesions of RAS* in experimental animals produce interminable sleep and coma. In human beings also, lesions in the RAS (e.g. tumours) cause prolonged sleep.
- *Many agents producing sedation*, hypnosis and anaesthesia (e.g. benzodiazepine and barbiturates) act by preventing synaptic transmission in RAS and thus inhibiting the RAS.

2. Selective attention and sensory inattention. The reticular formation is also responsible for selective attention and sensory inattention, through the corticofugal control of sensory input and due to habituation.

3. Conditioning and learning. Reticular formation is an integral part of the neural substrate for conditioning and learning (for details see page 75).

4. Control of muscle tone and regulation of postural reflex changes. Reticular formation modulates the tone of extensor (antigravity) muscles. The pontine (medial) reticulospinal tract has an excitatory and the medullary (lateral) reticulospinal tract has an inhibitory influence on the extensor muscle tone (for details see page 822).

5. Autonomic functions. The visceral regulating centres are an integral part of the reticular formation. The influence of higher neurons over the viscera and autonomic functions are mediated through the visceral centres in reticular formation.

6. Modulation of pain. Serotonergic neurons of the modulatory raphe nuclei form a part of the endogenous pain relief system. By affecting the transmission of pain impulses through the substantia gelatinosa of the spinal cord, these neurons modulate the perception of pain (for details see page 809).

7. Control of neuroendocrine system. The reticular formation projections play a role in the control of neuroendocrine systems in the hypothalamus.

ELECTRICAL ACTIVITY OF THE BRAIN

EVOKED CORTICAL POTENTIALS

Evoked potential refers to the surface electrical activity recorded from the surface of the scalp in response to a specific and adequate stimulus—auditory, visual or somatosensory. Stimulation by a specific adequate stimulus produces two types of electrical activity in the cerebral cortex known as primary evoked potential and diffuse secondary response.

Primary evoked potential. This is the initial, brief (lasting for few milliseconds) and localized response over the specific sensory cortex. For example in the foot area of the post-central gyrus, if electrical shock is given over the foot or over the occipital lobe after photopic stimulation, the primary evoked potential is characterized by (Fig. 10.11-2).

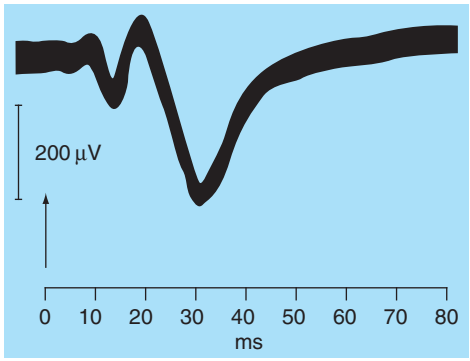


Fig. 10.11-2 Response evoked in contralateral sensory cortex by stimulation of sciatic nerve in a cat. The upward deflection is surface negative.

- Latency of about 5–12 ms (average 10 ms).
- First there appears a surface positive wave which is followed by a small negative wave.
- The primary evoked potential is highly specific in its location and can be observed only where the pathway from a particular sensory organ ends; that is, it is produced by the conduction of sensory signals through the specific sensory pathways.

Diffuse secondary response is characterized by:

- Latency of about 20–80 ms (average 50 ms), i.e. it appears about 50 ms of sensory stimulus.
- Positive–negative wave sequence of secondary diffuse response is frequently larger and more prolonged than the positive–negative wave sequence of primary evoked potential.
- The surface-positive diffuse secondary response unlike the primary is not highly localized. It can be recorded at the same time from most of the cerebral cortex. It is due to spread of impulses through the RAS to the cerebral cortex.

TYPES OF EVOKED POTENTIALS

Depending on the type of stimulus, the evoked potential can be:

- Visual evoked potential (VEP),
- Brain stem auditory evoked potential (BAEP),
- Somatosensory evoked potential and so on.

Depending upon the latency of response, evoked potential can be classified into:

- Stimulus-related potentials (short, mid and long latency) and
- Event-related or endogenous potentials.

Stimulus-related potentials

Stimulus-related potential refers to a series of waves that relates to the sensory modality. For example, the auditory

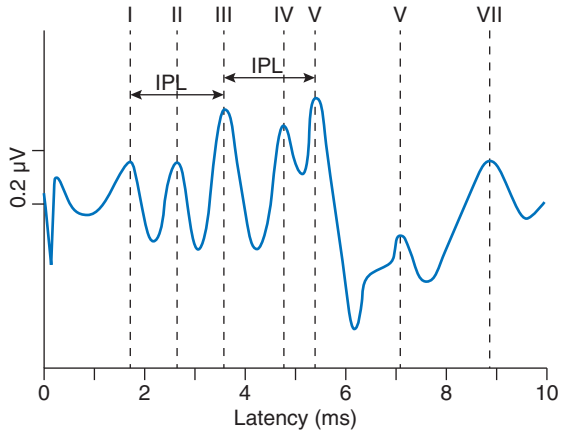


Fig. 10.11-3 Early latency response to an auditory stimulus of 60 dB at 10/s (brain stem auditory evoked potential).

stimulus-related response has been divided into three sequential time periods:

- Early latency response,
- Mid-latency response and
- Long latency response.

1. Early latency response (ELR) to an auditory stimulus is characterized by a latency of <10 ms and is named as 'BAEP'.

Waves of ELR. Early latency response consists of a series of waves named I–VII (Fig. 10.11-3).

- *Wave I* represents the volume conducted electrical activity from the auditory nerve,
- *Wave II* from the pons and
- *Wave III* from the mid brain.

Interpeak latency indirectly reflects neural conduction in the corresponding segment of the central auditory pathway.

2. Mid-latency response to an auditory stimulus is characterized by a latency of 10–50 ms. It is considered to represent the electrical activity arising in the thalamocortical radiations, the primary auditory cortex and the early association cortex.

3. Long latency response (LLR) to an auditory stimulus is characterized by a latency of more than 50 ms. It is negative–positive complex comprising:

- A large negative wave (N_1) and
- A large positive wave (P_2).

The visual and somatosensory stimuli too elicit a similar response. All these responses are stimulus evoked and reflect the functional integrity of the sensory pathways in the CNS. Unlike the event-related potentials (discussed below), they are largely independent of the subject's attention or level of arousal.

Event-related potentials

Event-related potentials (ERPs) are dependent upon the subject's attention or level of arousal (cf stimulus-related potential, as described above). The ERPs are elicited only when the subject is required to distinguish one stimulus (the target) from the other (the non-targets). Thus, ERPs are related to the cognitive events associated with the distinction of target from non-target stimuli. Long latency response in event-related potentials is also a negative–positive complex comprising:

- A negative wave (N_2) and
- A positive wave (P_3).

CLINICAL USES OF EVOKED POTENTIALS

Stimulus-related evoked potentials reflect the functional integrity of the sensory pathways from the receptor to the cortex. Therefore, any delay in conduction as depicted by delayed peak or interpeak latencies would be of diagnostic value. The lesions interrupting the conduction pathways in patients with optic neuritis due to multiple sclerosis (a demyelinating disorder), the abnormal VEP is diagnostic.

Event-related evoked potentials are related to cognitive behaviour. Therefore, use of ERPs in the clinical assessment of dementia and delirium is fairly well established by now. *Dementia* refers to an abnormal deterioration of intellect affecting several areas of cognitive functions, such as abstraction, orientation, judgement and memory.

ELECTROENCEPHALOGRAM

The term electroencephalogram (EEG) (introduced by the German psychiatrist Hans Berger) refers to the record of spontaneous electrical activity of the brain taken from the surface of scalp (cf evoked potential which is the surface electrical activity recorded in response to a specific and adequate stimulus). The spontaneous electrical activity of the brain is largely due to graded or summated post-synaptic potentials in the many hundreds or thousands of brain neurons that underlie the recording electrode at the surface of scalp. The electrical activity of the brain can also be recorded from the pial surface of the brain cortex after opening the skull (e.g. during brain surgery). The term *electrocorticogram* (ECoG) is used to denote such a record.

NORMAL ELECTROENCEPHALOGRAM

The EEG consists of waves which are oscillations in the electrical potential of brain having following characteristics:

- The oscillations differ in the frequency and amplitude at different points on the scalp and during different stages of mental alertness.

Table 10.11-1 Classification of brain waves depending on frequency of oscillations

Frequency (Hz)	Type of EEG wave	Amplitude (μV)
1–4	Delta (δ)	20–200
4–7	Theta (θ)	10
8–13	Alpha (α)	50
14–30	Beta (β)	5–10

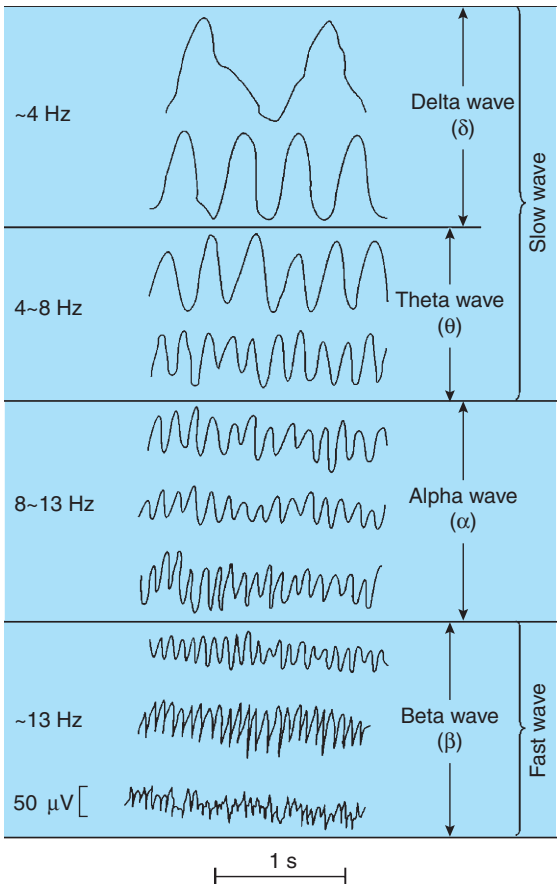


Fig. 10.11-4 Different types of normal EEG waves.

- *Frequencies* of brain waves range from 1 cycle/s to over 50 cycles/s.
- *Amplitude* of brain waves may vary from 50 to 200 μV.
- Much of the time brain waves are *irregular* and no general pattern is obtained. At other times distinct patterns do appear.
- Different waves recorded in a normal person, depending on their frequency, are classified as alpha, beta, theta, and delta (α, β, θ, δ) waves (Table 10.11-1, Fig. 10.11-4).

Waves of EEG (Fig. 10.11-4)

Alpha waves

These are the most prominent component of EEG obtained from adult humans who are awake but quiet and at rest with the eyes closed. Alpha waves are said to result from spontaneous activity of non-specific thalamocortical system.

Characteristic features of α waves are:

- *Frequency* of α waves varies from 8 to 13 Hz.
- *Amplitude* of these waves slowly waxes and wanes, but the average amplitude is about 50 μV.
- *Location*. Alpha waves are most marked in the parieto-occipital area of the scalp, though these are observed sometimes from other locations as well.
- *Disappear* during sleep.

Causes of decreased frequency of α waves are:

- Old age, due to decreased cerebral perfusion leading to decreased cerebral metabolism,
- Low blood glucose level,
- Low body temperature,
- Low levels of adrenal glucocorticoids,
- High arterial partial pressure of CO_2 and
- Sleep.

Causes of increased frequency of a waves are:

- High blood glucose level,
- Rise in body temperature,
- Low arterial pCO_2,
- High levels of adrenal glucocorticoids and
- Alerting states.

Alpha block

Alpha block or alpha attenuation refers to a phenomenon in which α wave attenuates and are replaced by the fast, irregular waves of low amplitude. Alpha block occurs when:

- The persons open their eyes,
- When the individuals engage in conscious mental activity, such as doing mathematical calculations and
- When any form of stimulation is applied (Fig. 10.11-5). The term *aroused* or *alerting response* is also used to denote α block, since it is correlated with arousal or alerting response.

The term desynchronization has also been suggested for α block because it represents breaking up of the obviously synchronized neural activity necessary to produce regular α waves. However, the term desynchronization is misleading as the fast EEG activity seen in alert state is also synchronized, but at a higher rate.

Beta waves

- *Frequency* of β waves is usually between 14 and 30 cycles/s.
- *Amplitude* (voltage) of β waves (5–10 μV) much lower than the α waves (50 μV).
- *Location*. They are frequently recorded from the parietal and frontal region.
- *Seen under* following conditions:
  - Tension and CNS activation,

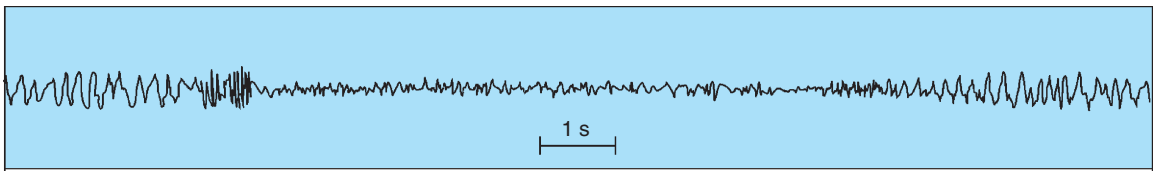


Fig. 10.11-5 Electroencephalography depicting α block produced by olfactory stimulus in a rabbit.

- Arousal response (or α block),
- Infants have fast β-like activity in EEG and occipital rhythm is slow 0.5–2/s
- Barbiturates induce β activity typically at a frequency of 18–24 Hz.

Theta waves

Theta waves usually do not occur in normal waking individual (except in newborn infants). Theta component persists in adult life in 10–15% of normal subjects. Usually, θ waves are seen under following conditions:

- Emotional stress in adults particularly during disappointment and frustration.
- Many brain disorders.
- Theta component of EEG often accentuates during crying in children.

Characteristic features of θ waves are:

- *Frequency* is between 4 and 7 Hz.
- *Amplitude* (10 μV) is slightly larger than the α waves.
- *Location.* They are recorded from the temporal and parietal region in children.

Delta waves

Delta waves do not occur in normal waking individuals. These are seen in following conditions:

- Deep sleep (stage III and IV of non-REM sleep),
- Infancy and
- Serious brain damage.

Characteristic features (Fig. 10.11-4) are:

- *Frequency* is less than 4 Hz,
- *Amplitude* is very high (20–200 μV),
- *Can be produced* by overbreathing,
- *Occur strictly* in the cortex independent of activities in lower regions of the brain, therefore they occur in sleep when cortex is released from the activating influence of lower centres.

NEUROPHYSIOLOGICAL BASIS OF EEG

The neurophysiological basis of EEG has not been fully elucidated. Some of the important points in this regard are:

Cortical grey matter along with its thalamic connection plays an important role in the EEG. Largely, the activity recorded in the EEG is that of rhythmically discharging cell bodies in the most superficial layers of cortical grey matter. Thalamus discharge synchronizes this activity.

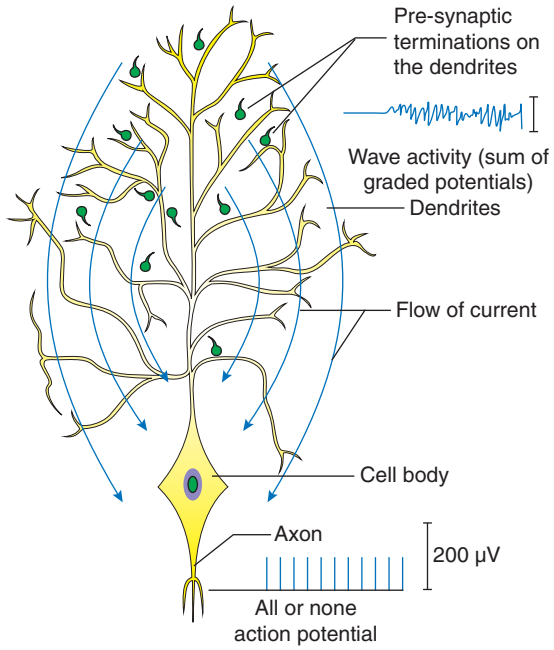


Fig. 10.11-6 Electrical activity recorded from vertically oriented dendritic tree of pyramidal cells in the cerebral cortex compared to that recorded from an axon.

Current flow in the fluctuating dipoles formed by the cell bodies and dendrites of the cortical cells accounts for the potential changes recorded as EEG. The dendrites are the sites of non-propagated de-polarizing and hyperpolarizing local potential changes. The cell dendrites relationship is, therefore, that of a constantly shifting dipole. Hence they become sites of current sink. The dense dendritic tree is present in a particular (vertical) orientation and this results in the brain wave patterns (Fig. 10.11-6). In general, when the sum of the dendritic activity is positive relative to the cell, the cell is hyperpolarized and less excitable, when it is negative the cell is depolarized and hyperexcitable. Thus the EEG is due to graded potentials which are summated post-synaptic potentials in the brain neurons (Fig. 10.11-6).

Synchronizing mechanisms. Synchronizing activity of neighbouring cells and rhythmic discharge from the thalamus are responsible for synchronizing mechanism.

1. Synchronizing activity of neighbouring cells is due to:

- The effect of parallel neural processes on each other in a volume conductor (Fig. 10.11-7).
- The interconnection of neurons by inhibitory pathway.

2. Rhythmic discharge from the thalamus also responsible for synchronization of EEG waves is evident from the following observations:

- Stimulation of certain thalamic nuclei of the frequency of about 8 cycles/s produces on EEG record with a similar frequency in greater part of the ipsilateral cerebral cortex. The amplitude of the waves also waxes and wanes, i.e. α rhythm is produced.
- Large lesions of the thalamus produce disturbances in the synchronized activity of the EEG on the side of lesion.

Desynchronizing mechanisms. Desynchronization, as mentioned earlier in α block, refers to the replacement of a rhythmic EEG pattern with irregular low-voltage activity (arousal reaction). It occurs due to sensory stimulation of RAS as is evident from following observations:

- Stimulation of specific sensory system up to mid brain only (up to which reticular formation is present) produces desynchronization and the stimulation of these systems above the mid brain do not produce desynchronization.
- Large lesions of the mid brain that interrupt the medial lemnisci and other ascending specific sensory systems fail to prevent the desynchronization produced by the specific sensory stimulation below the mid brain level.
- High-frequency stimulation of reticular formation in the mid brain feature like that of non-specific projection nuclei of the thalamus that produc es desynchronization and arousal in sleeping animals.

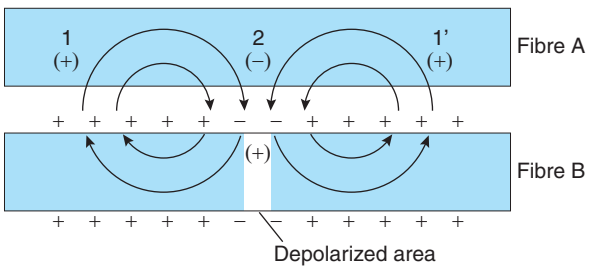


Fig. 10.11-7 The electrical property of two parallel placed nerve fibres. Note that current flows into the depolarized area of active fibre B from the surrounding membrane as an impulse passes along the nerve. At points 1 and 1′ on the membrane of inactive fibre A positive charges build up. Thus the membrane becomes slightly hypopolarized in these two regions. At point 2 of the fibre A positive charges are removed, so the membrane undergoes a slight depolarization at this point.

- Lesions in the mid-brain tegmentum that disrupt the RAS without damaging the specific systems are associated with asynchronized EEG pattern that is unaffected by the sensory stimulation.

Variations in the EEG wave formation with age

The EEG waveforms at rest in humans vary with age as:

- *In infants* (up to 1 year of age), the occipital rhythm is slow (0.5–2 Hz) than those of adults (8–13 Hz).
- *In children*, the occipital rhythm speeds up and the adult α pattern gradually appears during adolescence.
  - After 15 years of age, the EEG waveforms become almost the same as those of adults.

ABNORMAL EEG WAVEFORMS

1. Epilepsy

Electroencephalographic inspection is indispensable to the diagnosis of epilepsy. The waveforms of epilepsy (Fig. 10.11-8) include idiopathic abnormal waves, such as a spike, sharp wave and spike and slow wave complex. Between these abnormal waves, an irregular slow wave appears and the background waveform is disturbed.

Types of epilepsy

Grandmal epilepsy is a serious fit accompanied by convulsions with tonic muscle contractions, clonic jerks and

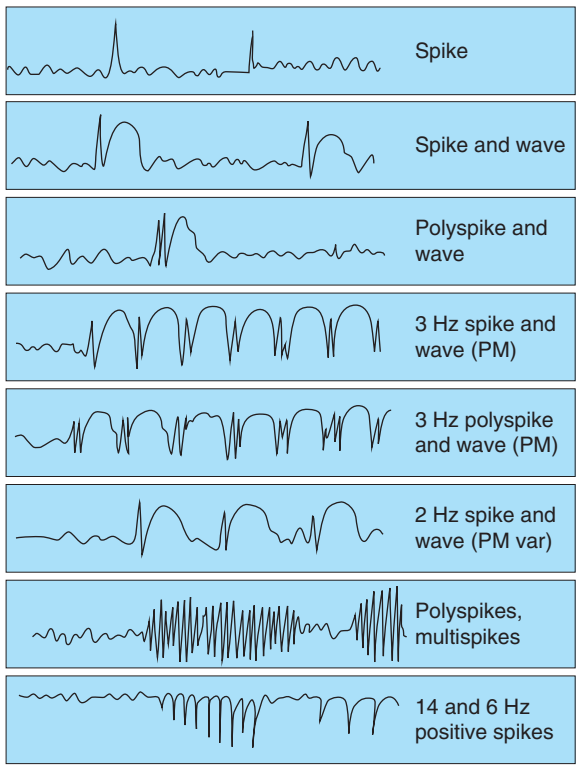


Fig. 10.11-8 EEG waveform of epilepsy. (PM = Petit mal.)

loss of consciousness. The EEG wave form shows continuous spike or sharp wave.

Petitmal epilepsy is characterized by a sudden loss of consciousness lasting only for few seconds, convulsive movements are absent but sometimes slight localized twitchings occur. EEG wave form generally shows spike slow wave complex of **3 Hz** (Fig. 10.11-8).

Psychomotor epilepsy. In this form of epilepsy, typical fit and loss of consciousness does not occur, but there are inappropriate movements accompanied by hallucinations. There is no typical change in the EEG waveform.

Activation

Activation is made to detect latent epilepsy, which cannot be detected by ordinary EEG recording. A flash stimulus, hyperventilation, sleep or drug administration are used to induce an epileptic attack.

2. Consciousness dysfunction

A slow wave appears in the case of consciousness dysfunction.

Disturbances of consciousness include coma, syncope and stupor.

Coma. Coma refers to a permanent state of sleep which is characterized by a loss of consciousness from which arousal cannot be elicited. It is produced by lesions blocking the connection between the ascending RAS and the thalamus.

During state of coma, stimulation of sensory pathways can cause a momentary desynchronization of the EEG but does not produce any behavioural changes.

Syncope (fainting) refers to a transient pathologic loss of consciousness.

Stupor is the more persistent loss of consciousness from which arousal can be obtained.

3. Organic brain dysfunction

An abnormal wave appears when in brain functional trouble occurs due to cerebral tumour, brain blood vessel trouble (bleeding, clogging, artery/vein leakage, hardened brain artery, etc.) or brain injury caused by an external wound of the head. In the case of brain tumour, for example, no waveform is generated from the tumour part, but a slow wave is generated from the surrounding organization. The EEG waveform shows the slow wave.

4. Brain death

Criteria for labelling brain death are important because of the desire to obtain organs for transplant operations and the desire to remove heroic life-support system. Individual is declared dead when brain cells stop activity, the EEG waveform becomes flat in all channels and finally disappears.

WAKEFULNESS AND SLEEP

WAKEFULNESS

The RAS of the reticular formation is the neural substrate of the consciousness and sleep-waking cycle. As described earlier, RAS is a complex polysynaptic pathway that projects diffusely from the brain stem reticular formation to the cerebral cortex both directly as well as via thalamus (for details see page 858).

NEURAL SUBSTRATE FOR WAKEFULNESS

In addition to the projections from the RAS, the wakefulness and consciousness are maintained by a continuous sensory input to the cortex from visceral as well as somatic systems via the non-specific thalamic system, subthalamus, hypothalamus and basal forebrain. This is proved by cerveau isole (transections separating the cerebrum, from the brain stem and spinal cord) which produces a sleep-like state with cortical slow waves. The neural substrate of wakefulness generating systems is described briefly:

1. Reticular activating system, as discussed above, is mainly responsible for the tonic maintenance of the cortical activation and behavioural arousal of the wakefulness. The cortical activation and behavioural arousal are controlled by tegmentum of mid brain.

2. Thalamus. The non-specific thalamic system formed by the ventromedial, intra-laminar and midline nuclei are involved in the activation of entire cerebral cortex. These nuclei get tonic drive from the reticular formation and in turn project diffusely to the cerebral cortex.

3. Hypothalamus and subthalamus. The ascending impulses from mid-brain reticular formation also relay to cerebral cortex through the posterior hypothalamus and subthalamus. In addition, the posterior hypothalamus also acts as waking centre, as its stimulation causes wakefulness. Conversely, lesions of posterior hypothalamus result in coma.

4. Basal forebrain. It comprises nucleus basalis of Meynert (substantia innominata), nuclei of the diagonal band and septum. The basal forebrain receives impulses from the reticular formation and in turn project to the cerebral cortex and is responsible for the cortical activation of wakefulness.

CHEMICAL MEDIATORS OF WAKEFULNESS

Chemical mediators of wakefulness include:

- Neurotransmitters,
- Cerebrospinal fluid (CSF)-borne peptides and
- Blood-borne peptides.

Neurotransmitters

1. Catecholamines. Norepinephrine neurons of the locus coeruleus and brain stem, which project diffusely to the forebrain, including the cortex, play an integral role in the cortical activating system. This fact is corroborated by following observations:

- *L-dopa* (a precursor of catecholamine) has been reported to cause an improvement in comatose states due to cerebral lesions.
- *Reserpine*, a drug which depletes catecholamines in nerve terminals induces drowsiness.
- *Amphetamine*, a sympathomimetic amine, produces intense arousal and cortical activation.
- *Cocaine* also produces increased arousal and alertness. It acts by blocking reuptake of norepinephrine at the nerve terminals.

2. Acetylcholine. Role of cholinergic neurons in wakefulness is also corroborated by following observations:

- Cholinergic agonists and anticholinesterases (e.g. neostigmine) promote cortical activation and wakefulness.
- *Acetylcholine antagonist* like atropine produces a decrease in vigilance due to loss of cortical activation.
- *Alzheimer's disease*, which is associated with loss of cholinergic innervation of cerebral cortex and degeneration of cholinergic neurons of the basal forebrain, in addition to dementia is also characterized by sleep disturbances.

3. Histamine. The histamine containing neurons are located in posterior hypothalamus. These account for:

- Arousing effect produced by an intraventricular administration of histamine and
- Sedative effect produced by the antihistaminic drugs.

4. Glutamate. It is an excitatory neurotransmitter released from the cerebral cortex in highest quantities during cortical activation of spontaneous waking or that induced by stimulation of the mid-brain reticular formation.

CSF-borne peptides

It has been presumed that wakefulness-promoting factors (probably peptides) are present in CSF. Some of the CSF-borne peptides which have been known to produce wakefulness include:

- Substance P,
- Hypothalamic-releasing factors and
- Vasoactive intestinal peptide.

Blood-borne peptides

Blood-borne peptides that act as wakefulness-promoting factors are:

- *Epinephrine and histamine*. These do not cross the blood–brain barrier but act on the circumventricular organs that lie outside the blood–brain barrier and mediate cortical arousal.
- *Glucocorticoids*. These readily cross the blood–brain barrier and act directly on the neurons to enhance arousal in stress.

SLEEP

Sleep refers to a state of unconsciousness from which the individual can be aroused by sensory or other stimuli. When asleep, an individual is not aware of the environment and is unable to perform activities that require consciousness. During sleep, the stimulus pulse transfer becomes less frequent between the reticular formation and cerebral cortex.

SLEEP-WAKE CYCLE AND FACTORS AFFECTING SLEEP

Sleep and wakefulness, like many of the body's regulatory mechanisms, have circadian rhythm of about 24 h. A newborn infant has many cycles of sleep and wakefulness in 24 h, but after the age of 2 years a single sleep-wake cycle is established. In a normal adult, the sleep-wake cycle consists of 7–8 h of sleep and 16–17 h of wakefulness.

Control of sleep-wake cycle

Sleep-wake cycle, like other circadian rhythms, is endogenous. The biological clock controlling the circadian rhythms is *suprachiasmatic nucleus* of the anterior hypothalamus. The circadian rhythms are endogenous and can persist without environmental cues; however, under normal circumstances the rhythms are modulated by external timing cues called *zeitgebers* (time givers) that adapt the rhythm to the environment. Sunlight is a powerful timing cue. Light entrains this rhythm by means of *retinohypothalamic tract*. Although the suprachiasmatic nucleus regulates the timing of sleep, it is not responsible for sleep itself.

Factors affecting sleep

Sleep time remains fairly stable from day to day even under widely varying conditions and is only modestly affected by variations in activity and sensory stimulation. However, the factors which minimize sensory stimulation and favour the onset of natural sleep are:

- Darkened room,
- Comfortable surrounding temperature,
- Silence,
- Physical and mental relaxation,
- Consumption of a basic urge, such as hunger or sex and
- Low-frequency stimulation, such as by patting or knocking in a cradle or sitting in a moving vehicle.

The above described factors have only a modest effect if any. The only behavioural factor that reliably and substantially

increases sleep is prior sleeplessness. On the other hand, anxiety and emotional stimuli by release of epinephrine cause activation of RAS and make sleep more difficult.

TYPES AND STAGES OF SLEEP

Sleep is of two types: non-REM sleep and REM sleep, which alternate in a sleep cycle.

Non-REM sleep

Non-REM sleep, i.e. non-rapid eye movement sleep is also known as *slow wave sleep* (SWS), because in this type of sleep brain waves are very slow.

In normal adults, sleep mostly begins with non-REM sleep. It is rest type of sleep which a person experiences during first hour of sleep after having been kept awake for many hours. The non-REM sleep alternates with REM sleep during the sleep cycle.

The non-REM sleep is discussed under following headings:

- Stages and EEG patterns of non-REM sleep,
- Physiological changes during non-REM sleep,
- Behavioural changes during non-REM sleep and
- Intellectual changes during non-REM sleep.

Stages and EEG patterns of non-REM sleep

Stage of wakefulness. As described above, the state of wakefulness and consciousness results due to stimulatory impulses from RAS to cerebral cortex.

EEG pattern during wakefulness is characterized by asynchronous and low-amplitude brain waves called β waves (Fig. 10.11-9A).

State of quiet, awake rest with eyes closed. State of quiet, awake rest with eyes closed is the period in between the stage of wakefulness and stage of sleep.

EEG pattern during quiet awake resting stage, as described earlier (page 861), is characterized by α waves which are highly synchronized, large waves having a frequency of 8–13 cycles/s (Fig. 10.11-9B).

State of non-REM sleep. When an individual from the state of quiet rest with eyes closed enters the state of non-REM sleep the consciousness is reduced. The non-REM sleep also known as slow-wave sleep progresses in an orderly way from light to deep sleep in four stages as:

Stage 1 of non-REM sleep (stage of very light sleep). EEG pattern in this stage is characterized by low amplitude mixed frequency activity (Fig. 10.11-9C). There is still considerable sensitivity to sensory stimuli. However, the mild to moderate stimuli are often unable to produce a full arousal.

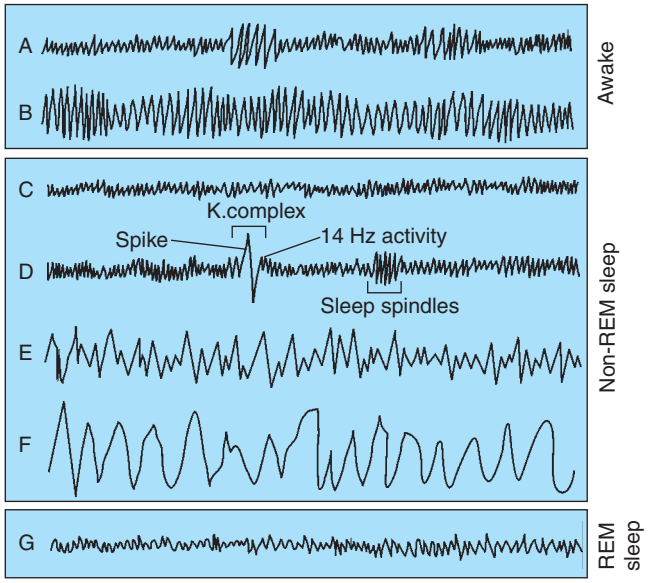


Fig. 10.11-9 EEG patterns of wakefulness and different stages of sleep: A, during wakefulness; B, during stage 1 of non-REM sleep; D, during stage 2 of non-REM sleep; E, during stage 3 of non-REM sleep; F, during stage 4 of non-REM sleep and G, during stage of REM sleep.

Stage 2 of non-REM sleep, also called stage of light sleep, is characterized by the appearance of sleep spindles. These are bursts of α-like 10–14 Hz, 50 μV waves, which periodically interrupt the α rhythm (Fig. 10.11-9D).

- Auditory stimuli during this phase readily evoke the K-complexes in the EEG. They also occur spontaneously during this stage. The K-complex consists of one or two high-voltage waves followed by a brief 14 Hz activity (Fig. 10.11-9D).

Stage 3 of non-REM sleep or stage of moderate deep sleep in characterized by an EEG that display high amplitude slow (0.5–2 Hz) waves called δ waves (Fig. 10.11-9E).

Stage 4 of non-REM sleep or stage of deep sleep produces EEG pattern dome-like very slow, large waves called δ *waves* (Fig. 10.11-9F). Thus, the characteristic of deep sleep is a pattern of rhythmic slow waves, indicating marked *synchronization*.

Physiological changes during non-REM sleep

- Muscle tone decreases progressively.
- Heart rate and blood pressure are decreased.
- Respiration rate is also decreased.
- Eyes begin slow, rolling movement until they finally stop in stage 4 (deep sleep) with eyes turned upwards.
- Body metabolism is lowered.
- Pituitary shows pulsatile release of growth hormone and gonadotropin.

Behavioural changes during non-REM sleep

Behaviourally, the non-REM sleep is characterized by:

- Progressive reduction in consciousness.
- An increasing resistance to being awakened, it is more difficult to wake up a person from stage 3 and 4 than from stage 1 and 2 of non-REM sleep.
- It is more difficult to wake up a young person than elderly from sleep because elderly person spends very little time in stage 3 and 4 of non-REM stage.
- When awaken person does not report dreaming.
- There is some response to meaningful stimuli even in sleep, which indicates that sensory processing continues at some level after the onset of sleep. This is apparent from the *discriminate responses* during sleep to meaningful versus non-meaningful stimuli. Examples of discriminate responses are:
  - Lower arousal threshold for one's own name versus someone else's name
  - A sleeping mother is more likely to hear her own baby's cry than the cry of an unrelated infant.
  - A captain wakes up to the cry of 'iceberg' in the midst of the din and bustle of a ship.

Intellectual functions during non-REM sleep

- *Thoughts* become illogical and incoherent towards the onset of sleep.
- *Retrograde amnesia* occurs during transition from wakefulness to sleep. This is because sleep inactivates the consolidation of short-term into long-term memory. Examples of retrograde amnesia include:
  - Inability to grasp the instant of sleep onset in memory,
  - Not remembering the ringing of alarm clock.

REM sleep

REM sleep, i.e. 'rapid eye movement' sleep is also called 'fast wave (desynchronized) sleep, or 'paradoxical sleep' or 'dream sleep' or 'deepest sleep' (as explained below). In adults, the REM sleep follows non-REM sleep, while in adults entry into sleep occurs via REM sleep.

EEG pattern of REM sleep

During REM sleep, EEG is characterized by a high-frequency and low-amplitude pattern (β rhythm), i.e. some desynchronized pattern that is seen in the waking state (Fig. 10.11-9G). Hence REM sleep is also called 'fast wave sleep' or 'desynchronized sleep'. However, the individual clearly is unresponsive to environment stimuli and thus is asleep. Further, it is usually more difficult to awake in REM sleep than in non-REM sleep. Because of EEG pattern of wakefulness, the REM is also called 'paradoxical sleep'.

In cats, REM sleep is also associated with ponto-geniculo-occipital (PGO) waves. The PGO waves are not detectable in humans by scalp EEG, but are recordable by depth EEG recordings. These waves originate in pons and pass rapidly to lateral geniculate body and then to cerebral cortex and hence the name PGO. These waves activate the reticular inhibiting area in the medulla producing hypotonia.

Behavioural changes during REM sleep

Arousal. As mentioned above, it is difficult to arouse an individual from REM sleep as it is from deep sleep. However, when awakened from REM sleep, the individual is immediately alert and aware of the environment. *Dreaming* occurs during REM sleep, so it is also called 'dream sleep'. There is vivid dream recall from approximately 80% of arousals from REM sleep.

Physiological changes during REM sleep

- *Rapid eye movements* are the hallmark of this state of sleep and that is why the name REM sleep. Rapid eye movements (saccadic eye movements) are bursts of small jerky movements that bring the eye from one fixation point to another to allow a sweeping of visual images of dreams.
- *Heart rate and respiration rate* become irregular.
- *Muscle tone* is reduced due to inhibition of spinal motor neurons via brain stem mechanisms. Snoring during sleep results from partial obstruction of airways caused by relaxed tongue (due to muscular atonia) in supine position.
- *Twitching of limb* musculature occurs occasionally. Because muscle tone is reduced tremendously during REM sleep, frequency and intensity of muscle twitching do not produce injuries or awaken the individual.
- *Middle ear muscles* are also active during REM sleep.
- *Penile erection* in males and engorgement of clitoris in females may occur during REM sleep.
- *Impaired thermoregulation*. Sweating or shivering during sleep in response to ambient temperature occurs in non-REM sleep and ceases in REM sleep.
- *Teeth grinding* (*bruxism*) may be seen in children.

SLEEP CYCLE

In a normal adult individual, the average sleep period of about 7–8 h is divided into about 5 cycles during which non-REM sleep and REM sleep alternate with each other. There is an orderly progression of sleep states and stages during a typical sleep cycle (Fig. 10.11-10):

Duration of sleep cycles and sleep stages

The average duration of each sleep cycle is about 90 min (range 70–120 min). Duration of different sleep stages are different in different cycles:

- Duration of non-REM sleep which is about 85 min (out of total 90 min) in first cycle decreases progressively in the next sleep cycles.
- About 25% of entire sleep period is passed in REM sleep.

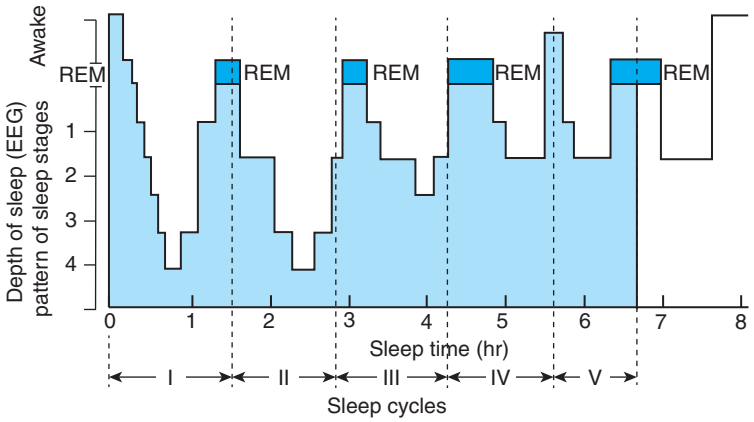


Fig. 10.11-10 Typical sleep cycles in an adult individual.

- Duration of REM sleep, which is about 5 min (out of total 90 min) in first cycle increases progressively in the next cycle.
- Duration of deeper stages (3 and 4) of non-REM sleep is maximum during first cycle and then decreases progressively and may even disappear altogether from the later cycles.
- Duration of second stage of non-REM sleep increases progressively from first cycle onwards and may even occupy most of the non-REM portion of the later cycles. About 50% of the entire sleep period is spent in second stage of non-REM sleep.
- As morning approaches, the individual may be periodically awaken during later sleep cycles.
- The approximate duration (%) of different stages of sleep during first cycle and during the entire sleep is as (Table 10.11-2):

Table 10.11-2 Approximate duration of different stages of sleep in first sleep cycle and during entire sleep

Stage of sleep	1st cycle (%)	Entire sleep (%)
Non-REM sleep		
Stage 1	5	4
Stage 2	20	50
Stage 3	30	6
Stage 4	40	15
REM sleep	5	25

Variations in sleep cycles

Variations in sleep cycle, from the typical adult pattern depicted in Fig. 10.11-10, occur under certain circumstances. In adults, onset of sleep with REM sleep occurs under special circumstances, such as in jet lag, chronic sleep deprivation, narcolepsy, acute withdrawal of REM suppressing drugs and endogenous depressions.

Variations in total sleep duration

Average sleep time per day differs according to the age:

- During infancy: 16 h,
- During childhood: 10 h,
- During adulthood: 7–8 h and
- During old age: <8 h.

Variation in time period of different stages of sleep

Effect of age

- Prematurely born infants spend about 80% of their sleep time in REM sleep.
- Full-term infants spend only 50% of their sleep time in REM sleep.
- The total time spent in REM sleep is reduced to about 1.5–2 h by puberty and remains unchanged there further.
- In adulthood, reduction in total sleep time to 8 h (2 h REM and 6 h non-REM sleep).
- *In old age*, there is very high variability in the type and duration of sleep. By the age of 60, SWS may no longer be present, particularly in men.

GENESIS OF SLEEP

The sleep state does not result from the passive withdrawal of arousal due to fatigue of RAS as thought earlier. Now, it is established that the sleep is produced by an active process which is different for non-REM sleep and REM sleep.

Genesis of non-REM sleep

The non-REM sleep is generated by interaction of neurons which are grouped as:

- Diencephalic sleep zone,
- Medullary synchronizing zone and
- Basal forebrain sleep zone.

Diencephalic sleep zone lies in the hypothalamus and the nearby intra-laminar and anterior thalamic nuclei. *A sleep*

facilitatory centre is considered to be located in the anterior hypothalamus, as its stimulation causes sleep. Posterior hypothalamus acts as a *waking centre*, as its stimulation causes wakefulness. The diencephalic sleep zone must be stimulated at low frequency (about 8 Hz) to produce sleep.

Medullary synchronizing zone is in the reticular formation of medulla oblongata at the level of nucleus of the tractus solitarius. Like diencephalic sleep zone, this zone also produces sleep when stimulated at low frequency.

Basal forebrain sleep zone includes the pre-optic area and the diagonal band of Broca. Unlike the other two zones, stimulation of this zone at low as well as high frequency produces sleep.

Activity of non-REM on cells

The *non-REM on cells* are GABAergic inhibitory neurons that mediate sleep-inducing action of the above described sleep zones. These cells are thought to produce sleep by inhibiting the *histaminergic* cells in the posterior hypothalamus as well as cells of nucleus reticularis pontis oralis (RPO) in the mid brain that mediate arousal.

Mechanism of production of sleep spindles and slow waves of non-REM sleep

The non-REM sleep is characterized by the EEG spindles and slow waves that are produced by synchronized postsynaptic potentials in the cortical neurons. These synchronized synaptic potentials are generated by the rhythmic firing of thalamic relay neurons that project to the cortex (Fig. 10.11-11). The rhythmic firing of relay neurons is a result of action of GABAergic inhibitory neurons in the nucleus reticularis that forms a shell around the thalamus.

Genesis of REM sleep

Rapid eye movement sleep is generated by the interaction of neurons in the caudal mid brain and pons with the neurons in the medulla and forebrain.

REM sleep as described earlier is characterized by:

- Blockage of EEG spindles and slow waves,
- Occurrence of PGO waves,
- Muscle atonia and
- Phasic motor action.

Genesis of the above components of REM sleep is discussed.

Role of cholinergic neurons of mid brain and the adjacent dorsal pons

These cells form an important component of the mid-brain arousal system and are maximally active during waking and REM sleep. Their activity contributes to the blocking of the slow waves of EEG.

Role of nucleus reticularis pontis oralis

The nucleus RPO forms another important neuronal machinery for genesis of REM sleep. Three classes of neurons in the RPO of particular interest are:

1. Cholinergic PGO-on cells. The discharge of these neurons produces the so-called PGO spikes that are characteristic of REM sleep (Fig. 10.11-11).

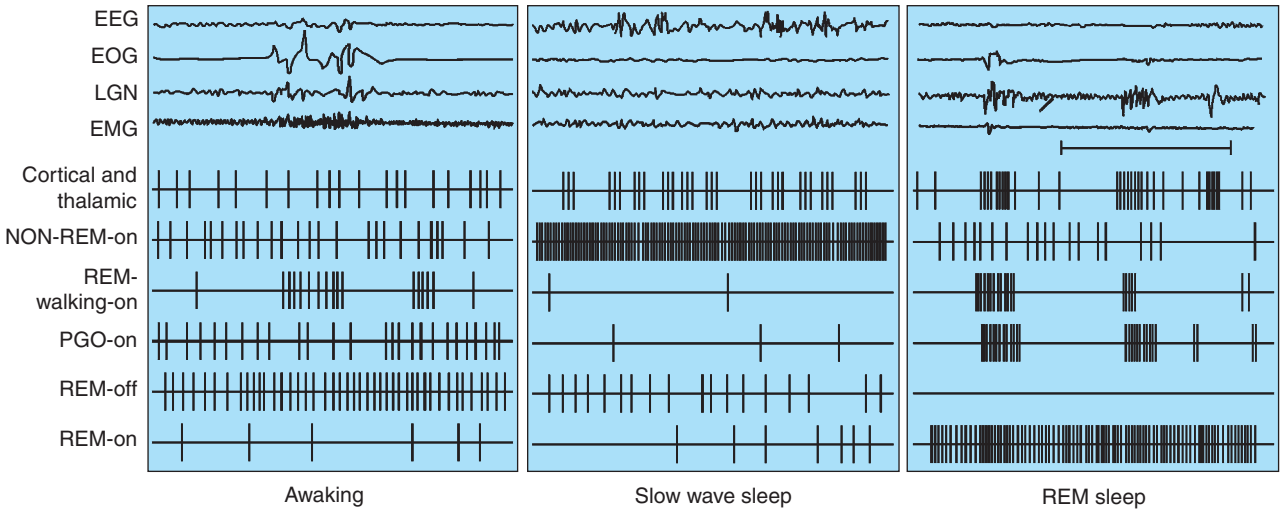


Fig. 10.11-11 The pattern of activity of key cell groups during waking and slow wave and REM sleep. Each vertical line represents an action potential. (EEG = Electroencephalogram; EOG = electro-oculogram depicting eye movement; LGN = recording from lateral geniculate nucleus showing ponto-geniculo-occipital (PGO) spikes activity during REM sleep; EMG = electromyography of dorsal neck cell.)

2. REM-waking-on-cells of RPO fire at high rate during active waking as well as during REM sleep (Fig. 10.11-11). Some of these cells project to the motor neurons in the spinal cord and others project to the motor neurons that drive the extraocular muscles.

- *Burst firing of REM-waking-on cells* during REM sleep produces rapid eye movements and muscle twitches.

3. REM-on-cells. REM-on-cells of RPO show high level of activity during REM sleep but have a very little or no activity during waking and non-REM sleep (Fig. 10.11-11). Although few in number, these cells play a key role in REM sleep.

Chemical mediators of sleep

Neurotransmitters employed by the neurons forming the neural substrate of sleep as discussed above include:

- Serotonin,
- Acetylcholine and
- Noradrenaline.

The substances that have been identified by an experiment on sleep-deprived animals as sleep-producing substances (S/S) are:

- *Muramyl dipeptide*, a chemical related to substances found in the bacterial cell walls,
- *Interleukin-1*, a cytokine that may mediate the effects of muramyl dipeptides as well as immune response,
- *Adenosine,*
- *Delta sleep-inducing peptide*, a substance isolated from the blood of sleeping rabbits,
- *Prostaglandin* D_2 and
- Arginine vasotoxin.

PHYSIOLOGICAL SIGNIFICANCE OF SLEEP

Sleep is an indispensable phenomenon. Its physiological significance is highlighted.

1. Sleep may serve as a period of body's rest and metabolic restoration as evidenced by following physiological changes during non-REM sleep:

- Pulsatile release of growth hormone and gonadotropins from the pituitary and
- Decrease in blood pressure, heart rate and respiration.

2. Sleep is necessary for certain forms of learning. In experimental animals, learning sessions do not improve performance until a period of SWS or SWS plus REM sleep has occurred. However, it is not known why sleep is necessary and there is as yet no clinical correlate to this experimental observation.

3. REM sleep is necessary for mental well-being. The correlation between dreaming and REM sleep indicates that the brain is highly active at this time. This may allow for the expression, through dreams, of concern in the subconscious and for long-term chemical and structural changes that brain must undergo to make learning and memory possible.

4. REM sleep plays an important role in homeostatic mechanism. It is evident from the observation that when the experimental animals are completely deprived of REM sleep for long periods, they loose weight in spite of increased caloric intake and finally die.

SLEEP DISORDERS

1. Insomnia refers to an inability to have sufficient or restful sleep despite an adequate opportunity for sleep. It is a subjective problem that occurs at one time or another in almost all adults. Insomnia can be relieved temporarily by sleeping pills, especially benzodiazepines. Prolonged use of these drugs can be habit-forming and can compromise day time performance.

2. Fatal familial insomnia is a serious disorder characterized by worsening insomnia, impaired autonomic and motor functions, dementia and eventually death. It is a progressive disease that occurs in both an inherited and a sporadic form.

3. Narcolepsy refers to an irresistible urge to sleep. As mentioned in the sleep cycle, in adults the sleep onset occurs with non-REM sleep, which is followed by REM sleep. However, in narcolepsy, REM sleep is entered directly from the waking states. Narcolepsy may manifest as:

- *Episodes of sudden sleep*. The individuals go to sleep while performing day time tasks.
- *Cataplexy*. In some narcoleptics, the profound reduction in the muscle tone characteristic of REM sleep can occur without loss of consciousness. During such an attack, called cataplexy, the individual suddenly becomes paralysed, falls to the ground and is unable to move.
- *Dream-like state during wakefulness* is another mode of manifestation of narcolepsy. Narcoleptics describe it as a hallucination.

4. Some sleep disorders associated with non-REM sleep (slow wave sleep), or more specifically, occurring during arousal from slow wave sleep are:

- *Sleep walking (somnambulism).* Episodes of sleep walking are more common in children than in adults and occur predominantly in males. These episodes may last for several minutes. Such individuals walk with their eyes open and avoid obstacles, but when awakened, they cannot recall the episode.

- *Bed-wetting* (nocturnal enuresis), i.e. involuntary voiding of urine, occurs in some children during slow wave sleep.
- *Nightmares (pavor nocturnus or episodes of night terror).* During a nightmare that occurs in slow wave sleep, an individual wakes up screaming and appears terrified. However, no reason for acute anxiety is recalled. By contrast, terrifying dreams that occur during REM sleep are graphically remembered.

5. REM behaviour disorder. It is a newly recognized condition in which REM sleep is not associated with inhibition of muscle tone. Consequently, such persons act out their dreams, that is, they thrash about and may even jump out off the bed, ready to do battle with imagined aggression. The generalized or localized muscle contraction associated with vivid visual imagery, i.e. the motor response to some of the dream events is referred to a *hypnoeic myoclonia*.

Chapter 10.12

Some Higher Functions of Nervous System

LANGUAGE AND SPEECH
- **Neurophysiology of language and speech**
  - Development of speech
  - Mechanism of speech and speech centres
- **Speech disorders**
  - Dysarthria
  - Aphasia

LEARNING AND MEMORY
- **Learning**
  - Incidental learning
  - Reflex learning
- **Memory**
  - Implicit memory
  - Explicit memory
  - Mechanism of memory
- **Inter-hemispheric transfer of learning and memory**
- **Applied aspects**
  - Drugs facilitating learning and memory
  - Amnesia
  - Alzheimer's disease and senile dementia

HIGHER INTELLECTUAL FUNCTIONS OF THE PRE-FRONTAL ASSOCIATION CORTEX
- Thought process
- Working memory and intellectual functions
- Episodic memory

LANGUAGE AND SPEECH

NEUROPHYSIOLOGY OF LANGUAGE AND SPEECH

Communication through language is a unique faculty which places the humans much above the animals. Language refers to that faculty of nervous system which enables the humans to understand the spoken and printed words, and to express ideas in the form of speech and writing. There are two aspects of communications: language input (the sensory aspect) and language output (the motor aspect). The sensory aspect of language includes the visual, auditory and proprioceptive impulses, while the motor aspect includes the mechanisms concerned with the expression of spoken (sound) language and written language.

DEVELOPMENT OF SPEECH

Development of speech involves co-ordinated activity of three important areas of cerebral cortex, namely Wernicke's area, Broca's area and motor areas of the categorical (dominant) hemisphere.

Development of speech in a child occurs in two stages:

First stage. In this stage, there occurs association of certain words with visual, tactile, auditory and other sensations, aroused by objects in the external world, which is stored in the memory.

Second stage. This stage of development of speech involves establishment of new neuronal circuits. When a definite meaning has been attached to certain words, pathway between the auditory area (area 41) and motor area for the muscles of articulation, which helps in speech (area 44) is established. And, the child attempts to formulate and pronounce the words, which are learnt.

MECHANISM OF SPEECH AND SPEECH CENTRES

Speech is of two types: spoken and written.

Spoken speech involves both understanding of spoken words as well as expressing ideas in the form of spoken words.

Written speech also involves both understanding of written words as well as expression of ideas in the form of written words. Mechanism of speech involves co-ordinated activities of central speech apparatus and peripheral speech apparatus. The central speech apparatus consists of cortical and subcortical centres. The peripheral speech apparatus includes larynx or sound box, pharynx, mouth, nasal cavities, tongue and lips. All the structures of peripheral speech

apparatus work in co-ordination with respiratory system under the influence of motor impulses from the respective motor areas of the cerebral cortex.

Mechanism of speech and the centres concerned with can be described separately for:

- Understanding of speech and
- Expression of speech.

Understanding of speech (sensory aspects of communication)

Different mechanisms are involved in the understanding of a spoken speech and written speech.

Understanding of spoken speech

Understanding of the spoken words is accomplished by following activities:

1. Hearing of the spoken words requires an intact auditory pathway from the ears to primary auditory areas.

Primary auditory areas, also called auditory sensory areas, include the Brodmann's area 41 and 42 and form the centre for hearing.

Location. Primary auditory areas are located in the middle of superior temporal gyrus on the upper margin and on its deep or insular aspect (Fig. 10.12-1).

Functions. This area perceives the nerve impulses as sound, i.e. auditory information, such as loudness, pitch, source and direction of sound.

2. Recognition and understanding of the spoken words is carried by *auditory association areas* (21 and 20) located in the middle and inferior temporal gyrus, respectively (Fig. 10.12-1).These areas receive impulses from the primary area and are concerned with interpretation and integration of auditory impulses.

3. Interpretation and comprehension of the speech ideas. It involves the activities of *Wernicke's area*. Wernicke's area (area 22) is a *sensory* speech centre located in the posterior part of the superior temporal gyrus behind the areas 41 and 42 (Fig. 10.12-1) in the categorical hemisphere, i.e. dominant hemisphere. *Functions* of this area are:

- Interpretation of the meaning of what is heard and
- Comprehension of the spoken language and the formation of idea that are to be articulated in speech.

Understanding of written speech

Understanding of the written speech is accomplished by following activities (Fig. 10.12-2):

1. Perception of written words requires an intact visual pathway from eyes to primary visual cortex.

Primary visual cortex, also called as striate area (area 17), or the centre of vision lies on the medial surface of occipital lobe in and near the calcarine sulcus occupying parts of lingual gyrus and cuneus. It also extends to the superolateral surface of the occipital pole limited by the lunate sulcus. *Afferents* to area 17 are fibres of the optic radiations which bring impulses from parts of both retinae and these parts are represented within the area in a specific orderly manner.

Functions. Primary visual cortex is concerned with perception of visual impulses.

2. Interpretation of written speech. *Visual association areas* (area 18 and 19), located in the walls and in front of lunate sulcus, are concerned with the interpretation of written words. These areas are involved in the recognition and identification of the written words in the light of past experience.

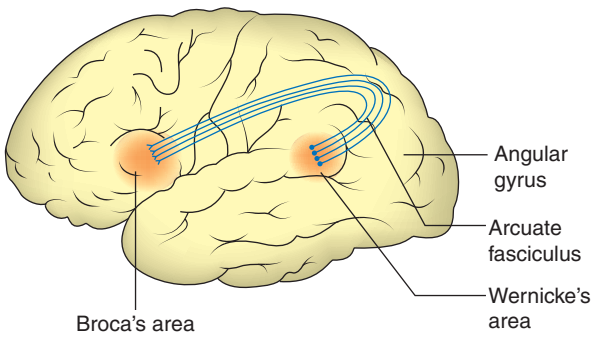


Fig. 10.12-1 Lateral surface of left (categorical) hemisphere showing location of primary areas of language.

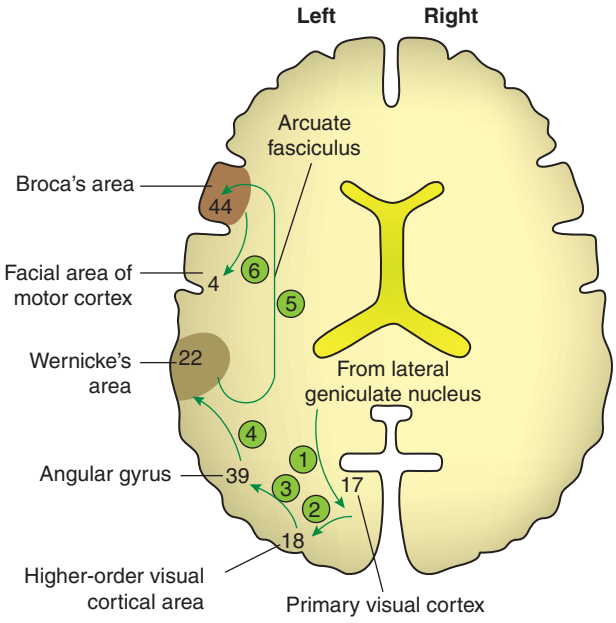


Fig. 10.12-2 Neural pathway in brain involved in the understanding and expression of written speech.

3. Generation of thoughts/ideas in response to written speech. *Dejerine area* (area 39), located in the angular gyrus behind the Wernicke's area in the dominant hemisphere, is involved in the activity of generation of thoughts/ideas in response to the written speech. This area is also called *visual speech centre* and along with the Wernicke's areas (auditory speech centre) forms the so-called sensory speech centre.

Expression of speech (motor aspect of communication)

Expression of speech in response to both spoken speech and written speech can be in the form of spoken speech or written speech or both. It involves the activities of *motor speech centres,* which include Broca's area (area 44) and Exner's area.

1. Expression in the form of spoken speech

Expression in the form of spoken speech involves the activities of motor speech (Broca's area) area.

Broca's area, or motor speech area (area 44), is a special area of the premotor cortex situated in the inferior frontal gyrus.

Functions. This area, especially in the dominant hemisphere (left hemisphere in right handed person) processes the information received from the sensory speech centres (Wernicke's area and Dejerine's area) into a detailed and co-ordinated pattern for vocalization, which is then projected by arcuate fasciculus to motor cortex for implementation. Thus, Broca's area is concerned with the movements of those structures which are responsible for the production of voice and articulation of speech, i.e. it causes activation of vocal cords simultaneously with movements of mouth and tongue during speech, lesions of this area cause motor aphasia.

2. Expression in the form of written speech

Expression in the form of written speech is the function of Exner's area (Fig. 10.12-2).

Exner's area (motor writing centre) is situated in the middle frontal gyrus in the categorical (dominant) hemisphere in the premotor cortex. It processes the information received from the Broca's area into detailed and co-ordinated pattern; and then along with the motor cortex (area 4) initiates the appropriate muscle movements of the hand and fingers to produce written speech.

Concept of dominant hemisphere for language

In human cerebral cortex, the interpretive functions of Wernicke's area, the angular gyrus and the frontal motor speech areas (i.e. the ability to understand or express oneself by spoken or written speech) are more highly developed in one hemisphere called the *dominant hemisphere*. How one hemisphere comes to be dominant is not yet understood.

It is important to note that:

- In approximately 95% of all individuals, the left hemisphere is dominant regardless of handedness.
- Since, the motor area concerned with the hand movements is closely associated with the centre for speech, this explains the right handedness in over 90% of the individuals.
- Right hemisphere dominance is seen in only 15% of left handers.
- Seventy percent of left handers also have left hemisphere dominance.
- The area in the non-dominant hemisphere that corresponds to the Wernicke's area is also involved in the language function. It is responsible for understanding the emotional content or intonation of spoken language. It also serves equally important functions of understanding and interpreting non-verbal, visual or auditory experiences, such as recognition of visual patterns or faces and interpretation of music.

Concept of categorical and representational hemisphere

Presently, it is believed that left hemisphere is not really dominant over the right hemisphere. In fact, the two halves of the brain have independent capabilities of consciousness, memory storage and control of motor activities and speech. The corpus callosum and anterior commissure connect the two halves of brain. By these connections, information stored in one hemisphere is made available to the other hemisphere and then the activities of two hemispheres are co-ordinated.

As summarized below some specialized higher functions are allowed to each hemisphere. Therefore, the terms 'dominant' and 'non-dominant' have been replaced by *categorical* and *representational* hemisphere, respectively.

Functions allotted to left hemisphere in a right-handed person

- Right-hand control,
- Spoken language,
- Written language,
- Mathematical skills,
- Scientific skills and
- Reasoning.

Functions allotted to right hemisphere in a right-handed person

- Left-hand control,
- Music awareness,
- Three-dimensional awareness,

- Art awareness,
- Insight and
- Imagination.

SPEECH DISORDERS

DYSARTHRIA

Dysarthria is a disorder of speech in which articulation of words is impaired, but the comprehension of spoken and written speech is not affected. It may be due to paresis, or inco-ordination of the muscles involved in the production of speech as seen in the lesions of pyramidal tract, cranial nerves, cerebellum or basal ganglia.

APHASIA

Aphasia refers to the inability to understand spoken or written speech or inability in expressing the spoken or written speech in the absence of mental confusion or motor deficit. Depending upon the site of lesion, the aphasia may be:

- Sensory aphasia,
- Motor aphasia, or
- Global aphasia.

Sensory aphasia

Site of lesion. Sensory or receptive aphasia, also known as Wernicke's aphasia, is the result of lesion in the Wernicke's area.

Characteristic features of sensory aphasia are:

1. Difficulty in understanding the meaning of speech. In this condition, the affected individuals are capable of hearing or identifying written or spoken words, but they do not comprehend the meaning of the words.

2. Motor speech is intact and the patients talk very fluently (or rather excessively), that is why, it is also called *fluent aphasia*. However, the speech does not make much sense and is often associated with:

- *Anomia,* i.e. inability to find an appropriate word to express a thought.
- *Neologism,* i.e. using or creating new words or new meanings for established words.
- *Paraphasias,* i.e. production of unintended words or phrases during effort to speak.

3. Impairment in reading and writing. Since the patient cannot comprehend the written words (*word blindness*) he/she is unable to read aloud or copy print into writing.

4. Conduction aphasia is another form of fluent aphasia in which patient can speak well and there is good auditory comprehension, but he cannot put parts of words together. It occurs due to the lesion of arcuate fasciculus connecting Wernicke's and Broca's areas or lesion in the auditory cortex (area 40, 41 and 42).

Motor aphasia

Site of lesion. Motor aphasia, also known as Broca's aphasia, results from lesions involving the Broca's motor speech area (area 44) in the frontal lobe.

Characteristic features of Broca's aphasia are:
1. Comprehension of written or spoken speech is good.

2. Difficulty in speaking. The affected individual is able to formulate verbal language in his mind but cannot vocalize the response. The defect is not in the control of musculature needed for speech but rather in the elaboration of the complex patterns of neural and muscle activation that is *effect,* which defines the motor aspect of language.

3. Speech is non-fluent, i.e. the patient utters only a few words with great difficulty. Because of this, motor aphasia is also known as *non-fluent aphasia*, or *expressive aphasia*.

4. Inability to write (agraphia).

Global aphasia

Global aphasia refers to the total inability to use language communication.

Site of lesion. This condition is produced as a result of loss of both Wernicke's and Broca's areas.

Dyslexia is a broad term applied to inability to read.

Common cause of aphasia

Aphasias are mostly produced by thrombosis or embolism of a blood vessel in the dominant hemisphere. Aphasias are commonly associated with right-sided motor and sensory deficit but may also occur independently, when the lesion is restricted to cortical association area.

Lesion in the representational hemisphere produces impairment of telling a story or a joke.

LEARNING AND MEMORY

LEARNING

Learning and memory are closely related. Learning is impossible without memory and memory has no meaning without learning. In fact, learning and memory are two sides of a coin. Learning refers to a neural mechanism by which the individual changes his or her behaviour on the basis of the past experience. Two patterns of learning are:

Incidental learning, in which the behavioural change is not immediately apparent. The individuals acquire information

about the world, while attending incidentally to sensory inputs and thereby develop the potential to behave differently. The two broad classes of learning are:

- Non-associative learning and
- Associative learning.

Reflex learning, in which the learning is associated with an immediate behavioural changes.

A. NON-ASSOCIATIVE LEARNING

In non-associative learning, the subject learns about the properties of a single stimulus. It results when an animal or person is repeatedly exposed to a single type of stimulus. Two forms of non-associative learning are common in everyday life: habituation and sensitization.

HABITUATION

Habituation refers to a decrease in response to a benign (neutral type) stimulus when the stimulus is presented repeatedly. When the stimulus is applied for the first time, it is novel and evokes reaction. This response is called *orientation reflex* or 'what is it' response. However, due to habituation lesser and lesser response is evoked on repeated stimulation. Eventually, the subject totally ignores the stimulus and thus gets habituated to it. For example, when a new clock is presented to a subject, at first the ticking noise may be annoying and may cause some difficulty in sleeping. However, after several nights the clock is no longer noticed.

IMPORTANT NOTE

Presentation of another, usually noxious stimulus results in recovery of the habituated response, i.e. *dishabituation.* Dishabituation is a major criterion to demonstrate that habituation has indeed occurred.

Cellular basis of habituation. Habituation is associated with a decrease in neurotransmitter released at the synapses, which in turn is due to the inactivation of Ca^{2+} influx at the axon endings. However, the mechanism of inactivation of Ca^{2+} channels is not known.

SENSITIZATION

Sensitization is opposite to habituation. In it repeated application of a distinctly pleasant or unpleasant (strong) stimulus produces greater and greater response. For example, an animal responds more vigorously to a mild tactile stimulus, after it has received a painful pinch. Thus, in sensitization, learning occurs in a direction opposite to that seen in habituation, presumably so that the behaviour becomes in lower animals directed toward escape from the stimulus. Moreover, a sensitizing stimulus can override the effects of habituation, i.e. can cause dishabituation (as described above).

Cellular mechanism of sensitization The sensitization is associated with increased release of neurotransmitters from the axonal endings of sensory neurons. This result due to the pre-synaptic facilitation of synaptic transmission brought about by a third neuron called facilitatory neuron. The transmitter released by pre-synaptic interneuron, is serotonin (5-HT).

B. ASSOCIATIVE LEARNING

In associative learning, the subject learns about the relationship between two stimuli or between a stimulus and a behaviour. Two forms of associative learning have been distinguished based on the experimental procedures used to establish the learning:

- Classical conditioning and
- Operant conditioning.

CLASSICAL CONDITIONING

Classical conditioning involves learning a relationship between two stimuli. Classical conditioning is also termed Pavlovian conditioning, conditioned reflex type I, respondent conditioning or type-S conditioning.

Characteristic features of a classical conditioned reflex are:

- A conditioned reflex is reflex response to stimulus that previously elicited little or no response, *acquired by repeatedly pairing the stimulus with another stimulus* that normally does produce the response. Thus, in classical conditioning, a temporal association is made between a neutral conditioned stimulus (CS) and an unconditioned stimulus (US) that elicits an unlearned response. It depends for its appearance on the *formation of new functional connections in CNS.*
- *Reinforcement*, i.e. a process of following a CS with the basic US is must for retaining a conditioned reflex otherwise it will *extinct.*

Pavlov's experiment to demonstrate classical conditioned reflex is:

- When food, i.e. an unconditional stimulus (US), is presented to a hungry dog, it produces salivation (an unconditioned response), or
- If a bell is rung (a conditioned stimulus (CS), just before the food (US) is presented, the dog learns to associate the bell (CS) with the food (US).
- Eventually, ringing the bell (CS) alone causes salivation.
- Of course, if the food fails to appear consistently when the bell is rung, the conditioned response fades away, a process called *extinction* or *internal inhibition.* Thus, a conditioned reflex needs to be reinforced frequently, otherwise it dies out.

Prerequisites for development of conditioned reflex

- *Alertness and good health.* The animal must be alert and in good health.
- *Timing of CS and US stimuli* is critical in classical conditioning. The CS must precede the US, often within an interval of about 0.5 s. If the CS follows US, no conditioned response is developed.
- *Duration of CS.* The CS must be allowed to continue to act so as to overlap the US.
- *Reinforcement.* For a conditioned reflex to continue, it is essential that CS should always be followed by US. As described above, when US fails to follow CS consistently, the conditioned reflex fades away soon. This phenomenon is known as *extinction* or *internal inhibition.*
- *External inhibition.* When the animal is disturbed by an external stimulus immediately after the CS is applied, the conditioned response may not occur. This is called external inhibition.
- *Type of US.* The conditioned reflexes are difficult to form when the US proves a pure motor response; since the motor responses are also under voluntary control.
- *Pleasant and unpleasant versus neutral US.* Conditioned reflexes are relatively easily formed when the US is associated with a pleasant or unpleasant effect than when associated with a neutral effect. For example, stimulation of the brain reward system is a powerful US; this is called pleasant or *positive reinforcement.* Similarly, stimulation of the avoiding system or a painful shock to the skin is also a powerful US; and this is called an unpleasant or *negative reinforcement.*

Physiological basis of conditioned reflexes

Physiologically, the occurrence of conditioned reflex is explained by the formation of a new functional connection in the nervous system. For example, in Pavlov's classical experiments, salivation in response to ringing of a bell indicates that a functional connection has developed between the auditory pathways and the autonomic centres controlling salivation.

Site of formation of functional connections can be intracortical as well as subcortical.

Evidences in favour of intracortical level are:

- In decorticate animals, the conditioned reflexes can be built up with great difficulty.
- Presence of sensory cortex is must to understand a complex sensory conditioned stimulus.

Evidence in favour of subcortical level. Non-discriminative conditioned reflexes to simple sensory stimuli can be formed in the absence of whole neocortex. This indicates that the new functional connections can also be formed at subcortical level.

OPERANT CONDITIONING

Operant conditioning is also termed as instrumental conditioning, type II conditioning, type-R conditioning or trial-and-error conditioning. It involves associating a specific behaviour with a reinforcement event. In it the organism's behaviour is instrumental in conditioning. Therefore, the organism learns which of its actions are responsible for the occurrence of reinforcement event.

Operant conditioning is of two types:

- *Reward conditioning.* In it a naturally occurring response is strengthened by positive reinforcement (reward).
- *Aversive conditioning.* In it a naturally occurring (innate) response is weakened by a negative reinforcement (punishment).

Experiment to demonstrate operant conditioning

A hungry animal (e.g. rat) is placed in a cage with a lever (bars) protruding in the cage. Because of naturally occurring (innate) response the rat will randomly press the lever.

- If pressing of lever is not associated with any event the pressing of the lever will be at a random rate.
- If pressing a lever is associated with a positive reinforce, i.e. reward (e.g. food) the rate of pressing the lever will be much more than the random rate (*reward conditioning*).
- If pressing of lever is associated with a negative reinforce, i.e. punishment (e.g. electric shock), the lever-pressing rate will be much less than the random rate (*aversive conditioning*).

Neural mechanism of operant conditioning

Because operant and classical conditioning involve different kinds of association—classical conditioning involves learning an association between two stimuli whereas operant conditioning involves learning the association between a behaviour and a reward—one might suppose the two forms of learning are mediated by different neural mechanisms. However, the laws of operant and classical conditioning are quite similar, suggesting that the two forms of learning may use the same neural mechanisms.

MEMORY

As mentioned earlier, memory and learning are closely related to each other. Memory refers to the acquisition, storage and retrieval of sensory information; while learning is the change in behaviour based on the sensory information stored in the brain. Brain has different sites and mechanisms for handling different types of information.

TYPES OF MEMORY

Memory can be classified in two ways:

I. Physiologically, on the basis of how information is stored and recalled

The memory can be classified as:

- Implicit memory
- Explicit memory

II. Depending upon permanency of storage memory is:

1. Short-term memory, also termed as *primary memory*, lasts for seconds to hours.

2. Intermediate long-term memory (or secondary memory) lasts for days to weeks but is eventually lost.

3. Long-term memory (or tertiary memory), which once stored, can be recalled years later or for a lifetime.

IMPLICIT MEMORY

Implicit memory, also termed as *reflexive* or *non-declarative memory*, refers to the information about how to perform something. It does not depend directly on conscious processes nor does recall require a conscious search of memory. This type of memory builds up slowly through repetition over many trials and is expressed primarily in performance, not in words. Examples of implicit memory include motor skills, habits, behavioural reflexes and the learning of certain types of procedures and rules which, once acquired, become unconscious and automatic. It also includes priming in which recall of words and objects is improved by prior exposure to them. Most forms of implicit memory are acquired through different forms of reflexive learning which comprise:

1. *Non-associative learning* that includes:
   - Habituation and
   - Sensitization.
2. *Associative learning* that includes:
   - Classical conditioning and
   - Operant conditioning.

Different forms of reflexive learning which comprise implicit memory have been described above. These involve different brain regions:

- Memory acquired through *fear conditioning*, which has an emotional component is thought to involve *amygdala*.
- Memory acquired through *operant conditioning* requires the *striatum* and *cerebellum*.
- Memory acquired through classical conditioning, sensitization and habituation involves changes in the sensory and motor systems involved in the learning.

EXPLICIT MEMORY

Explicit memory, also termed as *declarative* or *recognition memory*, refers to the factual knowledge of people, places, things and what these facts mean. This is recalled by a deliberate conscious effort. Explicit memory is highly flexible and involves the association of multiple bits and pieces of information. In contrast, implicit memory is more rigid and tightly connected to the original stimulus conditions under which the learning occurred.

Explicit memory can be further classified as *semantic memory* (a memory of facts) and *episodic memory* (a memory for events and personal experience).

Semantic (factual) memory

The semantic memory is that form of long-term explicit memory that embraces knowledge of objects, facts and concepts as well as words and their meaning. It includes the naming of objects, the definition of spoken words, and verbal fluency.

Semantic memory is stored in a distributed fashion in different association cortices. For example, the word alarm clock, immediately brings its features in our mind from our past experience (stored memory) as follows:

- *Visual memory* reminds us about its shape, needles depicting hours, minutes and seconds, and markings for 1–12 O'clock hours, etc.
- *Auditory memory* reminds us about its sound (ringing of alarm);
- *Somatosensory memory* reminds us that it is made of a plastic or metallic box, having a smooth transparent glass.

The visual, auditory and somatosensory memory, which reminds us about different attributes, is stored in different areas of neocortex. Whenever the information about the features of an alarm clock has to be recalled, the recall is built up from distinct bits of information, each of which is stored in specialized (dedicated) memory stores of neocortex. Thus, there is no general semantic memory store, i.e. semantic knowledge is not stored in a single region.

Damage to a specific cortical area leads to loss of specific information and therefore a fragmentation of knowledge as exemplified:

- *Associative visual agnosia* results from damage to the posterior parietal cortex. In it patient cannot name objects but can identify them by selecting the correct drawing and can faithfully reproduce detailed drawings of the object.
- *Appreciative visual agnosia* occurs in damage to occipital lobes and surrounding region. In it patients are unable to draw objects but they can name them if appropriate perceptual cues are available.

Episodic (autobiographical) memory

Episodic memory refers to memory of events and personal experiences. For example, we use episodic memory when we recall that last Sunday I visited my friend's house in Kailash Colony of New Delhi.

Episodic memory is stored in association areas of prefrontal cortex. These prefrontal areas work with other areas of the neocortex to allow recollection of when and where a past event occurred. Therefore, particularly striking symptom in patients with frontal lobe damage is *source amnesia*, i.e. tendency to forget how information was acquired. Since the ability to associate a piece of information with the time and place it was acquired is at the core of how accurately we remember the individual episodes of our lives, a deficit in source information interferes dramatically with the accuracy of recall of episodic knowledge.

MECHANISM (PHYSIOLOGICAL, AND CELLULAR OR MOLECULAR BASIS) OF MEMORY

Studies of memory retention and disruption of memory have revealed that both explicit and implicit memory are stored in stages by different mechanisms. Input to the brain is processed into *short-term memory* before it is transformed through one or more stages (intermediate long-term memory) into more permanent long-term storage.

MECHANISM OF IMPLICIT MEMORY

As mentioned earlier, most forms of implicit memory are acquired through different forms of reflexive learning (habituation and sensitization), and associative learning (classical and operant conditioning). Short-term storage of implicit memory for these simple forms of learning result from *changes in the effectiveness of synaptic transmission*:

- *Cellular basis of habituation* is described on page 786.
- *Cellular mechanism of sensitization* is described on page 786.
- *Physiological basis and cellular mechanism of classical conditioning* is described on page 786.

Mechanism of long-term storage of implicit memory

The process by which transient short-term memory is converted into a stable long-term memory is called consolidation. *Consolidation of long-term implicit memory* for simple forms of learning involves three processes:

- Gene expression,
- New protein synthesis and
- Growth (or prunning) of synaptic connections.

MECHANISM OF EXPLICIT MEMORY

Both semantic and episodic types of explicit memory are the result of at least four related but distinct types of processing; encoding, consolidation, storage and retrieval.

Mechanism of short-term explicit memory

Encoding refers to the process by which newly learned information is attended to and processed when first encountered. The extent and nature of this encoding are critically important for determining how well the learned material will be remembered at later times. For a memory to persist and be well remembered, the incoming information must be encoded thoroughly and deeply.

Neural substrate for encoding of explicit memory

As mentioned earlier, the explicit memory is associated with consciousness (or at least awareness) and is dependent for its retention on the *hippocampus* and other parts of *medial temporal lobes* of the brain.

Studies with human patients and with experimental animals suggest that the knowledge stored as explicit memory is processed as (Fig. 10.12-3):

- Sensory information is first acquired through processing in one or more of the three polymodal association cortices (the *prefrontal, limbic*, and *parieto-occipital-temporal cortices*) that synthesize visual, auditory and somatic information.
- From polymodal association cortices, the information is conveyed in series to the *parahippocampal* and *perirhinal cortices*, then the *entorhinal cortex*, the dentate gyrus,

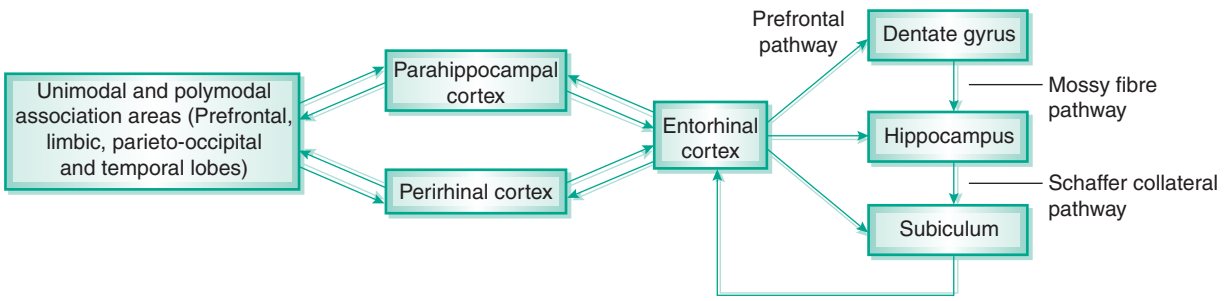


Fig. 10.12-3 Neural substrate for encoding of explicit memory (the input and output pathways of the hippocampal formation).

the hippocampus, the subiculum and finally back to the entorhinal cortex.
- From the entorhinal cortex the information is sent back to parahippocampal and perirhinal cortices and finally back to polymodal association areas of the neocortex.

Physiological processes in the neural substrate associated with the storage of short-term explicit memory (*hippocampus*) are:

- Continuous neural activity in reverberating circuits,
- Activation of synapses on pre-synaptic terminals that typically result in prolonged facilitation, i.e. long-term potentiation or prolonged inhibition i.e. long-term depression and
- Accumulation of calcium in axon terminals may eventually lead to enhanced synaptic output from the terminal.

Mechanism of intermediate long-term memory

Intermediate long-term memory can result from the temporary chemical or physical changes in either the pre-synaptic or post-synaptic membrane that can persist for a few minutes to several weeks. The newly stored sensory information is still labile during this stage, which is converted into long-term memory after the process of consolidation is complete.

Mechanism of long-term memory

Consolidation of memory. For memories to be converted to long-term memories, they must be consolidated. Consolidation refers to those processes that alter the newly stored and still labile information so as to make it more stable for long-term storage. In general, 5–10 min is required for minimal consolidation, whereas one or more hours may be needed for strong consolidation. If this time is not allowed for the consolidation to occur, the data in short-term memory is completely forgotten.

APPLIED ASPECTS

- In patients with concussion injury and after electro-convulsive therapy (ECT), who are unable to recall the events immediately preceding the concussion or convulsion. This phenomenon is called *retrograde amnesia*.
- A similar retrograde amnesia occurs before the onset of sleep. This is the reason one is unable to remember the precise time of one's own sleep onset.

Rehearsal mechanism is thought to represent the consolidation process. Rehearsal of the same information again and again in the mind potentiates the transfer from short-term to long-term memory. Over the time, the important features of sensory experience become progressively more fixed in memory stores. Also during consolidation, memories are codified into different classes of information. For example, new and old experiences related to a topic are compared for similarities and differences, and it is the later information that is stored.

Process of consolidation involves the expression of genes and synthesis of new proteins, giving rise to structural changes that store memory stably over time. The structural changes include:

- An increase in the number of synaptic vesicle release sites,
- An increase in the number of available synaptic vesicles,
- An increase in the number of synaptic terminals and
- Changes in the shape or number of postsynaptic spines.

Storage of memory refers to the mechanism and sites by which memory is retained over time. One of the remarkable feature about long-term storage is that it seems to have an almost unlimited capacity. In contrast short-term working memory is very limited.

Neural substrate for long-term storage memory

While the encoding process for short-term explicit memory involves the hippocampus, long-term memories are stored in the various parts of neocortex. Apparently, the various parts of the memories—visual, olfactory, auditory, etc. are located in the cortical regions concerned with these functions and the pieces are tied together by long-term changes in the strength of transmission at relevant synaptic junctions so that all the components are brought to consciousness when the memory is recalled.

Retrieval of memory

Retrieval refers to those processes that permit the recall and use of stored information. Retrieval involves bringing different kinds of information together that are stored separately in different storage sites.

Retrieval of information is most effective when it occurs in same context in which the information was acquired and in the presence of same cues (retrieval cues) that were available to the subject during learning. However, once established, long-term memories can be recalled or accessed by a large number of different associations. For example, the memory of a vivid scene can be evoked not only by a similar scene but also by a sound or smell associated with the scene (*dejavu phenomenon*, French word means already seen). Thus there must be multiple routes or keys to each stored memory.

Working memory

Both the initial encoding and the ultimate recall of explicit memory (and perhaps some forms of implicit memory as well) are thought to require recruitment of stored information into a special short-term memory store called

working memory. Working memory has three component systems:

- Attention control system,
- Rehearsal systems that include:
  - Articulatory loop and
  - Visuospatial sketch pad.

Attentional control system or (central executive) actively focuses perception on specific events in the environment. It is located in the prefrontal cortex and has a very limited capacity (less than a dozen items). It regulates the information flow to two rehearsal systems that are thought to maintain memory for temporary use.

Rehearsal systems include the articulatory loop and the visuospatial sketch pad.

- *Articulatory loop* is a storage system where memory for words and numbers can be maintained by subvocal speech. It is this system that allows one to hold in mind, through repetition, i.e. a new telephone number as one prepares to dial it.
- *Visuospatial sketch pad* represents both the visual properties and the spatial location of object to be remembered. This system, allows one to store the image of the face of a person one meets at a dinner party.

The two rehearsal memory systems are thought to be located in different parts of the posterior association cortices. The information processed in either of these systems has the possibility of entering long-term memory.

INTER-HEMISPHERIC TRANSFER OF LEARNING AND MEMORY

Much information is transferred between the two hemispheres through the corpus callosum, although some is transmitted through other commissures (e.g. the anterior commissure or hippocampal commissure).

Failure of inter-cortical transfer of learning and memory is seen in human patients who have had a surgical transection of the corpus callosum to prevent inter-hemispheric spread of epilepsy. Studies in subject indicate that the transfer of visual memory occurs in the posterior part of the corpus callosum, while transfer of auditory and somaesthetic memory occurs in the anterior part of the corpus callosum.

Further, functional capabilities of the two hemispheres when compared by exploring the performance of individuals with a transected corpus callosum have yielded following results:

- Right hemisphere specializes in spatial task, facial expression, body language and speech into notion.
- Patients with a transected corpus callosum lack coordination. For example, when they are dressing, one hand may button a shirt while other tries to unbutton it.
- From this experiment, it can be concluded that the two hemispheres can operate quite independently when they are no longer interconnected.

APPLIED ASPECTS

- Drug facilitating memory,
- Amnesia,
- Alzheimer's disease and senile dementia.

DRUGS FACILITATING MEMORY

- Learning and memory are reported to improve in animals when a variety of CNS stimulants are administered immediately before or after the learning sessions.

Common CNS stimulant that facilitates learning and memory are: caffeine, amphetamine, physostigmine, nicotine, pemoline, strychnine and pentylenetetrazol.

Mechanism of action. CNS stimulants act probably by facilitating consolidation of memory. For example, physostigmine acts by inhibiting acetylcholinesterase and hence preventing breakdown of acetylcholine, while nicotine stimulates cholinergic receptors.

AMNESIA

Amnesia refers to the loss of memory. It is of two types: antegrade amnesia and retrograde amnesia.

Antegrade amnesia refers to the inability of an individual to establish new long-term memories of those types of information that form the basis of intelligence. This usually occurs in lesions involving hippocampus.

Retrograde amnesia refers to the inability of an individual to recall past memories. Amnesia is much greater for events of recent past than those of remote past. Memories of distant past are rehearsed so many times that the memory traces are deeply engrained and elements of these memories are stored in the widespread areas of the brain. Retrograde amnesia occurs in lesions involving the temporal lobe (*temporal lobe syndrome*).

ALZHEIMER'S DISEASE AND SENILE DEMENTIA

Senile dementia

Senile dementia refers to a clinical syndrome in elderly people that is characterized by progressive impairment of memory and cognitive capacities. There are a number of diseases that are manifested by dementia in mid and late life.

Common causes of dementia in the elderly are:

- Alzheimer's disease,
- Cerebrovascular disease,
- Parkinsonism.

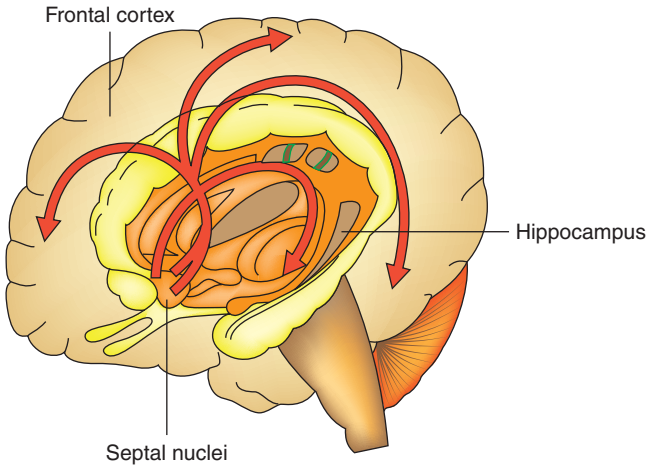


Fig. 10.12-4 Major cholinergic pathways involved in Alzheimer's disease.

Alzheimer's disease

Alzheimer's disease is the most common cause of dementia in the elderly persons.

Pathophysiology. Alzheimer's disease is a prototypical neurodegenerative disease. It is characterized by a series of abnormalities in the brain that selectively affect neurons in specific regions, particularly in the neocortex, the entorhinal area, hippocampus, amygdala, nucleus basalis, anterior thalamus, locus coeruleus and raphe complex. There is a severe loss of cholinergic neurosis in the affected areas. The major cholinergic pathways involved in Alzheimer's disease are shown in Fig. 10.12-4.

Alzheimer's disease is associated with cytoskeletal abnormalities in the affected nerve cells, most important being accumulation of neurofibrillary tangles in the neuronal cytoplasm. Amyloid plaque (fibrillar peptides) deposits are one of the hallmarks of Alzheimer's disease.

The sequence of events in the pathogenesis of Alzheimer's disease is depicted in Fig. 10.12-5.

Characteristic features of Alzheimer's disease are:

- Loss of recent memory in an otherwise alert individual,
- Impairment in other areas of cognition, such as language, problem solving, judgement, calculation, attention, perception and so on.
- Psychiatric symptoms begin to appear as the disease progresses.
- Extrapyramidal and akinetic hypertonic symptoms also appear in later stages.
- There may occur loss of spatial orientation.
- Finally, patient has to lead a vegetative life without memory, without thinking power, speechless, inability to understand anything, *apraxia* (inability to perform voluntary movements), *agnosia* (inability to recognize objects in spite of intact sensory modality).

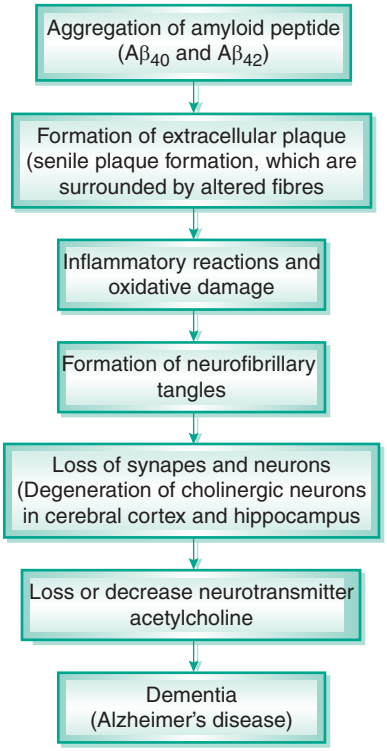


Fig. 10.12-5 The sequence of events in the pathogenesis of Alzheimer's disease.

Treatment. There is no effective treatment for Alzheimer's disease, as yet *physostigmine*, which inhibits cholinesterase causes some improvement. Presently, focus is on treating associated symptoms, such as depression, agitation, sleep disorders, hallucinations and delusions.

HIGHER INTELLECTUAL FUNCTIONS OF THE PREFRONTAL ASSOCIATION CORTEX

Prefrontal cortex refers to the portion of frontal lobes in front of the motor cortex. This area, like other association areas, is better developed in man than in any other species. The function of the prefrontal cortex is complex and multifactorial, and is typically explained by describing the deficits seen in individuals in whom the prefrontal lobotomy has been performed for tumour of this region. The functions thought to be performed by prefrontal cortex are:

1. Role in thought process

Prefrontal cortex gathers information from widespread area of the brain to develop solutions to problems, whether they require motor or non-motor responses. Without this function, thoughts lose their logical progression and the individual loses the ability to focus attention and is easily distracted

in the sequence of thoughts. Hence any activity involving a number of steps in sequence cannot be performed properly. In other words, there occurs *inability to progress towards goals* or *to carry through sequential thoughts*.

2. Site of working memory and intellectual functions

Prefrontal cortex is considered the site of 'working memory'. Working memory refers to the ability to hold and sort bits of information to be used in problem-solving function. By combining these stored bits of information, an individual can prognosticate, plan for the future, delay a response while further information is gathered, consider the consequences of actions before they are performed, correlate information from many different sources and control actions in accordance with societal or moral laws. All of these are considered *intellectual functions* of the highest order and seem to be definitive for the human experience.

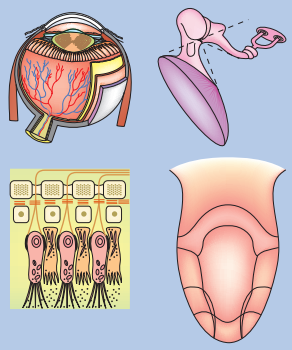
The patients with lesions of prefrontal cortex have great difficulty in abstract thinking, e.g. planning for future or considering the consequences of a particular motor activity beforehand. The patient cannot act within the norm of social or moral behaviour.

3. Role in episodic memory

Patients with prefrontal lesions show a difficulty in remembering the temporal sequence of events, i.e. he cannot remember how long ago, he save an event or picture card (episodic memory, i.e. a memory for events and personal experience).

Special Senses

SECTION 11



As we have studied, the sensory division of the human nervous system is concerned with collection of the information about the outside world and the changes occurring within the body itself. The term sensation refers to the conscious perception of sensory information reaching the brain. The sensations have been broadly divided into general and special sensations.

General sensations. These, depending upon their point of origin, can be classified into three main groups:

- Exteroceptive sensations, i.e. those arising from the skin, e.g. touch and temperature sensation,
- Visceral sensations, i.e. those arising from the viscera and
- Proprioceptive and kinaesthetic sensations, i.e. those arising from the muscles, tendons and joints.

Special sensations. There are a few organs in the body which collect information of special significance to us from the external environment and are, therefore, called the organs of special senses. These special sensory systems include:

- *The sense of vision,* which allows the animal to detect and analyse light,
- *The sense of hearing* that makes possible the detection and analysis of sound and
- *The chemical senses of taste and smell,* which are responsible for appreciation of chemical signals in the environment.

GENERAL VERSUS SPECIAL SENSATIONS

- *Location of receptors*. Receptors for general sensations are located throughout the body, e.g. touch, pain, pressure and temperature sensations, while the receptors for special senses are located at one place in the head near the nervous system, e.g. receptors for vision, hearing, taste and smell.
- *Response of receptors*. The receptors for general sensations get easily stimulated by different stimuli; however, they respond maximally to an adequate stimulus. Further, the receptors response is non-specific to different stimuli, while the receptor for special senses are specialized and respond only to one type of stimulus. Further, the receptor response is more complex and makes co-ordination within the central nervous system.

“This page intentionally left blank"

Chapter 11.1

Sense of Vision

INTRODUCTION AND FUNCTIONAL ANATOMY

INTRODUCTION

Sense of vision, the choicest gift from the Almighty to the humans and other animals, is a complex function of the two eyes and their central connections. The eyeballs are able to perform their function with the help of following physiological activities:

- Maintenance of clear media of the eye,
- Maintenance of normal intraocular pressure,
- The image forming mechanism,
- Physiology of vision,
- Physiology of binocular vision,
- Physiology of pupil, and
- Physiology of ocular motility.

Before discussing the details of the above physiological considerations, it will be worthwhile to be conversant with the broad outlines of functional anatomy of the eyeball and related structures.

FUNCTIONAL ANATOMY

There are two eyeballs, each being suspended by *extraocular muscles* and fascial sheaths in a quadrilateral pyramid-shaped bony cavity called *orbit.* Each eye is protected anteriorly by two shutters called the *eyelids.* The anterior part of the sclera and posterior surface of the eyelids are lined by a thin membrane called *conjunctiva.* For smooth functioning, the cornea and conjunctiva are to be kept moist by tears, which are produced by the lacrimal gland and drained by the lacrimal passages, which together form the *lacrimal apparatus.* The eyelids, the eyebrows, the conjunctiva and the lacrimal apparatus are collectively known

as the *appendages of the eye.* A brief account of anatomy of the eyeball and its related structures is given.

THE EYEBALL

Each eyeball (Fig. 11.1-1) is a cystic structure kept distended by the pressure inside it. Although, generally referred to as a globe, the eyeball is not a sphere but an oblate spheroid.

Coats of the eyeball

The eyeball comprises three coats: outer (fibrous coat), middle (vascular coat) and inner (nervous coat).

1. The outer fibrous coat

The fibrous coat (Fig. 11.1-1) is a dense strong wall which protects the intraocular contents. Anterior one-sixth of this fibrous coat is transparent and is called cornea. The posterior five-sixth opaque part is called sclera. Junction of the cornea and sclera is called limbus.

Cornea. The cornea is a transparent, avascular, watch-glass-like structure with a smooth shining surface. The average diameter of the cornea is 11–12 mm. Its thickness in the central part is 0.52 mm and in the peripheral part is 0.67 mm.

Sclera. The sclera is a strong, opaque, white fibrous layer. It is a relatively avascular structure about 1 mm in thickness. It is pierced by nerves and vessels entering in the eyeball.

2. The middle vascular coat

The middle vascular coat (Fig. 11.1-1) also known as uveal tract, from anterior to posterior, can be divided into three parts: iris, ciliary body and choroid. The blood supply of the uveal tract is derived from the short posterior ciliary arteries, long posterior ciliary arteries and anterior ciliary arteries.

Iris. Iris is a coloured, circular diaphragm with a central aperture of 3–4 mm size known as pupil. The pupil regulates the light reaching the retina. The pupil constricts and dilates by the contraction of sphincter pupillae and dilator pupillae muscles of the iris, respectively. The sphincter pupillae is supplied by the parasympathetic nerves, while the dilator pupillae is supplied by the sympathetic nerves.

Ciliary body. The ciliary body is the middle part of the uveal tract. In cut section, it is triangular in shape with base forwards. Anteriorly, the iris is attached to about the middle of the base of ciliary body. Posteriorly, the ciliary body becomes continuous with the choroid.

The ciliary body contains a non-striated muscle called the ciliary muscle which is supplied by parasympathetic fibres and takes part in the process of accommodation of the eye.

There are about 70–80 finger-like projections from the ciliary body. These are called *ciliary processes* and are the site of aqueous humour production—a watery fluid which maintains the intraocular pressure of the eyeball.

Choroid. Choroid is a dark brown highly vascular layer situated in between the sclera and retina. It supplies nutrition to the outer layers of retina.

Note. The inflammations of choroid invariably involve the underlying retina.

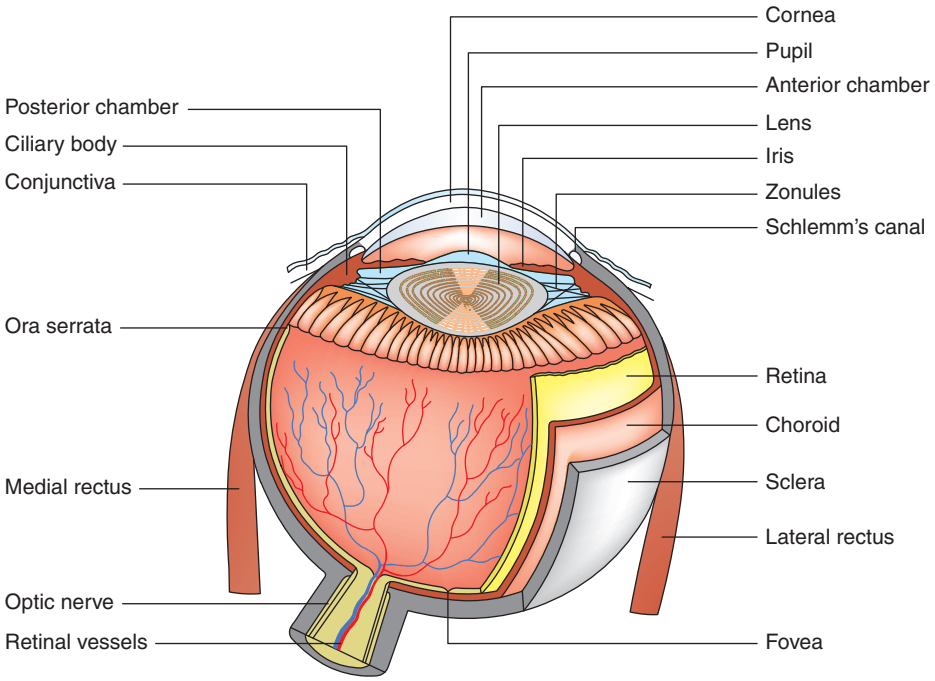


Fig. 11.1-1 Gross anatomy of eyeball.

3. The inner nervous coat (retina)

Retina, the innermost tunic of the eyeball, is a thin, delicate, transparent membrane. It is the most highly developed tissue of the eye. It is concerned with the visual functions (details on page 898).

Interior of the eyeball

Interior of the eyeball consists of anterior and posterior chambers containing the aqueous humour, the lens and the vitreous.

Anterior and posterior chambers

Anterior chamber is the space bounded anteriorly by the back of cornea and posteriorly by the anterior surface of iris. *Posterior chamber* is the space between the front of crystalline lens and the back of iris. Through pupil, anterior and posterior chambers communicate with each other. *Aqueous humour* is a watery fluid present in the anterior and posterior chambers of the eyeball.

MAINTENANCE OF CLEAR REFRACTIVE MEDIA OF THE EYE

The main prerequisite for visual function is the maintenance of clear refractive media of the eye. The major factor responsible for transparency of the ocular media is their avascularity. The structures forming refractive media of the eye from anterior to posterior are:

- Tear film,
- Cornea,
- Aqueous humour (see page 920),
- Crystalline lens and
- Vitreous humour.

PHYSIOLOGY OF TEARS

Lacrimal apparatus

The lacrimal apparatus (Fig. 11.1-2) comprises the structures concerned with the formation (main lacrimal gland and accessory lacrimal glands) and drainage (lacrimal passages: puncta, canaliculi, lacrimal sac and nasolacrimal duct) of tears.

Tear film and its functions

Tear film. Tear film refers to the fluid covering the cornea and conjunctiva. Tears are composed of 98% water and 1.5% sodium chloride (which give the tears their salty flavour). It also contains antibacterial substances like lysozyme, beta lysin and lactoferrin.

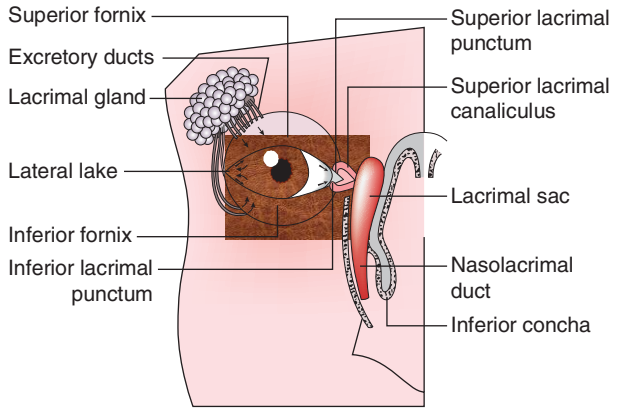


Fig. 11.1-2 The lacrimal apparatus.

Functions of tear film

1. It keeps the cornea and conjunctiva moist.
2. It provides oxygen to the corneal epithelium.
3. It washes away debris and noxious irritants.
4. It prevents infection due to presence of antibacterial substances.
5. It facilitates movements of the lids over the globe.

PHYSIOLOGY OF CORNEA

Cornea forms the main refracting medium of the eye. It is a transparent watch-glass-like structure, the anterior surface of which is bathed with tears and endothelial surface is bathed in the aqueous humour.

Corneal transparency

The main physiologic function of the cornea is to act as a major refracting medium, so that a clear retinal image is formed. Maintenance of corneal transparency of high degree is a prerequisite to perform this function.

Factors responsible for corneal transparency are:

- Avascularity of cornea,
- Absence of pigment in the cornea,
- A peculiar regular arrangement of the stromal lamellae and
- Relative dehydration of stroma.

PHYSIOLOGY OF CRYSTALLINE LENS

Structure of lens

The lens is a transparent, biconvex, crystalline structure. Its diameter is 9–10 mm and thickness varies with age from 3.5 mm (at birth) to 5 mm (at extreme of age). It consists of (Fig. 11.1-3):

1. Lens capsule. It is a thin, transparent, hyaline membrane surrounding the lens.

2. Anterior epithelium. It is a single layer of cuboidal cells, which lies deep to the anterior capsule.

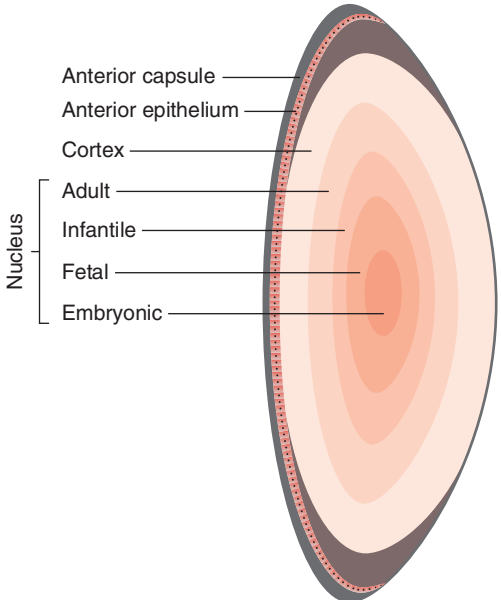


Fig. 11.1-3 Structure of crystalline lens.

3. Lens fibres. These form the main bulk of the lens and are arranged compactly as nucleus and cortex of the lens.

- Nucleus is the central part containing the oldest fibres.
- *Cortex* is the peripheral part which comprises the youngest fibres.

4. Suspensory ligaments of lens (zonules of Zinn). These consist essentially of a series of fibres by which lens is suspended from the ciliary body.

Lens transparency

Factors that play a significant role in maintaining outstanding clarity and transparency of lens are:

- Avascularity,
- The arrangement of lens protein,
- Auto-oxidation, high concentration of reduced glutathione (GSH) in the lens maintains the lens proteins in a reduced state and ensures transparency.

Metabolism

Source of nutrient supply. The crystalline lens, being an avascular structure is dependent for its metabolism on chemical exchanges with the aqueous humour.

Pathways of glucose metabolism. Glucose is very essential for the normal working of the lens. In the lens, 80% glucose is metabolised anaerobically by the glycolytic pathway, 15% by pentose hexose monophosphate shunt and a small proportion via oxidative Krebs' citric acid cycle.

APPLIED ASPECTS

Cataract. Any opacity in the lens or its capsule is called cataract. Three basic mechanisms which cause cataract are:

- Damage to the lens capsule that changes its membranous properties,
- Change in the lens fibre protein synthesis.

Senile cataract. The main biochemical changes in senile cataract occur are decreased levels of total proteins, amino acids and potassium associated with an increased concentration of sodium and marked hydration of the lens, followed by coagulation of proteins.

PHYSIOLOGY OF VITREOUS HUMOUR

Vitreous humour is an inert, transparent, colourless, jelly-like structure that fills the posterior four-fifth of the cavity of eyeball and is about 4 mL in volume.

Structure. The normal youthful vitreous gel is composed of a network randomly oriented collagen fibrils interspersed with numerous spheroidal macromolecules of hyaluronic acid.

Functions. The vitreous gel mainly serves the *optical function*. In addition, it mechanically stabilizes the shape and volume of globe, and is a pathway for nutrients to reach the lens and retina.

THE IMAGE FORMING MECHANISM

The functioning of the eye as an optical instrument can be compared with a close-circuit colour television camera (Fig. 11.1-4) in which:

Eyelids act as a shutter of the camera.

Cornea and crystalline lens act as a focussing system of the camera.

Iris acts as a diaphragm, which regulates the size of the aperture (pupil) and therefore the amount of light entering the eye.

Choroid and pigment epithelium of retina help in forming the darkened interior of the camera.

Neural retina acts as a light-sensitive plate or film on which images of the objects in the environment are focussed. The light rays striking the retina generate potentials in the rods and cones. Thus the eye converts energy in the visible spectrum into action potentials in the optic nerve.

Optic nerve and its connections convey the impulses generated in the retina to the occipital region of the cerebral cortex where they produce sensation of vision.

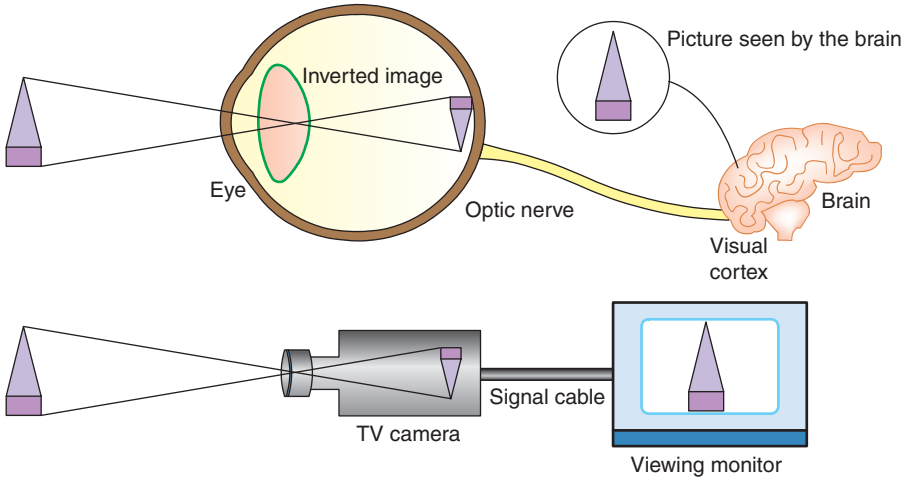


Fig. 11.1-4 The sense of sight in many ways is similar to a close-circuit colour TV system. It is superior in all respects except ease of replacement.

To understand the image forming mechanism of eye (i.e. optics of eye) and its abnormalities, it is imperative to have some knowledge about the light and geometrical optics.

PRINCIPLES OF OPTICS

LIGHT

Light is the visible portion of the electromagnetic radiation spectrum. It lies between ultraviolet and infrared portions, from 400 nm at the violet end of the spectrum to 700 nm at the red end. The white light consists of seven colours denoted by VIBGYOR (violet, indigo, blue, green, yellow, orange and red).

Light ray is the term used to describe the radius of the concentric wave forms. A group of parallel rays of light is called a *beam of light*.

Reflection of light

Reflection of light is a phenomenon of change in the path of light rays without any change in the medium (Fig. 11.1-5). The light rays falling on a reflecting surface are called *incident rays* and those reflected by it are *reflected rays*. A line drawn at right angle to the surface is called the *normal*.

Laws of reflection are (Fig. 11.1-5):

1. The incident ray, the reflected ray and the normal at the point of incident, all lie in the same plane.
2. The angle of incidence is equal to the angle of reflection.

Refraction of light

Refraction of light is the phenomenon of change in the path of light, when it goes from one medium to another. The basic cause of refraction is change in the velocity of light when passing from one medium to the other.

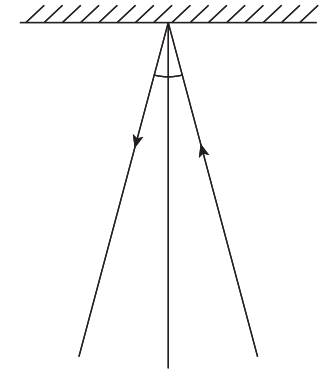


Fig. 11.1-5 Reflection of light.

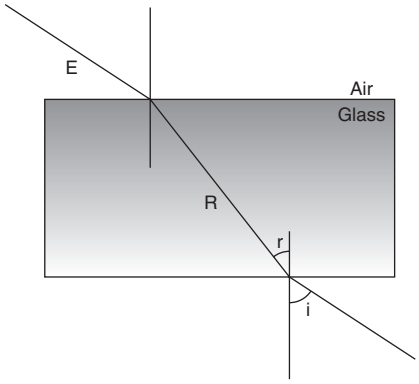


Fig. 11.1-6 Laws of refraction. N_1 and N_2 (normals); I (incident ray); i (angle of incidence); R (refracted ray, bent towards normal); r (angle of refraction); E (emergent ray, bent away from the normal).

Laws of refraction are (Fig. 11.1-6):

1. The incident and refracted rays are on the opposite sides of the normal and all the three are in the same plane.
2. The ratio of sine of angle of incidence to the sine of angle of refraction is constant for the part of media in contact.

This constant is denoted by the letter n and is called *'refractive index'* of the medium 2 in which the refracted ray lies with respect to medium 1 (in which the incident ray lies), i.e., sin i/sin $r = {}'n_2$. When the medium 1 is air (or vacuum), then n is called the refractive index of the medium 2.

Lenses

A lens is a transparent refracting medium, bounded by *two* surfaces which form a part of a sphere (spherical lens) or a cylinder (cylindrical or toric lens).

Cardinal data of a lens (Fig. 11.1-7)

1. *Centre of curvature* (C) of the spherical lens is the centre of the sphere of which the refracting lens surface is a part.
2. *Radius of curvature* of the spherical lens is the radius of the sphere of which the refracting surface is a part.
3. *The peripheral axis* (AB) of the lens is the line joining the centres of curvatures of its surfaces.
4. *Optical centre* (O) of the lens corresponds to the nodal point of a thick lens. It is a point on the principal axis in the lens, the rays passing from where do not undergo deviation.
5. *The principal focus* (F) of a lens is that point on the principal axis where parallel rays of light, after passing through the lens, converge (in convex lens) or appear to diverge (in concave lens).
6. *The focal length* (f) of a lens is the distance between the optical centre and the principal focus.
7. *Power of a lens* (P) is defined as the ability of the lens to converge a beam of light falling on the lens. For a converging (convex) lens the power is taken as positive and for a diverging (concave) lens power is taken as negative. It is measured as a reciprocal of the focal length in metres, i.e. P = 1/f. The unit of power is dioptre (D). One dioptre is the power of a lens of focal length 1 m.

Types of lenses

Lenses are of two types: the spherical and cylindrical (tonic or astigmatic).

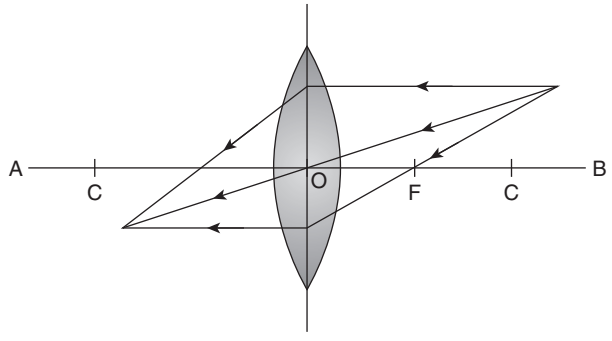


Fig. 11.1-7 Cardinal points of a convex lens: optical centre, O; principal focus, F; centre of curvature, C and principal axis, AB.

1. Spherical lenses. Spherical lenses are bounded by two spherical surfaces and are mainly of two types: convex and concave.

(a) Convex lens or plus lens is a converging. It may be of biconvex, plano-convex or concavo-convex (meniscus) type (Fig. 11.1-8).

Identification of a convex lens

(i) The convex lens is thick on the centre and thin at the periphery.
(ii) An object held close to the lens appears magnified.
(iii) When a convex lens is moved, the object seen through it moves in the opposite direction to the lens.

(b) Concave lens or minus lens is a diverging lens. It is of three types: biconcave, plano-concave and convexo-concave (meniscus) (Fig. 11.1-9).

Identification of concave lens

(i) It is thin at the centre and thick at the periphery.
(ii) An object seen through it appears minified.
(iii) When the lens is moved, the object seen through it moves in the same direction as the lens.

Uses of concave lens. It is used (i) for correction of myopia; (ii) as Hruby lens for the fundus examination with a slit-lamp.

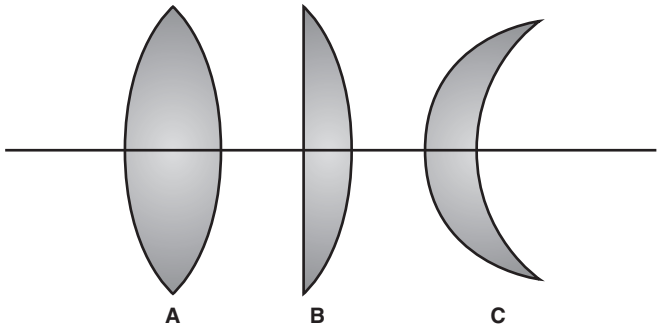


Fig. 11.1.8 Basic forms of a convex lens: A, biconvex; B, plano-convex and C, concavo-convex.

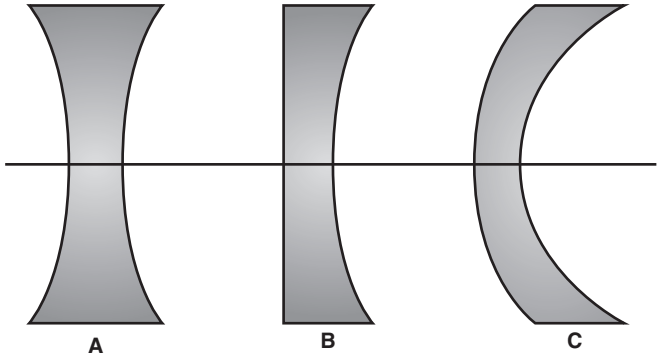


Fig. 11.1-9 Basic forms of a concave lens: A, biconcave; B, plano-concave and C, convexo-concave.

2. Cylindrical lens. A cylindrical lens acts only in one axis, i.e. power is incorporated in one axis, the other axis having zero power. A cylindrical lens may be convex (plus) or concave (minus) (Fig. 11.1-10). The axis of a cylindrical lens is parallel to that of the cylinder of which it is a segment. The cylindrical lens has a power only in the direction at right angle to the axis. Therefore, the parallel rays of light after passing through a cylindrical lens do not come to a point focus but form a focal line (Fig. 11.1-11).

Identification of a cylindrical lens. When the cylindrical lens is rotated around its optical axis, the object seen through it becomes distorted.

Uses. Cylindrical lenses are prescribed to correct astigmatism.

OPTICS OF THE EYE

As an optical instrument, the focusing system of eye is composed of several refracting structures. The refractive indices of the media of eye are:

Cornea:	1.37
Aqueous humour:	1.33
Crystalline lens:	1.42
Vitreous humour:	1.33

These constitute a homocentric system of lenses, which when combined in action form a very strong refracting system of a short focal length. The total dioptric power of the eye is about +60 D out of which about +44 D is contributed by cornea and +16 D by the crystalline lens.

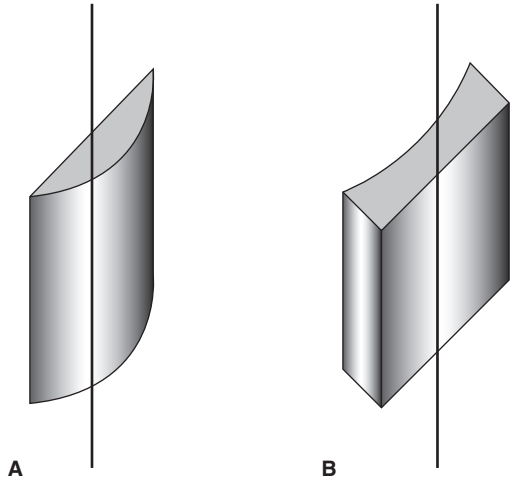


Fig. 11.1-10 Cylindrical lenses: A, convex and B, concave.

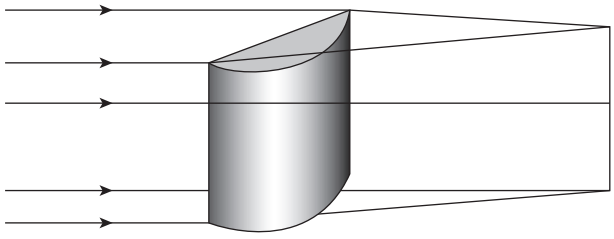


Fig. 11.1-11 Refraction through a convex cylindrical lens.

The reduced eye

The optics of eye otherwise is very complex. However, for understanding, Listing has simplified the data by choosing single principal point and single nodal point lying midway between two principal points and two nodal points, respectively. This is called *Listing's reduced eye*. The simplified data of this eye (Fig. 11.1-12) are:

- Total dioptric power +60 D.
- The principal point (P) lies 1.5 mm behind the anterior surface of cornea.
- The nodal point (N) is situated 7.2 mm behind the anterior surface of cornea.
- The anterior focal point is 15.7 mm in front of the anterior surface of cornea.
- The posterior focal point (on the retina) is 24.4 mm behind the anterior surface of cornea.
- The anterior focal length is 17.2 mm (15.7 + 1.5) and the posterior focal length is 22.9 mm (24.4 – 1.5).

Axes of the eye

The eye has three principal axes (Fig. 11.1-13):

1. The *optical axis* is the line passing through the centre of the cornea (P), centre of the lens (N) and meets the retina (R) on the nasal side of the fovea.

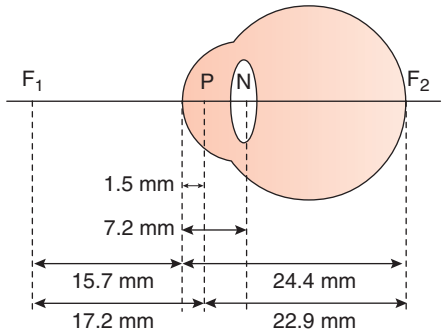


Fig. 11.1-12 Cardinal points of Listing's reduced eye.

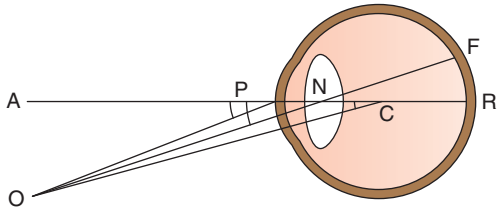


Fig. 11.1-13 Axes of the eye: AR, optical axis; OF, visual axis and OC, fixation axis.

2. The *visual axis* is the line joining the fixation point (O), nodal point (N) and the fovea (F).
3. The *fixation axis* is the line joining the fixation point (O) and the centre of rotation (C).

ACCOMMODATION

Definition of accommodation and related terms

Accommodation

As we know that in an emmetropic eye, parallel rays of light coming from infinity are brought to focus on the retina with accommodation at rest. Our eyes have been provided with a unique mechanism by which we can even focus the diverging rays coming from a near object on the retina in a bid to see clearly (Fig. 11.1-14). This mechanism is called accommodation. In it there occurs an increase in the power of the crystalline lens.

Far point, near point, range and amplitude of accommodation

The nearest point at which small objects can be seen clearly is called near point or *punctum proximum* and the distant (farthest) point is called far point or *punctum remotum*. The distance between the near point and the far point is called *range of accommodation*. The difference between the dioptric power needed to focus at near point (P) and to focus at far point (R) is called *amplitude of accommodation* (A). Thus, A = P – R.

In an emmetropic eye, far point is at infinity and near point varies with age (Table 11.1-1). Thus the amount that the eye can alter its refraction is greatest in childhood and slowly decreases with age.

Mechanism of accommodation

As we know, accommodation is a process by which one can focus the objects at different distances in a bid to have a clear vision. Its mechanism varies from species to species. Just for an interest examples of a few species are:

- Some fishes retract their lenses to focus on distant objects.
- Snakes and frogs have a mechanism to move the lens forward for near vision.
- Horses, by moving their heads, tilt the retina so that different regions lie at appropriate distances behind the lens.

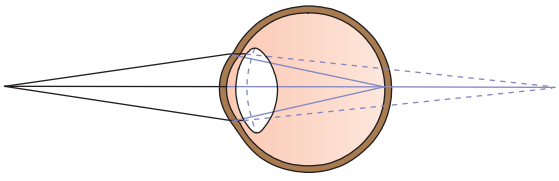


Fig. 11.1-14 Effect of accommodation on divergent rays entering the eye.

- In man, the process of accommodation is achieved by a change in the shape of the lens.

Ocular changes in accommodation

The changes which take place in the eye during accommodation are:

Slackening of the zonules. Zonules are normally tense and keep the lens flat. They slacken during accommodation due to contraction of the ciliary muscle.

Changes in the curvature of lens surface. The principal change in the lens during accommodation is seen in the anterior surface of the lens (Fig. 11.1-15). At rest, the radius of curvature of the anterior surface of the lens is 11 mm and

Table 11.1-1 Near point and amplitude of accommodation at different ages

Age (years)	Near point (cm)	Amplitude of accommodation (dioptre)
10	7	14.0
20	9	11.0
30	12	8.0
40	22	4.5
45	28	3.5
50	40	2.5
60	85	1.5
70	100	1.0

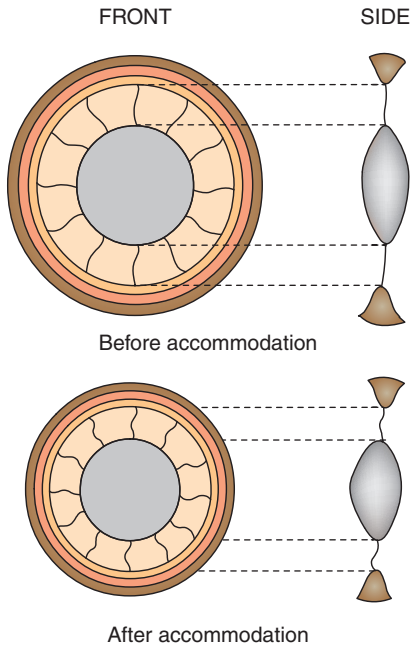


Fig. 11.1-15 Changes in the ciliary body ring zonules and shape of lens during accommodation.

that of posterior surface is 6 mm. In accommodation, the curvature of posterior surface remains almost the same, but the anterior surface changes, so that in strong accommodation its radius of curvature becomes about 6 mm in the periphery and 3 mm in the central part which bulges more.

Pupillary constriction and convergence of eyes. In addition to the changes in the lens and zonular system, the pupil constricts and the eyes converge almost simultaneously. These changes occur in a bid to achieve clear vision for near objects.

OPTICAL ABERRATIONS OF THE EYE

The eye, in common with many optical systems in practical use, is by no means optically perfect; the lapses from perfection are called aberrations. Physiological optical defects in a normal eye include the following:

1. Diffraction of light

Diffraction is bending of light caused by the edge of an aperture or the rim of a lens. Even a perfect lens, free from aberrations, will not focus light to a point due to diffraction. The actual pattern of a diffracted image point produced by a lens with a circular aperture or pupil is a series of concentric bright and dark rings.

2. Spherical aberrations

Spherical aberrations occur owing to the fact that spherical lens refracts peripheral rays more strongly than the paraxial rays, which in the case of a convex lens brings the more peripheral rays to focus closer to the lens (Fig. 11.1-16).

The factors which contribute in diminishing the spherical aberrations of human eye are:

- Peculiar curvature of the cornea, i.e. flatter periphery than the centre.
- Peculiar structure of the crystalline lens wherein the central portions have a greater density and are arranged in layers of greater curvature than the peripheral portion.
- Iris blocks the peripheral rays to enter the eye and thus in ordinary circumstances, refraction of only paraxial rays of light takes place.

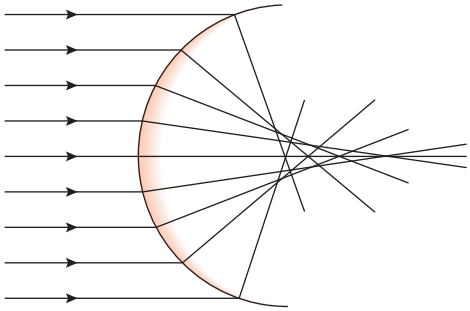


Fig. 11.1-16 Spherical aberration.

3. Chromatic aberrations

Chromatic aberrations result owing to the fact that the index of refraction of any transparent medium varies with the wavelength of incident light. In human eye, which optically acts as a convex lens, blue light is focused slightly in front of the red (Fig. 11.1-17). In other words, the emmetropic eye is, in fact, slightly hypermetropic for red rays and myopic for blue and green rays.

COMMON DEFECTS OF THE IMAGE FORMING MECHANISM

Emmetropia

Emmetropia (optically normal eye) can be defined as a state of refraction when the parallel rays of light coming from infinity are focused at the sensitive layer of retina with the accommodation being at rest (Fig. 11.1-18).

Ametropia

Ametropia (a condition of refractive error) is defined as a state of refraction when the parallel rays of light coming from infinity (with accommodation at rest) are focused either in front or behind the sensitive layer of retina, in one or both the meridia. The ametropia includes myopia, hypermetropia and astigmatism.

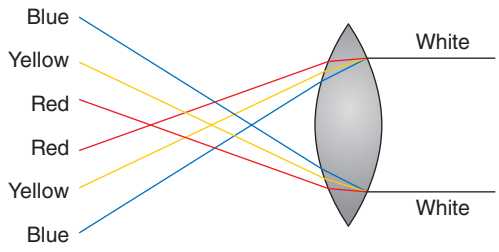


Fig. 11.1-17 Chromatic aberration.

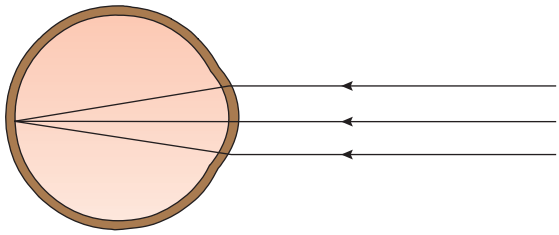


Fig. 11.1-18 Refraction in an emmetropic eye.

Hypermetropia

Hypermetropia (hyperopia) or long sightedness is the refractive state of the eye wherein parallel rays of light coming from infinity are focused behind the retina with accommodation being at rest (Fig. 11.1-19).

Mechanism of production

Aetiologically, hypermetropia may be axial, curvatural, index, positional and due to absence of lens.

1. Axial hypermetropia is by far the most common form. In this condition, there is an axial shortening of eyeball. About 1 mm shortening of the anteroposterior diameter of the eye results in 3 dioptres of hypermetropia.

2. Curvatural hypermetropia is the condition in which the curvature of cornea, lens or both is flatter than the normal. About 1 mm increase in radius of curvature results in 6 dioptres of hypermetropia.

3. Index hypermetropia occurs due to the change in refractive index of the lens in old age. It may also occur in diabetics under treatment.

4. Positional hypermetropia results from posteriorly placed crystalline lens.

5. Absence of crystalline lens either congenitally or acquired (following surgical removal or posterior dislocation) leads to aphakia—a condition of high hypermetropia.

Characteristic features

- *Far-sightedness.* Persons with mild to moderate hypermetropia, in their young age, can see the distant objects clearly using their accommodation. This is why hypermetropia is also called far sightedness or long sightedness.
- *Near point* of vision moves further away and the patient may have sometimes problem in near vision, when most of the accommodation is used for correcting for vision. Because of this, hypermetropia requires presbyopic correction at younger age.

Optical correction

Basic principle of treatment of hypermetropia is optical correction with convex (plus) lenses, so that the light rays are brought to focus on the retina (Fig. 11.1-20).

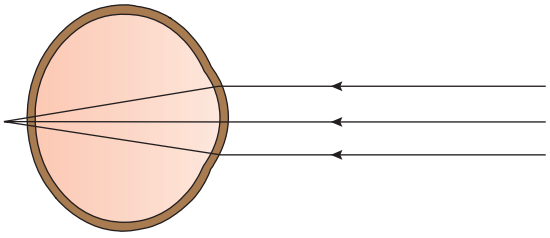


Fig. 11.1-19 Refraction in a hypermetropic eye.

Myopia

Myopia or short-sightedness is a type of refractive error in which parallel rays of light coming from infinity are focused in front of the retina when accommodation is at rest (Fig. 11.1-21).

Mechanisms of production

1. Axial myopia results from an increase in the anteroposterior length of the eyeball. It is the most common form.

2. Curvatural myopia occurs due to increased curvature of the cornea, lens or both.

3. Positional myopia is produced by anterior placement of crystalline lens in the eye.

4. Index myopia results from an increase in the refractive index of crystalline lens associated with nuclear sclerosis.

5. Myopia due to excessive accommodation occurs in patients with spasm of accommodation.

Characteristic features

- *Short-sightedness.* Far point of vision is a finite point in front of the eye (at infinity in emmetropes). Therefore, the myopic persons cannot see the distant objects. This is why myopia is also called *short sightedness.*

Optical correction

Basic principle of treatment of myopia is optical correction with concave (minus) lenses, so that clear image is formed on the retina (Fig. 11.1-22).

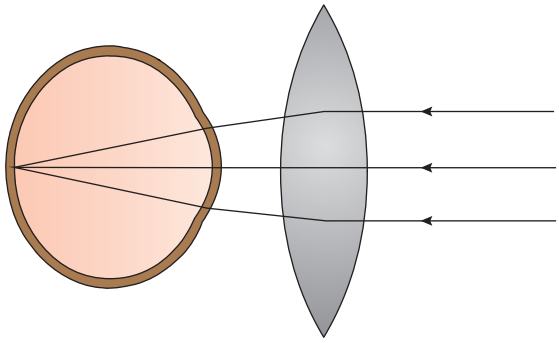


Fig. 11.1-20 Refraction in a hypermetropic eye corrected with convex lens.

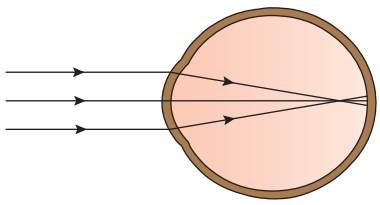


Fig. 11.1-21 Refraction in a myopic eye.

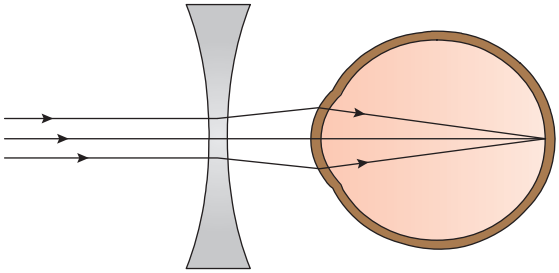


Fig. 11.1-22 Refraction in a myopic eye corrected by concave lens.

IMPORTANT NOTE

A myopic patient may not need glasses for near vision in old age, because his near point may be at a reading distance (which recedes back in emmetropic presbyopes).

Astigmatism

Astigmatism is a type of refractive error wherein the refraction varies in the different meridia. Consequently, the rays of light entering the eye cannot converge to a point focus but form focal lines. Broadly, there are two types of astigmatism: regular and irregular.

Regular astigmatism

The astigmatism is regular when the refractive power changes uniformly from one meridian to another (i.e. there are two principal meridia).

Treatment. Optical treatment of regular astigmatism comprises the prescribing appropriate cylindrical lens.

Irregular astigmatism

It is characterized by an irregular change of refractive power in different meridia. There are multiple meridia which admit no geometrical analysis.

Treatment. Optical treatment of irregular astigmatism consists of contact lens which replaces anterior surface of cornea for refraction.

Presbyopia

Presbyopia (eyesight of old age) is not an error of refraction but condition of physiological insufficiency of accommodation, leading to failing vision for near due to: (i) Decrease in the elasticity and plasticity of the crystalline lens (which results from age-related sclerosis). (ii) Age-related decrease in the power of ciliary muscles. To understand the condition of presbyopia, a working knowledge about accommodation is mandatory, see page 894.

Since, we usually keep the book at about 25 cm, so we can read comfortably up to the age of 40 years. After the age of 40 years, the near point of accommodation recedes beyond the normal reading or working range. This condition of failing near vision due to age-related decrease in the amplitude of accommodation is called presbyopia.

Symptoms

- Difficulty in near vision (to start with in the evening and in dim light and later even in good light).
- Asthenopic symptoms due to fatigue of the ciliary muscle are also complained after reading or doing any near work.

Treatment

The treatment of presbyopia is the prescription of appropriate convex glasses for near work.

PHYSIOLOGY OF VISION

Physiology of vision is a complex phenomenon which is still poorly understood.

The main mechanisms concerned with vision are:

- *Initiation of vision* (phototransduction), a function of photoreceptors (rods and cones),
- *Processing and transmission of visual sensation,* a function of the image processing cells of retina and visual pathway.
- *Visual perceptions,* a function of visual cortex and related areas of cerebral cortex. It is based on the activities of serial processing stations in the visual pathway and parallel processing pathways.

For the purpose of understanding, the description of physiology of vision can be organized as:

- Retina, photoreceptors and visual pigment,
- Phototransduction,
- Processing and transmission of visual impulse in retina,
- Processing and transmission of visual impulse in visual pathway,
- Processing and analysis of visual impulse in the visual cortex and
- Concept of serial and parallel processing of visual information.

RETINA, PHOTORECEPTORS AND VISUAL PIGMENTS

RETINA

Gross anatomy

Retina, the innermost tunic of the eyeball is a thin, transparent membrane. It is concerned with the visual functions.

Grossly, retina exhibits three distinct areas: optic disc, macula lutea and peripheral retina (Fig. 11.1-23):

Optic disc. It is a well-defined, circular, pink coloured disc of 1.5 mm diameter. It has only nerve fibre layer, so it does not excite any visual response. It produces *blind spot* in the field of vision.

Macula lutea (yellow spot). It is a comparatively dark area situated at the posterior pole temporal to the optic disc. Its central depressed area 1.5 mm in diameter is called *fovea centralis,* which is the most sensitive part of the retina. Visual acuity is maximum in this part of the retina.

Ora serrata. It is the anterior serrated margin where the retina ends.

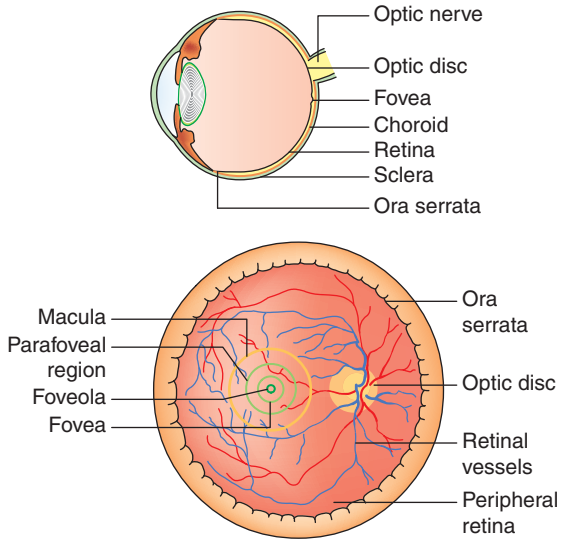


Fig. 11.1-23 Gross anatomy of retina.

Microscopic structure

Retina consists of ten layers, which from without inwards are (Fig. 11.1-24):

1. Layer of pigment epithelium. It is a single layer of hexagonal cells containing melanin pigments. It serves following *functions*:

- Absorbs stray light and thereby reduces light scatter.
- Phagocytose the ends of the outer segments of rods which are continuously shed.
- Reconvert the metabolized photopigment into a form that can be reused after it is transported back to the photoreceptor.
- Tight junction between the cells form outer blood–retinal barrier.

2. Layer of rods and cones. It consists of the outer segments of the photoreceptors (rods and cones). Photoreceptors are the end organs of vision.

3. External limiting membrane. It is not a separate membrane. In fact, the numerous connections made between Muller cells and inner segments of photoreceptors give the appearance of a continuous membrane under light microscopy.

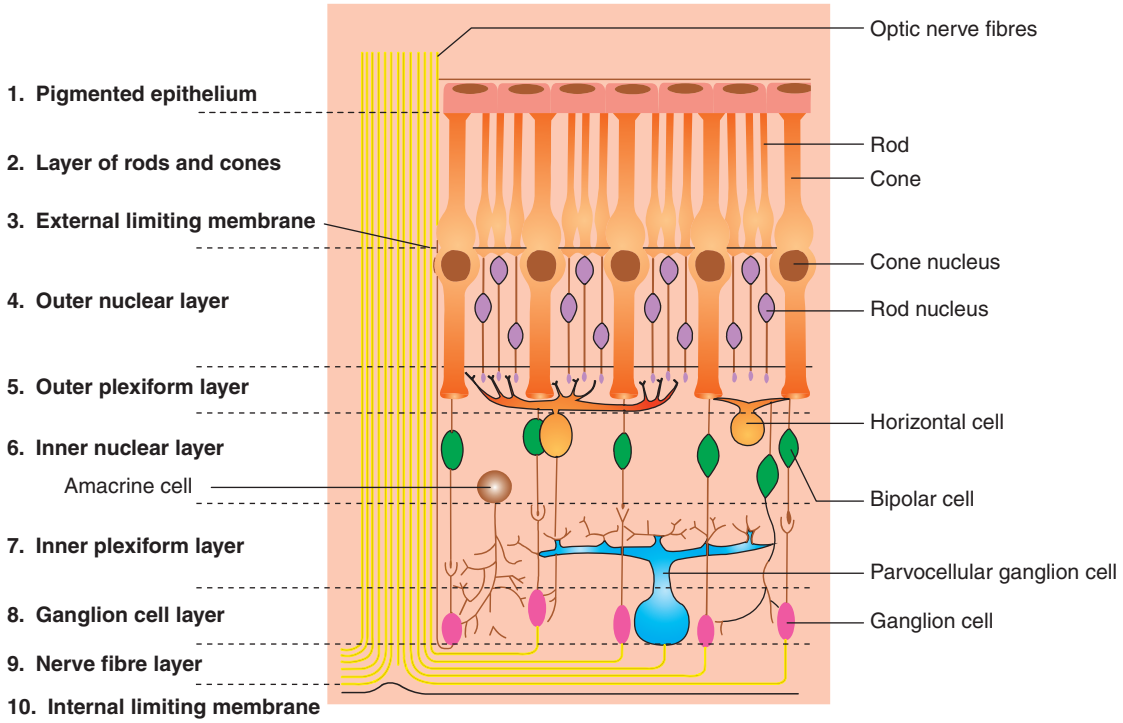


Fig. 11.1-24 Microscopic structure of retina.

4. Outer nuclear layer. This layer contains the nuclei of rods and cones.

5. Outer plexiform layer. This layer contains pre-synaptic and post-synaptic elements of synapses that exist between the photoreceptors, bipolar cells and horizontal cells.

6. Inner nuclear layer. It contains the cell bodies and nuclei of bipolar cells, amacrine cells and horizontal cells.

7. Inner plexiform layer. It is the layer of synapse between bipolar cells, ganglion cells and amacrine cells.

8. Ganglion cell layer. It consists of ganglion cells, which are the output cells of the retina. They transmit visual information to the brain.

9. Nerve fibre layer. It consists of the axons of ganglion cells which pass through lamina cribrosa to form the optic nerve. These fibres remain unmyelinated in the retina, but become myelinated in the optic nerve.

10. Inner limiting membrane. It is formed by projections of the Muller's cells and separates the retina from vitreous.

Structural characteristics of fovea centralis

Foveal region has the highest visual resolution because of following structural characteristics:

- Rods are absent and cone density is maximum.
- The most central part of fovea (foveola) is devoid of even capillaries, while the rest of fovea contains fine capillaries but no large vessels which encircle this area.
- There is no convergence of efferents of the foveal cones. Each foveal cone relays to single ganglion cell. Hence, there is a disproportionate large representation of the fovea in the visual cortex.

Blood supply

- *Outer four layers of the retina*, viz., pigment epithelium, layer of rods and cones, external limiting membrane and outer nuclear layer get their nutrition from the choroidal vessels.
- *Inner six layers* get their supply from the central retinal artery, which is a branch of the ophthalmic artery.
- *Central retinal artery* emerges from the centre of the physiological cup of the optic disc and divides into four branches, namely the superior-nasal, superior-temporal, inferior-nasal and inferior-temporal. These are end arteries, i.e. they do not anastomose with each other.
- *The retinal veins*. These follow the pattern of the retinal arteries. The central retinal vein drains into the cavernous sinus directly or through the superior ophthalmic vein.

PHOTORECEPTORS

Density and distribution of photoreceptors

- Rods and cones *(photoreceptors)* are the end organs of vision which transform light energy into visual (nerve) impulse.
- Rods contain a photosensitive substance visual purple *(rhodopsin)* and subserve the peripheral vision and vision of low illumination (scotopic vision).
- Cones also contain a photosensitive substance and are primarily responsible for highly discriminatory central vision (photopic vision) and colour vision.
- There are about 120 million rods and 6.5 million cones.
- The highest density of cones is at fovea with an average of 199,000 cones/mm^2. The number of cones falls off rapidly outside the fovea.
- Rods are absent at the fovea in an area of 0.35 mm (rod-free zone) which corresponds to 1.25° of the visual field; but are present in a large number (160,000/mm^2) in a ring-shaped zone, 5–6 mm from the fovea.

IMPORTANT NOTE

The photoreceptors (rods and cones) get their nourishment from choroidal papillary plexus, therefore, in retinal detachment the receptor cells suffer most and leads to blindness.

Structure of photoreceptor

Each *photoreceptor* consists of a cell body and a nucleus (which lie in the outer nuclear layer), a cell process that extends into outer plexiform layer and inner and outer segments (which form the layer of rods and cones) (Fig. 11.1-25). The long axis of the photoreceptor is oriented perpendicular to the retinal surface.

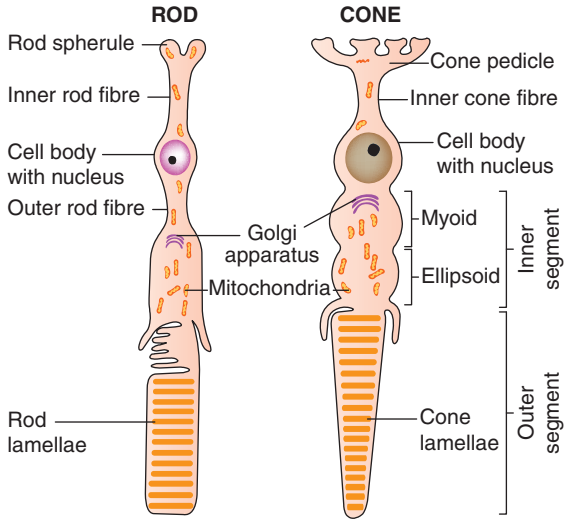


Fig. 11.1-25 Microscopic structure of rod and cone cells.

The rod cell

Each rod is about 40–60 µm long.

The *outer segment* of the rod is cylindrical, highly refractile, transversely striated and contains visual purple. It is composed of numerous lipid protein lamellar discs stacked one on top of the other and surrounded by a cell membrane.

The *inner segment* of the rod is thicker than the outer segment. It consists of two regions: ellipsoid and myoid. *An outer rod fibre* arises from the inner end of rod, which passes through the external limiting membrane and swells into a densely staining nucleus—the rod granule (lies in the outer nuclear layer); and then terminates as *inner rod fibre* (lies in the outer molecular layer) which, at its end has got an end bulb called the rod spherules that are in contact with the cone foot.

The cone cell

- Each cone cell is 40–80 µm long. It is longest at the fovea (80 µm) and shortest at the periphery (40 µm).
- The *cone outer segment* is conical in shape, much shorter than that of rod and contains the iodopsin. The lamellar discs, which are narrower than those of the rods, are, in fact, infoldings of plasma membrane. There are about 1000–1200 discs/cone.
- The *cone inner segment* and cilium are similar to the rod structures; however, the cone ellipsoid is very plump and contains a large number of mitochondria.
- Unlike rod the inner segment of the cone becomes directly continuous with its nucleus and lies in outer nuclear layer. A stout cone inner fibre runs from the nucleus which at the end is provided with lateral processes called *cone foot* or *cone pedicle* (lies in the outer plexiform layer).

VISUAL PIGMENTS

Visual pigments are those substances which have the property of absorbing light. These include *rhodopsin* and cone pigments.

Rhodopsin (visual purple)

Rhodopsin is the photosensitive visual pigment present in the discs of the rod outer segments. It consists of a protein opsin (called scotopsin) and a carotenoid called retinal or retinene$_1$ (the aldehyde of vitamin A).

Human rhodopsin has a molecular weight of 40,000. It is one of the many serpentine receptors coupled to G proteins.

The absorption spectrum of rhodopsin, as shown in Fig. 11.1-26, depicts that its peak sensitivity to light lies within the narrow limits of 493–505 nm. It absorbs primarily yellow wavelength of light, transmitting violet and red to appear purple by transmitted light; it is therefore also called visual purple.

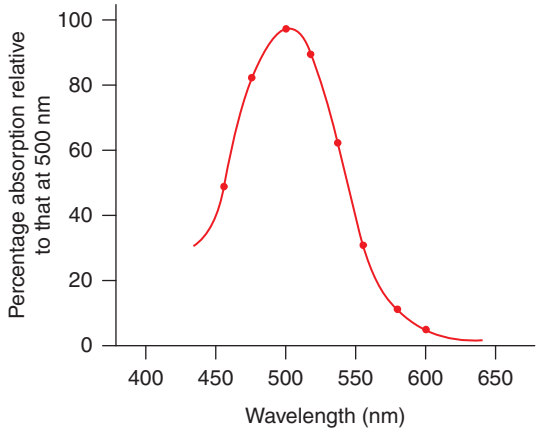


Fig. 11.1-26 Absorption spectrum of rhodopsin.

Note the term retinene$_1$ is used to differentiate it from retinene$_2$ (a compound present in the eyes of animal species).

Cone pigments

The visual pigments present in the cones have not been so intensively studied as the rhodopsin. There are three kinds of cones in primates. Cone pigments are somewhat different from the rhodopsin, in that they respond to specific wavelength of light, giving rise to colour vision. These differences are present in the opsin portion of the molecule, whereas the chromophore 11-cis-retinal remains the same. The peak absorbance wavelength of the 'blue', 'green' and 'red' sensitive cones lie at about 440, 535 and 565 nm, respectively.

Light-induced changes

Light falling upon the retina is absorbed by the visual pigments and initiate *photochemical changes* which in turn trigger a sequence of events that cause phototransduction. The photochemical changes occurring in the rods and cones are similar, but they have been studied in detail in the rods and can be described under three headings:

- Rhodopsin bleaching,
- Rhodopsin regeneration and
- Visual cycle.

Rhodopsin bleaching. As mentioned earlier, the rhodopsin consists of a protein called *opsin* and a carotenoid called retinene (vitamin A aldehyde or 11-cis-retinal). The light absorbed by the rhodopsin converts its *11-cis-retinal* into *all-trans-retinal.* This light-induced isomerization of 11-cis-retinal into all-trans-retinal occurs through formation of many intermediates which exist for a transient period (Fig. 11.1-27). One of the intermediate compounds (*metarhodopsin II*, also called as activated rhodopsin) of the above isomerization chain reaction acts as an enzyme to

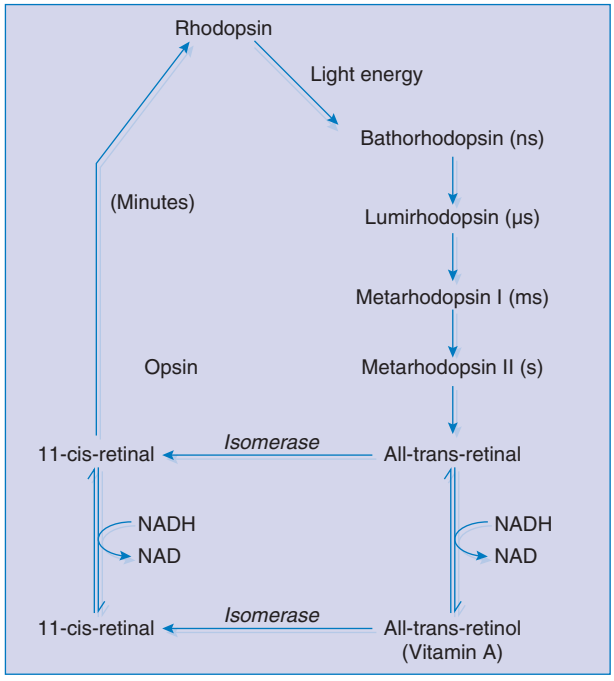


Fig. 11.1-27 Light-induced changes in rhodopsin.

activate many molecules of transducin. The activated transducin triggers the phototransduction.

The all-trans-retinal (produced from light-induced isomerization of 11-cis-retinal) can no longer remain in combination with the opsin and thus there occurs separation of opsin and all-trans-retinal. This process of separation is called *photodecomposition* and the rhodopsin is said to be bleached by the action of light.

Rhodopsin regeneration. The all-trans-retinal separated from the opsin (as above) subsequently enters into the chromophore pool existing in the photoreceptor outer segment and the pigment epithelial cells (for this, close approximation of retinal pigment epithelium (RPE) and photoreceptor is must). The all-trans-retinal may be further reduced to retinol by alcohol dehydrogenase, then esterified to re-enter the systemic circulation.

The first stage in the reformation of rhodopsin, as shown in Fig. 11.1-27, is isomerization of all-trans-retinal back to 11-cis-retinal. The process is catalyzed by the enzyme *retinal isomerase*. Energy for the regeneration process is supplied by the overall metabolic pool of the photoreceptor outer segment. The 11-cis-retinal in the outer segments of photoreceptors reunites with the opsin to form rhodopsin. This whole process is called regeneration of the rhodopsin. Thus the bleaching of the retinal photopigments occurs under the influence of light; whereas the regeneration process is independent of light, proceeding equally well in light or darkness. The amount of rhodopsin in the rods, therefore, varies inversely with the incident light.

Visual cycle. In the retina of living animals, under constant light stimulation, a steady state must exist under which the rate at which the photochemicals are bleached is equal to the rate at which they are regenerated. This equilibrium between the photodecomposition and regeneration of visual pigments is referred to as *visual cycle.*

IMPORTANT NOTE

It is important to note that a small number of photoreceptors contain no photopigment (rhodopsin or cone pigment) but they contain *melanopsin*. The axons of these neurons project to suprachiasmatic nuclei and lateral geniculate nuclei and regulate pupillary response to light and thus they are responsible for circadian responses to dark–light changes. When the gene for melanopsin is knocked out, the circadian responses are abolished.

PHOTOTRANSDUCTION

Phototransduction refers to the conversion of light energy into nerve impulse. It involves a cascade of biochemical reactions in following steps:

Activation of rhodopsin. As described above, following exposure to light the rhodopsin undergoes a series of spontaneous transformation, leading to formation of an active form of rhodopsin, the *metarhodopsin II.*

Activation of transducin. The activated rhodopsin acts as an enzyme to activate many molecules of transducin (G-protein). When transducin gets replaced by GTP and the α subunit separates.

Conversion of cGMP to GMP. The α subunit activates the many molecules of the enzyme phosphodiesterase which catalyses conversion of cGMP to GMP leading reduction in the concentration of cGMP within the photoreceptor.

Production of receptor potential. Reduction in the cGMP is responsible for producing receptor potential as explained (Fig. 11.1-28):

In dark, the Na^+ channels present in the cell membrane of the outer segment of photoreceptor are kept open by cGMP. So, a net influx of Na^+ results in a continuous current called *dark current.* The dark current causes the receptor cell to be maintained in a constant state of depolarization (the resting potential is about –40 mV). The intracellular Na^+ concentration is kept at a steady state level by sodium pump (Na^+–K^+–ATPase) located in the inner segment.

When light strikes the photoreceptor, the amount of cyclic GMP in the photoreceptor is reduced (as discussed in photochemistry of vision), so some of the Na^+ channels (which were kept open by cyclic GMP in dark) are closed, and the result is a hyperpolarizing receptor potential.

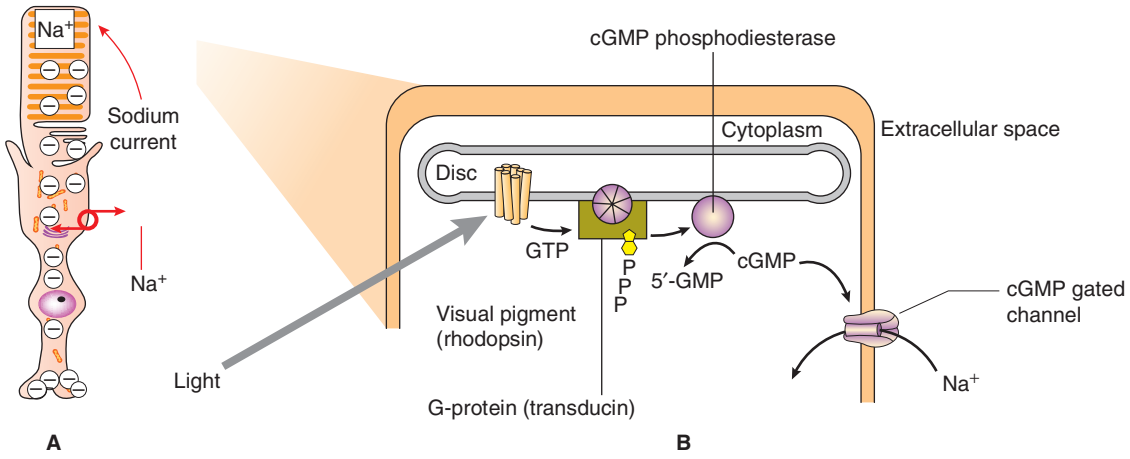


Fig. 11.1-28 Potential changes in a photoreceptor: A, in dark Na^+ channels are opened by the cGMP and due to Na^+ influx (dark current) results and membrane is kept depolarized (at resting membrane potential of −40 mV) and B, when light falls on the retina, the activated rhodopsin reduces intracellular levels of cGMP and consequently, the Na^+ channels are blocked. This results in hyperpolarization (photoreceptor potential).

IMPORTANT NOTE

The photoreceptor potential is different from the receptor potentials in almost all other sensory receptors in that the excitation of photoreceptor causes increased negativity of the membrane potential (hyperpolarization), rather than decreased negativity (depolarization), which is characteristic of all other receptors.

Normally, in dark the electronegativity inside the rod membrane is about 40 mV and after excitation it approaches about 70–80 mV. Further, the eye is unique in that the receptor potential of the photoreceptors is local graded potential, i.e. it does not propagate and does not follow the 'all or none law'.

The sequence of events in photoreceptors by which incident light leads to production of a nerve impulse (*phototransduction*) is summarized in Fig. 11.1-29.

PROCESSING AND TRANSMISSION OF VISUAL IMPULSE IN RETINA

The receptor potential generated in the photoreceptors is transmitted by *electrotonic conduction* (i.e. direct flow of electric current, not action potential to other cells of the retina viz. horizontal cells, amacrine cells and ganglion cells). However, the ganglion cells transmit the visual signals by means of action potential to the neurons of lateral geniculate body and the later to the primary visual cortex.

Role of different cells in the processing of retinal image can be discussed in terms of following concepts which have been evolved in physiology of vision:

- Concept of receptive field,
- Concept of serial processing of the image (see page 909) and
- Concept of parallel processing pathway (see page 908).

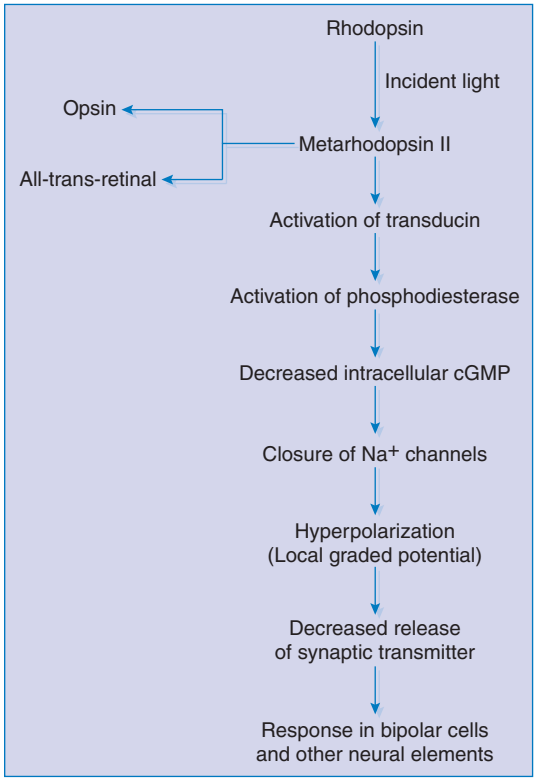


Fig. 11.1-29 Sequence of events involved in phototransduction process in the photoreceptors.

Concept of receptive field

The concept of receptive field has been evolved to explain the processing of visual signal. In general sense, the receptive field is defined as the influence area of a sensory neuron. It is circular in configuration.

Receptive field of individual photoreceptor is small and circular. Light falling in the receptive field hyperpolarizes

the cell (as described above). In the dark, i.e. when the photoreceptor is depolarized a neurotransmitter (glutamate) is released from its terminal. When hyperpolarized, the photoreceptor will therefore release less neurotransmitter.

Horizontal cells have a very large receptive field in comparison to the photoreceptor cell. A horizontal cell transmits signals horizontally in the outer plexiform layer from rods and cones to the bipolar cells. Their main function is to *enhance the visual contrast by causing lateral inhibition, i.e.* they play a role in processing of *spatial information* (Fig. 11.1-30).

Bipolar cells. There are two types of bipolar cells, one type of cells (which are inhibited by glutamate) are depolarized while the other (which are excited by glutamate) are hyperpolarized when the photoreceptors are excited (Figs 11.1-30 and 11.1-31). Thus, the two different types of bipolar cells provide opposing excitatory and inhibitory signals in the visual pathway.

- *Receptive field* of the bipolar cell is also circular in configuration but has got a *centre-surround antagonism*. As shown in Fig. 11.1-31 in case of centre depolarizing cells (also called 'on cell'), the light striking the centre of receptive field activates and the light striking the 'surround' inhibits bipolar cell output. The reverse occurs in the centre hyperpolarizing cell (also called as 'off cell'), i.e. the light striking the 'centre' is inhibitory and the light striking the 'surround' is excitatory to bipolar cell output. The size of the centre of the bipolar cell receptive field is determined by the reach of its dendrites and that of the much larger 'surround' is determined by the spread of interconnected horizontal cells.
- The importance of the above described reciprocal relationship between the depolarizing and hyperpolarizing bipolar cells is that it provides a second mechanism for lateral inhibition (*spatial information processing*) in addition to the horizontal cell mechanism. Further, this reciprocal relationship allows half of the bipolar cells to transmit positive signals and the other half to transmit negative signals, both of these have a useful role in transmitting visual information to the brain.

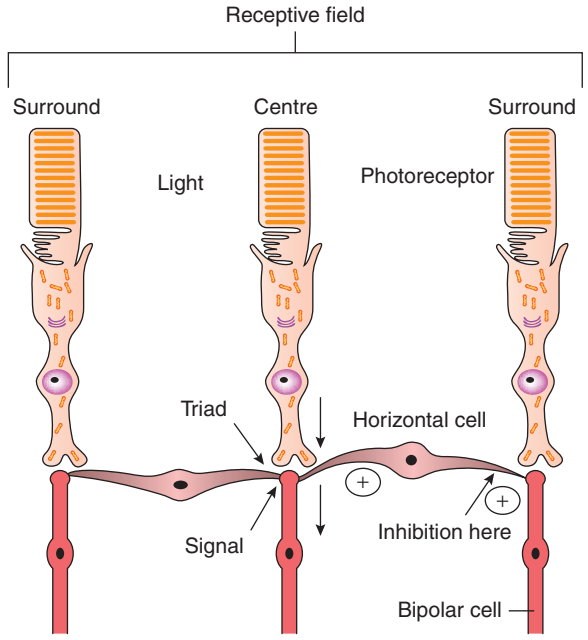


Fig. 11.1-30 Horizontal cells showing phenomenon of lateral inhibition in the surrounding receptive plexiform layer. The central photoreceptor has been stimulated with light and inner portion of the cell membrane becomes more negative. The signal is transmitted upward to bipolar cell and horizontally to horizontal cells. This horizontal transmission results in inhibition of receptor bipolar synapse of neighbouring photoreceptor element. The stimulated bipolar cell may be hyperpolarized or depolarized.

Amacrine cells. Amacrine cells receive information at the synapse of bipolar cell axon with ganglion cell dendrite and use this information for *temporal processing.* Further, these cells receive input from different combinations of on-centre and off-centre bipolar cells. Therefore, the *receptive fields of amacrine cells* are mixture of on-centre and off-centre regions.

The amacrine cells produce transient depolarizing potentials and spikes at the onset and offset of visual stimulus. Therefore these are the first cells in the visual pathway for generating the impulse.

Ganglion cells. The electrical response of bipolar cells (local graded potential) after modification by the amacrine cells is transmitted to the ganglion cells which in turn transmit their signals by means of action potentials to the brain.

Receptive field of ganglion cells like that of bipolar cells has got a centre surround antagonism. Further, like bipolar cells, the ganglion cells are also of two types in terms of their centre response: '*on-centre' cells* that increase their

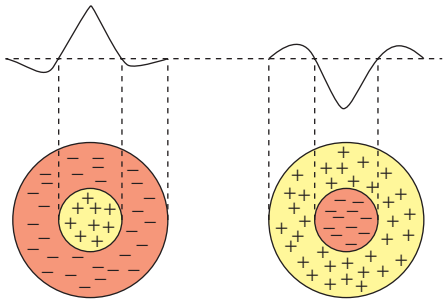


Fig. 11.1-31 Diagram showing the centre surround response to light in 'on' or centre depolarizing bipolar cell (left) and 'off' or centre hyperpolarizing bipolar cell (right). Plus (+) signs indicate region giving depolarizing response and minus (–) signs, a hyperpolarizing one.

discharge and '*off-centre*' cells that decrease their discharge upon illumination of the centre of their receptive fields.

Functionally, the ganglion cells are of two types:

- M ganglion cells (also called large ganglion cells or Y cells) are concerned with movements and stereopsis.
- P ganglion cells (also called small ganglion cells or X cells) are concerned with shape, colour and texture of the image.

Concept of parallel processing pathway

See page 908.

Concept of serial processing of image in retina

See page 909.

Synaptic mediators in the retina

Various types of synaptic transmitters found in retina are: acetylcholine (secreted only by amacrine cells of retina), glutamate, GABA, serotonin, dopamine, glycine, substance P, TRH, GnRH, somatostatin, enkephalin, β endorphin, CCK, VIP, glucagon and neurotensin.

PROCESSING AND TRANSMISSION OF VISUAL IMPULSE IN VISUAL PATHWAY

The retina relays the visual information to the brain (occipital cortex) via visual pathway, which comprises the optic nerve, optic chiasma, optic tract, geniculate body and optic radiations (Fig. 11.1-32).

1. Optic nerve

Optic nerve fibres are axons of the retinal ganglion cells and carry the total output of retina. Arrangement of nerve fibres in the optic nerve head and distal region of the optic nerve (behind the eyeball) is exactly same as in the retina (Fig. 11.1-33), i.e.

Macular fibres, which form papillomacular bundle pass straight in the temporal part of optic disc.

Temporal fibres of retina arch above and below the papillomacular bundle as superior and inferior *arcuate fibres* and occupy upper temporal and lower temporal quadrants of the optic disc.

Nasal fibres of retina come directly to the nasal half of the disc as superior and inferior radiating fibres.

2. Optic chiasma

It is flattened structure lying above the pituitary fossa. Fibres originating from the nasal halves of the retinae decussate at the chiasma while the fibres from temporal halve of retinae remain uncrossed. It is to be noted that the

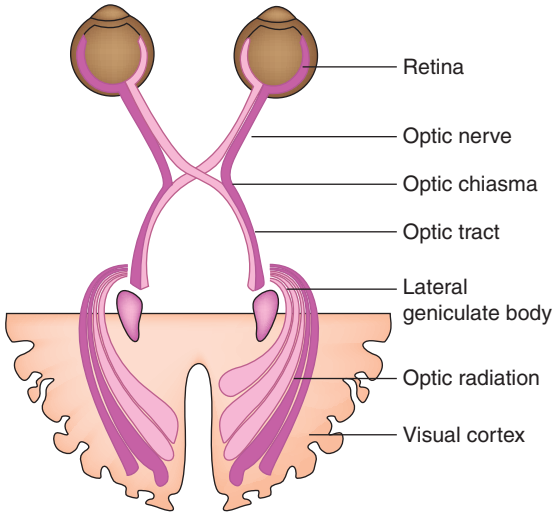


Fig. 11.1-32 Components of visual pathway.

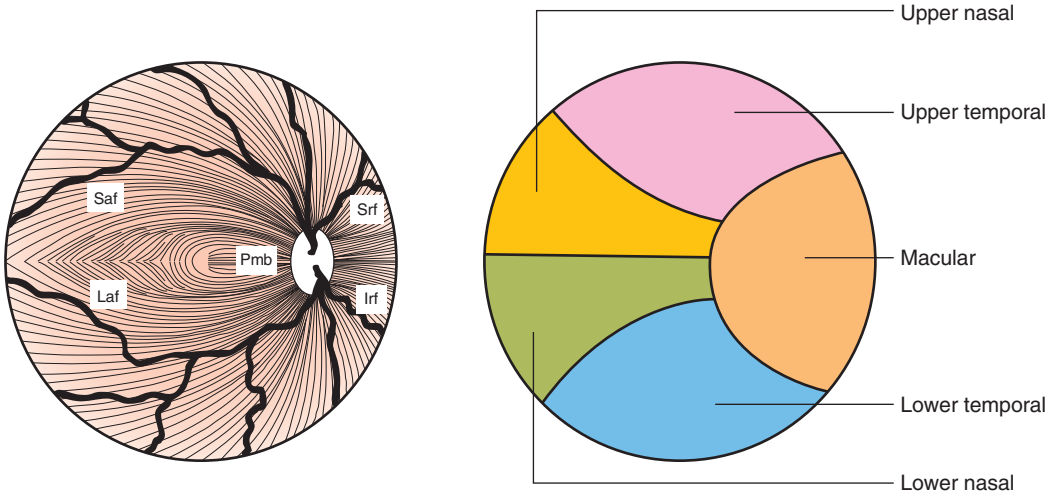


Fig. 11.1-33 Arrangement of nerve fibres: A, in the retina, optic disc and distal part of optic nerve and B, in proximal region of optic nerve.

nasal and temporal halves of retina are demarcated by a vertical line passing through the fovea and not through the optic disc. This implies that visual impulse from the temporal half of visual field goes to the opposite side while the input from the nasal half of the visual field remains in the same side (Fig. 11.1-32).

3. Optic tracts

These are cylindrical bundles of nerve fibres which originate from the posterolateral angle of chiasma and run outwards and backwards to end in the lateral geniculate body (LGB). They consist of temporal fibres of the same side and nasal fibres of the opposite side.

4. Lateral geniculate bodies

These are oval structures situated at the posterior termination of the optic tracts.

Retinotopic projection. The optic tract fibres, which are axons of retinal ganglion cells project a detailed spatial representation of the retina on the lateral geniculate body, with precise point-to-point localization.

Lamellar structure of lateral geniculate body. Each LGB contains six well-defined layers. On each side, layers 1,4 and 6 receive input from the nasal half of the contralateral eye, while layers 2, 3 and 5 receive input from the temporal half of the ipsilateral eye (Fig. 11-1-34). In each layer, there is precise point-to-point representation of the retina.

Magnocellular and parvocellular layers. The layers 1 and 2 of LGB have large cells and are called magnocellular layers, whereas layers 3–6 have small cells and are called parvocellular layers. The inputs to magnocellular layer come from the M ganglion cells of retinae, while inputs to parvocellular layer come from the P ganglion cells of retinae.

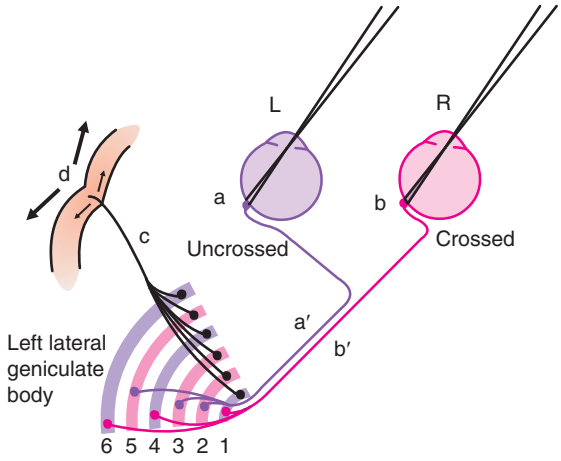


Fig. 11.1-34 Arrangement of termination of axons of ganglion cells (second-order neurons) of two eyes in the lateral geniculate body (LGB) (for explanation see text).

Functions of LGB. Two principal functions served by LGB are:

1. Relay station. Lateral geniculate body serves as a relay station to relay the visual information from the ganglion cells to the visual cortex via parvocellular and magnocellular pathways, which travel through optic radiations to the visual cortex. The relay function is very accurate, so much so that there is exact point-to-point transmission with a high degree of spatial fidelity all the way from the retina to visual cortex. The signals from the two eyes are kept apart in LGB.

2. Visual perception and to 'gate' the transmission of signals, i.e. to control how much of the signals be allowed to pass to the cortex. It is worth noting that only 10–20% of the input to the LGB comes from the retina. The major inputs (80–90%) come as corticofugal fibres from the primary visual cortex and other brain regions. The feedback pathway from the visual cortex has been shown to be involved in the visual processing related to the perception of orientation and motion. These inputs also control the flow of visual information from the retina to the cortex.

5. Optic radiations

The optic radiations are composed of axons of the lateral geniculate relay cells, which project to visual cortex on the same side. The optic radiations maintain a *retinotopic organization* in their passage to visual cortex.

PROCESSING AND ANALYSIS OF VISUAL IMPULSE IN THE VISUAL CORTEX

RETINOTOPIC ORGANIZATION

Just as the ganglion cell axons project a detailed spatial representation of the retina on the LGB, the LGB projects a similar point-to-point representation on the visual cortex. The visual cortex is, therefore, also called the *cortical retina,* since a true copy of the retinal image is formed here. It is only in the visual cortex that impulses originating from corresponding points of two retinae meet. Thus, the right visual cortex is concerned with the perception of objects situated to the left of the vertical median line in the visual fields, and the left visual cortex with the objects situated to the right.

FUNCTIONAL ANATOMY AND ORGANIZATION OF THE VISUAL CORTEX

Visual areas

Classical nomenclature

Classically, the visual cortex has been divided into:

- Primary visual cortex, or striate cortex (area 17),
- Peristriate cortex or visual association area (area 18) and
- Parastriate cortex or visual association area (area 19).

1. Primary visual cortex, also called striate cortex or area 17, lies in the medial surface of the occipital lobe in and near the calcarine sulcus occupying parts of the lingual sulcus. Retina is represented in primary visual cortex as (Fig. 11.1-35):

- *Peripheral part* in the anterior part of area 17, upper quadrants projects on the upper wall of the calcarine sulcus and lower quadrants on the lower walls of sulcus.
- *Macular part* projects mainly to the posterior part of area 17 and anteriorly to a thin strip along the calcarine sulcus. The macular area occupies nearly one-third of area 17.

2. Peristriate cortex or visual association area 18 lies in the walls of lunate sulcus.

3. Parastriate cortex or visual association area 19 lies in the cortex in front of the lunate sulcus.

Modified nomenclature of visual areas

It is now believed that some other parts of the brain are also involved in visual processing and so a modified nomenclature recognizing five visual areas has been described as:

- *V_1 (Visual area 1).* It mainly includes primary visual cortex or Brodmann area 17.
- *V_2 (Visual area 2).* It occupies the greater part of Brodmann area 18, but not the whole of it.
- *V_3 (Visual area 3).* It occupies a narrow strip over the anterior part of area 18.
- *V_4 (Visual area 4).* It occupies the area 19.
- *V_5 (Visual area 5)* or middle temporal (MT) area. It is located at the posterior end of superior temporal gyrus.

Histological layers of primary visual cortex

Primary visual cortex, like other portions of cerebral cortex, has six distinct layers (Fig. 11.1-36). Layers I, II and III are thin and contain pyramidal cells. Layer IV is thickest and contains stellate cells. Layer IV may be further subdivided into layers, a, b, c_α and c_β. Layers V and VI are again relatively thin.

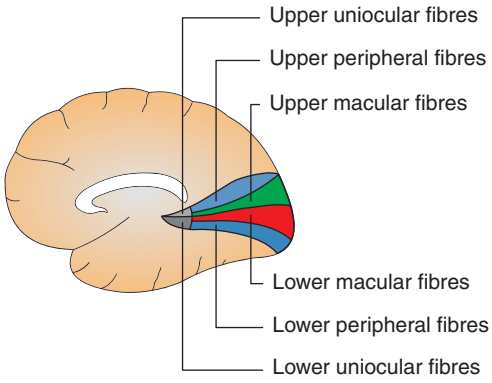


Fig. 11.1-35 Arrangement of termination of retinal fibres in the visual cortex.

PHYSIOLOGICAL CONSIDERATIONS OF VISUAL CORTEX

The present information available on physiology of the visual cortex is just a tip of iceberg. The much credit goes to Hubel and Wiessel for the present day knowledge. Some of the aspects of physiology of visual cortex can be discussed under following headings:

- Concept of receptive field of striate cortex.
- Columnar organization of striate cortex.
- Serial versus parallel analysis of visual image.
- Role of extrastriate cortex in visual functions.
- Psychophysiological aspects of visual functions.

Concept of receptive field of striate cortex

Unlike retinal ganglion cells and lateral geniculate neurons (which respond to both diffuse retinal stimulation and spot stimulus), the cortical neurons prefer stimuli in the form of straight line, bar or edge presented in the proper spatial orientation. Thus, in visual cortex the orientation and configuration of receptive field differ from those earlier points in the visual pathway. Depending upon the peculiarities of receptive fields the cortical cells have been classified into three types:

- Simple cells,
- Complex cells and
- Hypercomplex cells.

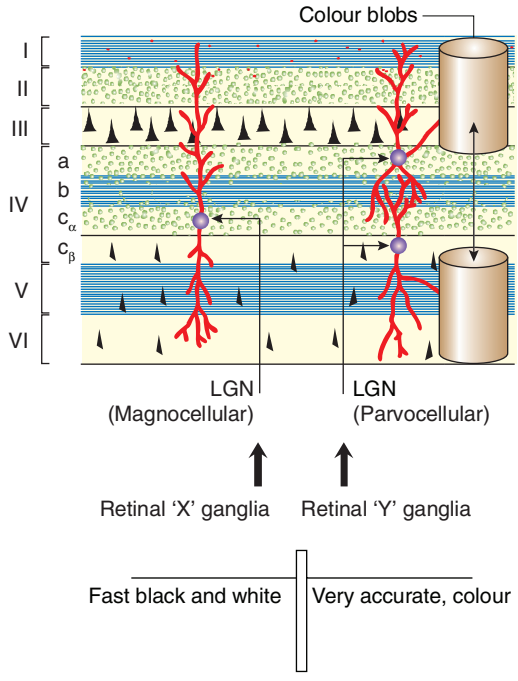


Fig. 11.1-36 Layers of visual cortex.

Simple cells

Simple cells are found mainly in layer IV of the primary visual cortex (area 17) and form the first relay station within the visual cortex. These respond to bars of light, lines or edges, but only when they have a particular orientation. The orientation of a stimulus that is most effective in evoking a response is called the *receptive field axis orientation*. The receptive fields of simple cells are arranged in parallel bands of 'on' and 'off' regions, rather than concentric centre-surround arrangement of geniculate body (Fig. 11.1-37A and B). Receptive fields of simple cells often have a central band that is either an 'on-region or an 'off-region', with parallel flanking region on two sides that are opposite (Fig. 11.1-37C–G).

Thus, the simple cell receptive fields play an important role not only in the detection of lines and borders in the different areas of retinal image, but also detects the orientation of each line or border—that is whether it is vertical, or horizontal or lies at some degree of inclination. It is assumed that for each such orientation of a line a specific neuronal simple cell is stimulated.

Complex cells

These cells are found in the cortical layers above and below layer IV of areas 17, 18, and 19 of visual cortex, and only rarely in layer IV itself. They often respond maximally when a linear stimulus is moved laterally without a change in its orientation. Complex cells often receive input from both eyes and are thus called binocular. The receptive fields of a given binocular complex cell are on corresponding parts of the two retinae and have identical receptive field properties.

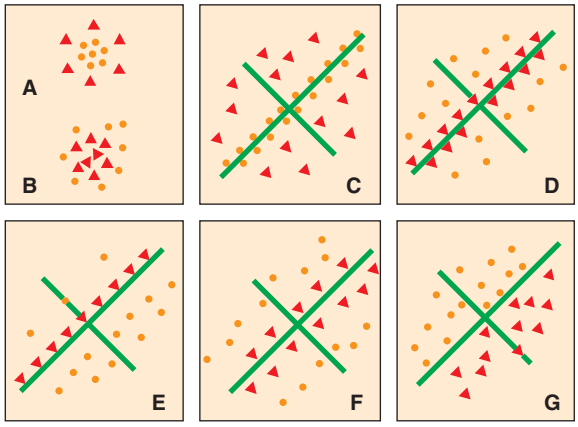


Fig. 11.1-37 Diagram showing arrangement of receptive fields of lateral geniculate body and primary visual cortex: A, 'on centre' geniculate receptive field; B, 'off centre' geniculate receptive field and C to G, arrangement of receptive field of simple cell. Areas give excitatory responses (on response) and areas give inhibitory responses (off response). Receptive field axes are shown by continuous lines through the field centres.

By means of simple and complex cells, the person perceives the features, orientation and movements of the objects. Therefore, simple and complex cells together are known as 'feature detectors'.

Hypercomplex cells

These are found in cortical layers II and III of the areas 17, 18 and 19. These cells retain all the properties of complex cells but also have the added feature of requiring the line stimulus to be of a specific length.

Thus, the hypercomplex cells play a role in the detection of lines of specific length, angles or other shapes.

Columnar organization of the striate cortex (Fig. 11.1-38)

The primary visual cortex is organized into vertically oriented functional modules called the *hypercolumns,* each of which processes visual information from a specific region of the visual field. Each hypercolumn includes sets of three types of vertical columns, which are (Fig. 11.1-38):

- Orientation columns,
- Ocular dominance columns and
- Colour blobs.

Orientation columns

Like the somatic sensory cortex, the primary visual cortex is organized into narrow columns of cells, running from the pial surface to the white matter. Thus, the orientation column is the unit of organization in the visual cortex which can be defined as *vertical grouping of cells with identical orientation specificity.* The visual cortex is thus organized into several million vertical columns of neuronal cells, each being about 30–100 μm wide and 2 mm deep.

Thus, it is possible to speculate that for each ganglion cell receptive field in the visual field, there is collection of column in a small area of visual cortex representing the possible preferred orientation at small intervals throughout the full 360°.

Ocular dominance columns

Ocular dominance columns refer to an independent system of columns which exist in the visual cortex with respect to the binocular input to cortical cells. Although, most cortical neurons are binocularly activated, there remains a strong monocular dominance. Neurons with receptive fields dominated by one eye are grouped alternately into left eye and right eye columns that are 0.25–0.5 mm in width (Fig. 11.1-39).

Ocular dominance column existence may have something to do with the *binocular stereoscopic vision.*

The colour blobs

Interspersed among the primary visual columns are special column-like areas called *colour blobs* (Fig. 11.1-38). These

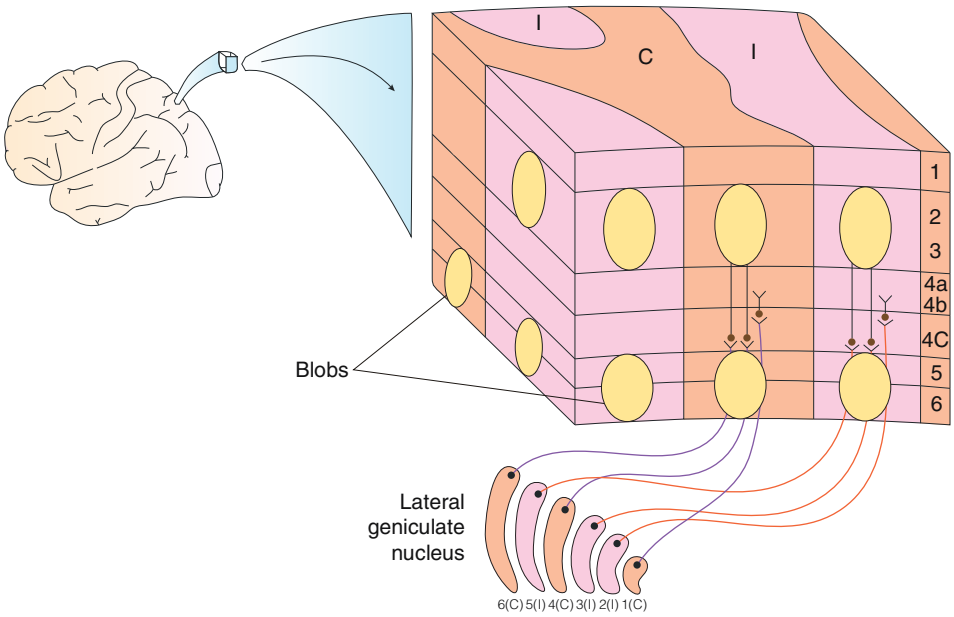


Fig. 11.1-38 Organization of the orientation columns, ocular dominance columns and blobs in the primary visual cortex. (I = Ipsilateral inputs; C = contralateral inputs.)

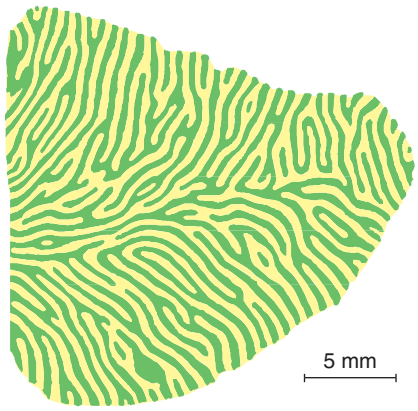


Fig. 11.1-39 Representation of ocular dominance columns in a relatively large segment of monkey striate cortex of right occipital lobe. View is of layer IVc seen from above; ocular dominance column for one eye are in green and those for the other eye in yellow. The foveal representation is to the right.

Table 11.1-2 Differences in the sensitivity of M and P cells to stimulus features

Stimulus feature	Sensitivity	
	M cell	P cell
Colour contrast	No	Yes
Luminance contrast	Higher	Lower
Spatial frequency	Lower	Higher
Temporal frequency	Higher	Lower

receive lateral signals from the adjacent visual column and respond specifically to colour signals. Therefore, it is presumed that these blobs are the primary areas for deciphering colour. Also in certain secondary visual areas additional colour blobs are found, which presumably perform still higher levels of colour deciphering.

CONCEPT OF PARALLEL AND SERIAL PROCESSING OF VISUAL INFORMATION

Parallel processing pathways

The visual pathway is now being considered to be made of two lanes: one made of the large cells is called magnocellular pathway and the other of small cells is called parvocellular pathway. These can be compared to two lanes of a road. The M pathway and P pathway are involved in the *parallel processing* of the image, i.e. analysis of different features of the image. There are striking differences between the sensitivity of M and P cells to stimulus features (Table 11.1-2).

Magnocellular pathway

It is formed by the M cells and their processes (Fig. 11.1-40). M ganglion cells of the retina project to the magnocellular layers of the lateral geniculate nucleus (layer 1 and 2).

The magnocellular projections from the LG nucleus project to the striate cortex first to layer IV c_α and then project directly to the MT. From MT, the M pathway extends to the posterior parietal cortex as dorsal cortical pathway.

Functions. Magnocellular pathway is concerned with the processing and detection of movement, depth and flicker feature of visual information.

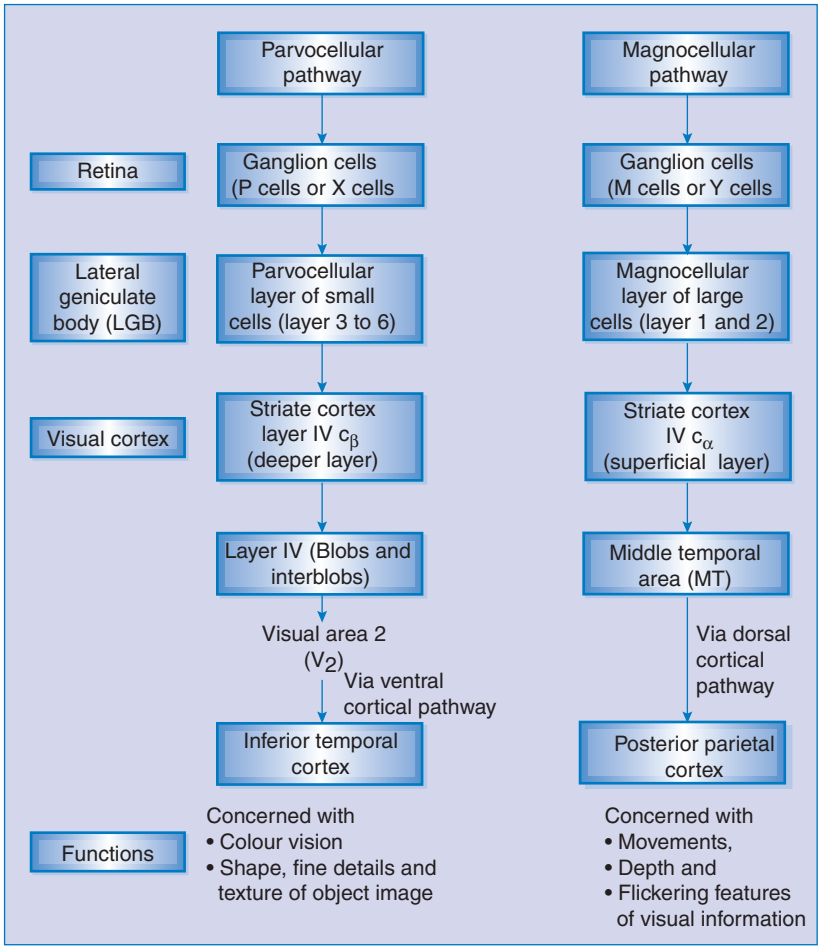


Fig. 11.1-40 The magnocellular (M) and parvocellular pathway (P) from retina project through lateral geniculate body (LGB) to visual area 1 (V_1). Separate pathways to temporal and parietal cortices course through the extrastriate cortex beginning in visual area 2 (V_2).

Parvocellular pathway

It consists of P cells of visual system and their processes (Fig. 11.1-40):

- P ganglion cells of the retina project to the parvocellular layers of lateral geniculate nucleus (layers 3–6).
- Parvocellular projections from the LG nucleus project to the layer IV c_β of striate cortex, from which cells project to the blobs and interblobs of V_1.
- The blobs send a strong projection to the thin stripes in V_2, whereas interblobs send strong projection to the interstripes in V_2.

Function. The parvocellular pathway is concerned with colour vision, texture, shape, and fine details.

Concept of serial processing of visual information

The successive cells in the visual pathway starting from the photoreceptors to the cells of lateral geniculate body are involved in increasingly complex analysis of image. This is called sequential or serial processing of visual information.

Serial processing in the retina

In a sense, the processing of visual information in the retina involves the formation of three images:

First image is formed by the action of light on the photoreceptor. Photoreceptors break up the image into small spots of light or darkness (much like a scanner that breaks down a picture into small pixels).

Second image. First image is converted into second image by bipolar cells. In the formation of second image, the signal is altered by the horizontal cells, which cause *spatial summation* by lateral inhibition.

Third image. The second image is converted into third image by the ganglion cells. In the formation of third image, the signal is altered by amacrine cells, which cause *temporal summation.* Thus, image processing in ganglion cells result in the sharpening of the image contrast. The image is thus analysed mostly in terms of contours of the light-darkness boundaries and areas of uniform light or darkness elicit very little neural response. There is a little change in the

impulse pattern in the LGBs, so the third image reaches the occipital cortex.

Serial analysis of visual image in the visual cortex

A hierarchical model for cell interconnections has been suggested in the visual cortex. The sequence from simple to complex to hypercomplex forms a system of serial analysis with more and more details being deciphered. A complex cell is thought of as receiving input from several simple cells of the same orientation whose receptive fields are overlapping to produce the complex cell receptive field. Since the complex cells are binocular and simple cells are mainly monocular, this adds support to the idea that complex cells are at a more advanced stage of processing.

VISUAL PERCEPTION

It is a complex integration of light sense, form sense, sense of contrast and colour sense. The receptive field organization of the retina and cortex are used to encode this information about a visual image.

THE LIGHT SENSE

It is awareness of the light. The range of luminance to which human eye responds is summarized in Fig. 11.1-41. The minimum brightness required to evoke a sensation of light is called the *light minimum.* It should be measured when the eye is dark adapted for at least 20–30 minutes.

The human eye in its ordinary use throughout the day is capable of functioning normally over an exceedingly wide range of illumination by a highly complex phenomenon termed as the *visual adaptation.* The process of visual adaptation primarily involves:

- *Dark adaptation* (adjustment in dim illumination) and
- *Light adaptation* (adjustment to bright illumination).

Dark adaptation

It is the ability of the eye to adapt itself to a decreasing illumination. When one goes from a bright sunshine into a dimly lit room, one cannot perceive the objects in the room until some time has elapsed. During this period, eye is adapting to low illumination. The time taken to see in dim illumination is called *dark adaptation time.* The rods are much more sensitive to low illumination than cones. Therefore, rods are used more in dim light (*scotopic vision*) and cones in bright light (*photopic vision*).

Dark adaptation curve

Dark adaptation curve plotted with an illumination of test object in vertical axis and duration of dark adaptation along the horizontal axis shows that visual threshold falls progressively in the darkened room for about half an hour until

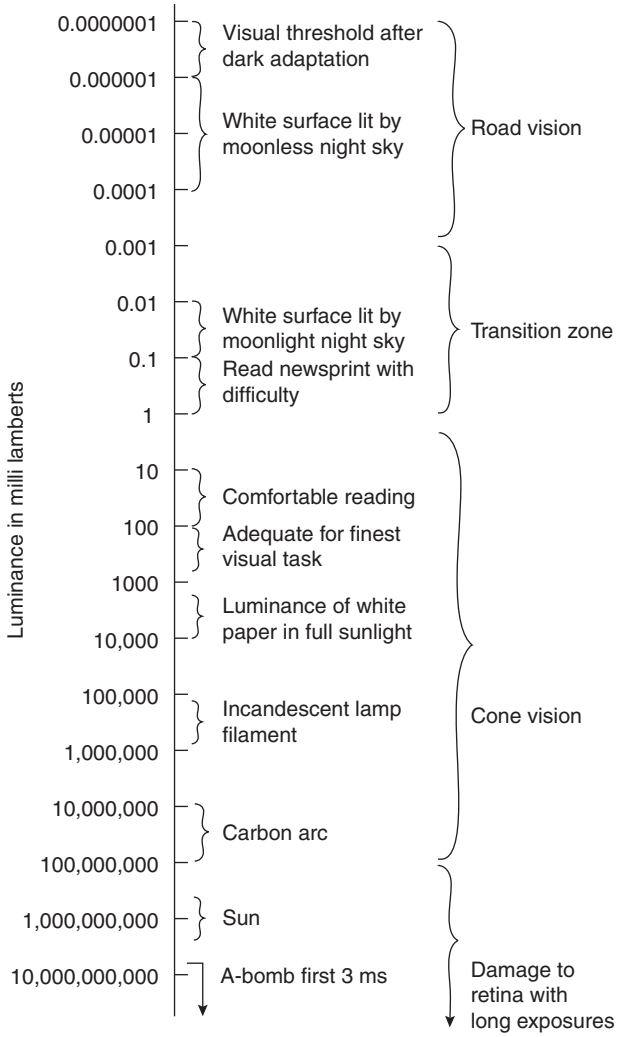


Fig. 11.1-41 The ranges of luminance to which human eye responds.

a relative constant value is reached (Fig. 11.1-42). The dark adaptation curve plotted with retinal sensitivity along the vertical axis and duration of dark adaptation along the horizontal axis (Fig. 11-1-43) shows that sensitivity of retina is very low on first entering the darkness, but within 1 min the sensitivity has increased ten-fold, that is, the retina can respond to light of one-tenth the previously required intensity. At the end of 20 min, the sensitivity has increased about 6000-fold, and at the end of 40 min, it has increased about 25,000-fold.

It can be seen that the decrease in threshold of the retina (Fig. 11.1-42), i.e. an increase in sensitivity of retina (Fig. 11.1-43), proceeds in two steps: (i) the first is rapid, of short duration, and small in extent and (ii) the second is slow, more prolonged and larger. This indicates that two processes are at work, each having different characteristics and that the break in the curve is the point at which one process is about to finish and the second one is just commencing. The analyses have revealed that the first plateau of the curve represents cone threshold (reached in about 5 min)

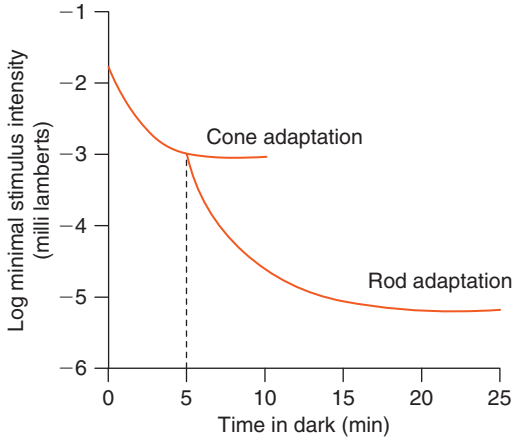


Fig. 11.1-42 Dark adaptation curve plotted with illumination of test object along the vertical axis and duration of dark adaptation along the horizontal axis.

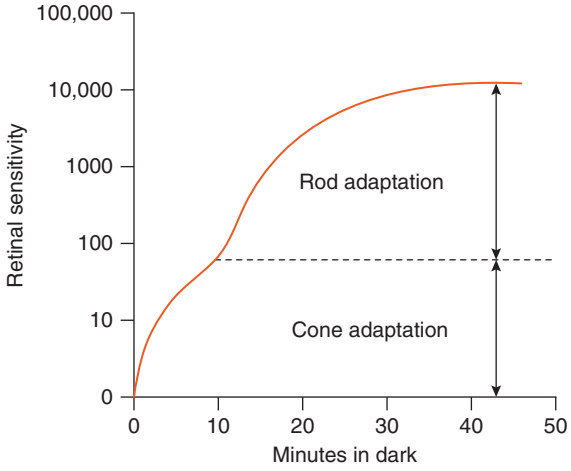


Fig. 11.1-43 Dark adaptation curve plotted with retinal sensitivity along the vertical axis and duration of dark adaptation along the horizontal axis.

and the second plateau represents rod threshold (reached after about 30 min). The inflection of the dark adaptation curve where the rod limb begins is called the cone-rod break or alpha point, and it usually occurs after 7–10 min of adaptation. The final rod phase of adaptation does not begin until 93% of rhodopsin has already regenerated.

Dark adaptation and vitamin A deficiency

Severe deficiency of vitamin A elevates the threshold for dark adaptation curve due to depletion of the photosensitive pigment. *Nyctalopia,* i.e. night blindness is another important feature of vitamin A deficiency. Other causes of nyctalopia are retinitis pigmentosa and congenital night blindness.

Light adaptation

When one passes suddenly from a dim to a brightly lighted environment, the light seems intensely and even uncomfortably bright until the eyes adapt to the increased illumination and the visual threshold rises. The process by means of which retina adapts itself to the bright light is called *light adaptation.* Unlike dark adaptation, the process of light adaptation is very quick and occurs over a period of 5 min. Strictly speaking, light adaptation is merely the disappearance of dark adaptation.

THE FORM SENSE

It is the ability to discriminate between the shapes of the objects. Cones play a major role in this faculty. Therefore, form sense is most acute at the fovea, where there is maximum number of cones and decreases very rapidly towards the periphery. Visual acuity recorded by Snellen's test chart is a measure of the form sense.

Components of visual acuity

In clinical practice, measurement of the threshold of discrimination of two spatially separated targets (a function of the fovea centralis) is termed visual acuity.

Resolution (ordinary visual acuity)

Discrimination of two spatially separated targets is termed resolution. The minimum separation between the two points, which can be discriminated as two, is known as *minimum resolvable (minimum separable).* The distance between the two targets is specified by the angle subtended at the nodal point of the eye. The normal angular threshold of discrimination for resolution measures approximately 30–60 s arc; it is usually called the minimum angle of resolution. The distance on retina separating two images is approximately 4.5 μm.

The clinical tests determining visual acuity measure the form sense or reading ability of the eye. Thus, broadly, resolution refers to the ability to identify the spatial characteristics of a test figure. The test targets in these tests may either consist of letters (Snellen's chart) or broken circle (Landolt ring). More complex targets include gratings and checker board patterns.

Snellen's test types

To measure the minimal resolvable, Snellen constructed certain test types, which are now routinely used to test the distant central visual acuity.

The fact that two distant points can be visible as separate only when they subtend an angle of 1 min at the nodal point of the eye, forms the basis of Snellen's test types. It consists of a series of black capital letters on a white board, arranged in lines, each progressively diminishing in size. The lines comprising the letter have such a breadth that they will subtend an angle of 1 min at the nodal point. Each letter of the chart is so designed that it fits in a square, the

sides of which are five times the breadth of the constituent lines. Thus at the given distance, each letter subtends an angle of 5 min at the nodal point of the eye (Fig. 11.1-44). The letter of the top line of Snellen's chart (Fig. 11.1-45) should be read clearly at a distance of 60 m. Similarly, the letters in the subsequent lines should be read from a distance of 36, 24, 18, 12, 9, 6, 5 and 4 m.

For testing distant visual acuity, the patient is seated at a distance of 6 m from the *Snellen's chart*, so that the rays of light are practically parallel and the patient exerts minimal accommodation. The chart should be properly illuminated (not less than 20 ft candles). The patient is asked to read the chart with each eye separately and the visual acuity is recorded as a fraction, the numerator being the distance of the patient from the letters, and the denominator being the smallest letters accurately read.

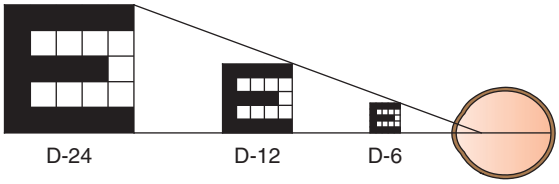


Fig. 11.1-44 Principle of Snellen's test type.

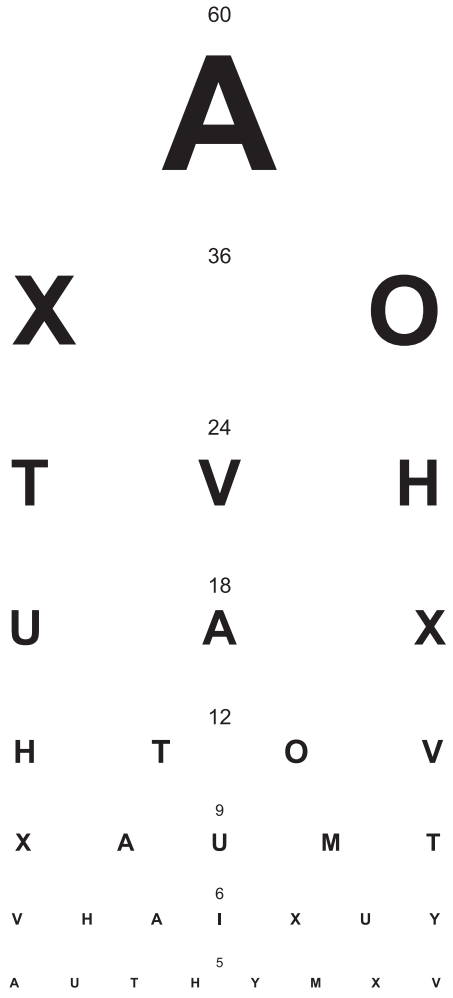


Fig. 11.1-45 Snellen's test types.

When the patient is able to read up to 6 m line, the visual acuity is recorded 6/6, which is normal. Similarly, depending upon the smallest line which the patient can read from the distance of 6 m, his vision is recorded as 6/9, 6/12, 6/18, 6/24, 6/36 and 6/60, respectively.

Visual acuity is influenced by variety of factors, which include:

(i) *Optical factors,* such as state of image forming mechanism. Visual acuity is low in patient having refractive errors and optic aberration.

(ii) *Retinal factors,* such as state of cones. Visual acuity is maximum at fovea centralis, whereas at the periphery of retina it is less than 1/30th as that of fovea centralis.

(iii) Stimulus factors, such as:
- Distance of the object
- Size of the object
- Illumination and brightness
- Contrast between stimulus and background
- Duration for which the subject is exposed to the stimulus

Visual acuity for near

Near vision is tested by asking the patient to read the near vision chart (Fig. 11.1-46) kept at a distance of 25 cm in good illumination, with each eye separately. In near vision charts, a series of different sizes of printer type are arranged in an increasing order and marked accordingly. Commonly used near vision chart is *Jaeger's chart.*

Critical flicker fusion frequency

When intermittent light stimuli are presented to the eye, a sensation of 'flicker' is evoked. As the frequency of presentation of the stimuli is increased, a point is reached at which flicker sensation fuses to form the sensation of continuous stimulation. This frequency is known as the critical flicker fusion (CFF) frequency. The CFF frequency serves as a measure of the temporal resolving power of the visual system under the particular condition of stimulation. Motion pictures move because the frames are presented at a rate above the CFF and movies began to flicker when the projector slows down.

THE COLOUR SENSE

Colour sense is the ability of the eye to discriminate between colours excited by light of different wavelengths. Some broad facts about colour vision are:

- Colour vision is a function of cones and thus better appreciated in photopic vision.

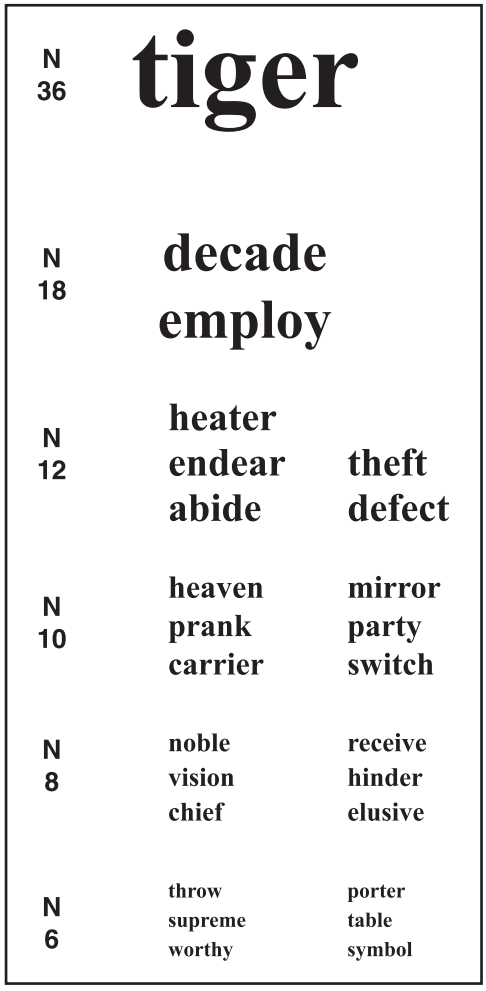


Fig. 11.1-46 Near-vision chart.

- There are three different types of cones viz. red sensitive, green sensitive and blue sensitive, which combinedly perform the function of colour vision.
- All colours are a result of admixture in different proportion of three primary colours: the red (723–647 nm), green (575–492 nm) and blue (492–450 nm).
- Colours have three attributes: hue, intensity and saturation.
- For any colour there is a complementary colour that, when properly mixed with it, produces a sensation of white.
- The colour perceived depends in part on the colour of other objects in the visual field. Thus, for example, a red object is seen red if the field is illuminated with green or blue light but as pale pink or white if the field is illuminated with red light.
- A normal person can see all wavelengths between violet and red. If the wavelength is shorter than that of violet, the light becomes ultraviolet (UV) and is beyond visibility. If the wavelength is greater than 750 nm, the light is infrared and is again beyond visibility.
- In dim light, all the colours are seen as grey; this is called Purkinje shift phenomenon.

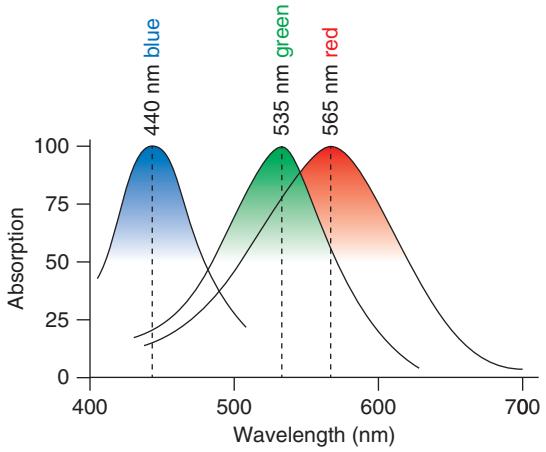


Fig. 11.1-47 Absorption spectrum of three cone pigments.

Mechanism (neurophysiology) of colour vision

Theories of colour vision

The process of colour analysis begins in the retina and is not entirely a function of brain. Many theories have been put forward to explain the colour perception, but two have been particularly influential:

1. Trichromatic theory. The trichromacy of colour vision was originally suggested by Young and subsequently modified by Helmholtz. Hence it is called the Young-Helmholtz theory. It postulates the existence of three kinds of cones, each containing a different photopigment and maximally sensitive to one of three primary colours viz. red, green and blue. The sensation of any given colour is determined by the relative frequency of the impulse from each of the three cone systems. In other words, a given colour consists of admixture of the three primary colours in different proportions. The correctness of the Young-Helmholtz's trichromacy theory of colour vision has now been demonstrated by the identification and chemical characterization of each of the three pigments by recombinant DNA technique, each having different absorption spectrum as (Fig. 11.1-47):

- *Red sensitive cone pigment,* also known as *erythrolabe* or long wavelength sensitive (LWS) cone pigment, absorbs maximally in a yellow portion with a peak at 565 nm.
- *Green sensitive cone pigment*, also known as *chlorolabe* or medium wavelength sensitive (MWS) cone pigment, absorbs maximally in the green portion with a peak at 535 nm.
- *Blue sensitive cone pigment*, also known as *cyanolabe* or short wavelength sensitive (SWS) cone pigment, absorbs maximally in the blue-violet portion of the spectrum with a peak at 440 nm.

It has been studied that the gene for human rhodopsin is located on chromosome 3 and the gene for the blue-sensitive cone is located on chromosome 7. The genes for the red and green sensitive cones are arranged in tandem array on the q arm of the X chromosomes.

2. Opponent colour theory. According to the opponent colour theory of Herring, there are two main types of colour opponent ganglion cells:

(a) Red-green opponent colour cells use signals from red and green cones to detect red/green contrast within their receptive field.

(b) Blue-yellow opponent colour cells obtain a yellow signal from the summed output of red and green cones, which is contrasted with the output from blue cones within the receptive field.

Analysis of colour signals in the visual cortex

Colour information from the parvocellular portion of the LGB is relayed to the layer IVc of the *striate cortex* (area 17). From there, the information passes to the blobs in layers II and III. The neurons in the *blobs* respond to colours. Like the ganglion cells and LGB cells, they are *centre-surround cells.* Many are *double-opponent cells,* which for example, are stimulated by green centre and inhibited by green surround and are inhibited by red centre and stimulated by red surround. From the blobs colour information is relayed to thin strips in the visual association area and from there to a *specialized area concerned* with colour, which in human is in the lingual and fusiform gyri of occipital lobe.

Colour blindness

An individual with normal colour vision is known as *'trichromate'*. In colour blindness, faculty to appreciate one or more primary colours is either defective (anomalous) or absent (anopia). It may be congenital or acquired.

1. Congenital colour blindness

It is an inherited condition affecting males more (3–4%) than females (0.4%). Colour blindness like haemophilia is also inherited as recessive sex linked X chromosome abnormality (see page 162). The abnormal gene responsible for colour blindness is located on X chromosome. The females are the carrier, they show defect only when both the X chromosome show abnormal gene. However, the female children of a man with X linked colour blindness pass the defect on to half of their sons. Therefore, X-linked colour blindness skips generation and appears in males of every second generation. It may be of the following types:

- Dyschromatopsia and
- Achromatopsia

(i) Dyschromatopsia. Dyschromatopsia, literally means colour confusion due to deficiency of mechanism to perceive colours. It can be classified into:

- Anomalous trichromatism and
- Dichromatism.

(a) Anomalous trichromatism. Here the mechanism to appreciate all the three primary colours is present but is defective for one or two of them. It may be of following types:

- *Protanomalous,* i.e. defective red colour appreciation,
- *Deuteranomalous,* i.e. defective green colour appreciation.
- *Tritanomalous,* i.e. implies defective blue colour appreciation.

(b) Dichromatism. In this condition, faculty to perceive one of the three primary colours is completely absent. Such individuals are called dichromates and may have one of the following types of defects:

- *Protanopia,* i.e. complete red colour defect.
- *Deuteranopia,* i.e. complete defect for green colour.
- *Tritanopia,* i.e. absence of blue colour appreciation.

Red-green deficiency (protanomalous, protanopia, deuteranomalous, and deuteranopia) is more common. Such a defect is source of danger in certain occupations, such as drivers, sailors and traffic police. *Blue deficiency* (tritanomalous and tritanopia) is comparatively rare.

(ii) Achromatopsia. It is an extremely rare condition presenting as cone monochromatism or rod monochromatism.

Cone monochromatism is characterized by the presence of only one primary colour and thus the persons are truly colour blind.

Rod monochromatism may be complete or incomplete. It is inherited as an autosomal recessive trait. It is characterized by:

- Total colour blindness and
- Day blindness (visual acuity is about 6/60).

Acquired colour blindness

It may follow damage to macula or optic nerve.

Tests for colour vision

Commonly employed colour vision tests are:

1. Pseudo-isochromatic chart test. It is the most commonly employed test using Ishihara's plates. In this, there are patterns of coloured and grey dots which reveal one pattern to the normal individuals and another to the colour deficients.

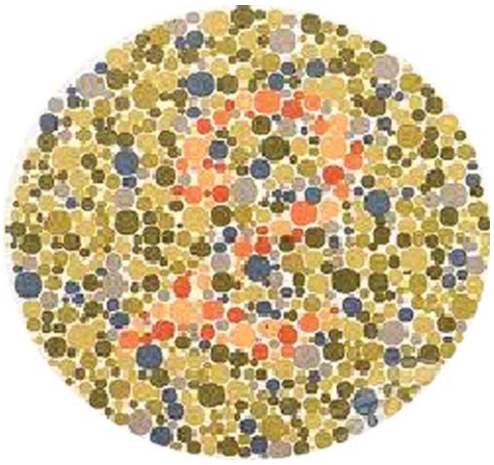


Fig. 11.1-48 Ishihara charts.

It is quick method of screening colour blinds from the normal individuals (Fig. 11.1-48).

2. The lantern test. In this test, the subject has to name the various colours shown to him by a lantern and the judgement is made by the mistake he makes. *Eldridge-Green lantern* is most popular.

3. Holmgren wool test. In this, the subject is asked to make a series of colour matches from a selection of skeins of coloured wools.

CONTRAST SENSITIVITY

Contrast sensitivity is the ability to perceive slight changes in luminance between regions, which are not separated by definite borders and is just as important as the ability to perceive sharp outlines of relatively small objects. It is only the latter ability which is tested by means of the Snellen's test types. In many diseases, loss of contrast sensitivity is more important and disturbing to the patient than the loss of visual acuity.

Encoding of contrast

Contrast is encoded when one ganglion cell is stimulated and its neighbour is inhibited.

ELECTROPHYSIOLOGICAL TESTS

The electrophysiological tests allow objective evaluation of the retinal functions. These include: electroretinography (ERG), electro-oculography (EOG) and visually evoked response (VER).

Electroretinography (ERG)

Electroretinography (ERG) is the record of changes in the resting potential of the eye induced by a flash of light. It is measured in a dark adapted eye with the active electrode (fitted on contact lens) placed on the cornea and the reference electrode attached on the forehead.

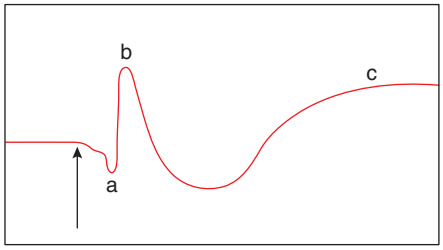


Fig. 11.1-49 Components of normal electroretinogram (ERG).

Normal record of ERG consists of the following waves (Fig. 11.1-49):

- *a-wave.* It is a negative wave possibly arising from the rods and cones.
- *b-wave.* It is a large positive wave which is generated by Muller cells, but represents the activity of the bipolar cells.
- *c-wave.* It is also a positive wave representing the metabolic activity of pigment epithelium.

Both scotopic and photopic responses can be elicited in ERG. Foveal ERG can provide information about the macula.

Uses. ERG is very useful in detecting functional abnormalities of the outer retina (up to bipolar cell layer), much before the ophthalmoscopic signs appear. However, ERG is normal in diseases involving ganglion cells and the higher visual pathway, such as optic atrophy.

Electro-oculography (EOG)

Electro-oculography is based on the measurement of resting potential of the eye, which exists between the cornea (+ve) and back of the eye (–ve).

Normally, the resting potential of the eye decreases during dark adaptation and reaches its peak in light adaptation.

Uses. Since the EOG reflects the pre-synaptic function of the retina, any disease that interferes with the functional interplay between the RPE and the photoreceptors will produce an abnormal or absent light rise in the EOG.

Visually evoked response (VER)

As we know when light falls on the retina, a series of nerve impulses are generated and passed on to the visual cortex via the visual pathway. The VER is nothing but the EEG

recorded at the occipital lobe. *Visually evoked response* is the *only clinically objective technique available to assess the functional state of the visual system beyond the retinal ganglion cells.* Since there is disproportionately large projection of the macular area in the occipital cortex, the VER represents the macula-dominated response.

FIELD OF VISION AND BINOCULAR VISION

FIELD OF VISION

The visual field is a three-dimensional area that can be seen around an object of fixation. The visual field of the eye is not circular because it is cutoff medially by nose, superiorly by roof of the orbit and inferiorly by the cheek bone. The extent of normal visual field with a 5 mm white colour object is superiorly 60°, inferiorly 70°, nasally 60° and temporally 90° (Fig. 11.1-50). The field for blue and yellow is roughly 10° less and that for red and green colour is about 20° less than that for white. The visual field can be divided into central and peripheral fields.

- Central field includes an area from the fixation point to circle 30° away. The central zone contains physiological blind spot on the temporal side.
- Peripheral field of vision refers to the rest of the area beyond 30° to the outer extent of the field of the vision.

Methods of estimating the visual fields

A. Peripheral field charting
- Confrontation method
- Perimetry: Lister's, Goldmann's and automated.

B. Central field charting
- Campimetry or scotometry

Confrontation method. This is a rough but rapid and extremely simple method of estimating the peripheral visual field. Assuming the examiner's fields to be within the normal range, they are compared with patient's visual fields.

Perimetry. It is the procedure for estimating extent of the visual fields.

Lister's perimeter. It has a metallic semicircular arc, graded in degrees, with a white dot for fixation in the centre. The arc can be rotated in different meridia. With the help of this perimeter, extent of peripheral field is charted (Fig. 11.1-51).

Campimetry (scotometry) is done to evaluate the central and paracentral area (30°) of the visual field. The Bjerrum's screen is used and can be of size 1 or 2 m^2 (Fig. 11.1-52).

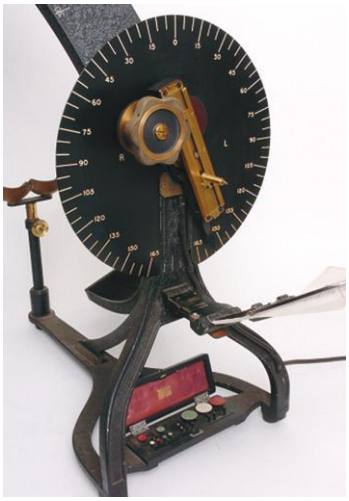


Fig. 11.1-51 Perimeter.

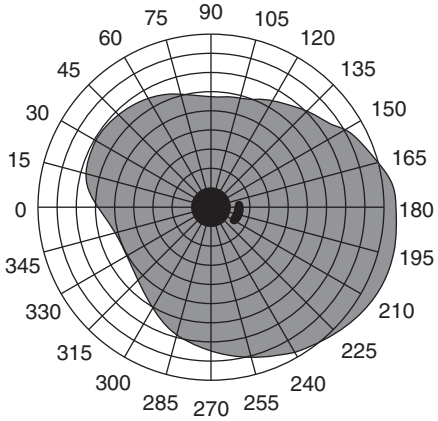


Fig. 11.1-50 Extent of normal visual field.

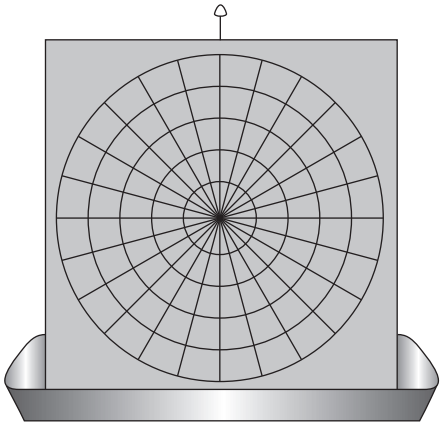


Fig. 11.1-52 Bjerrum's screen.

Initially, the *blind spot* physiological scotoma is charted, which is normally located about 15° temporal to the fixation point. Dimensions of blind spots are horizontally 7–8° and vertically 10–11°. *Central/paracentral scotomas* can be found in optic neuritis and open angle glaucoma.

Automated perimeters are computer assisted and test visual fields by a static method. They automatically test suprathreshold and threshold stimuli and quantify depth of field defect.

Common causes of defects in the field of vision

The most common causes of field defects are:

I. Glaucomatous field defects

A typical pattern of field defects is seen in patients with chronic glaucomas.

II. Fields defects in lesions of visual pathway

Salient features and important causes of lesions of the visual pathway at different levels (Fig. 11.1-53) are as follows:

1. Lesions of the optic nerve. These are characterized by a marked loss of vision or complete blindness on the affected side associated with abolition of the direct light reflex on the ipsilateral side and consensual light reflex on the contralateral side. Near (accommodation) reflex is present.

2. Lesions through proximal part of the optic nerve. Salient features of such lesions are: ipsilateral blindness, contralateral hemianopia and abolition of direct light reflex on the affected side and consensual on the contralateral side. Near reflex is intact.

3. Sagittal (central) lesions of the chiasma. These are characterized by bitemporal hemianopia and bitemporal

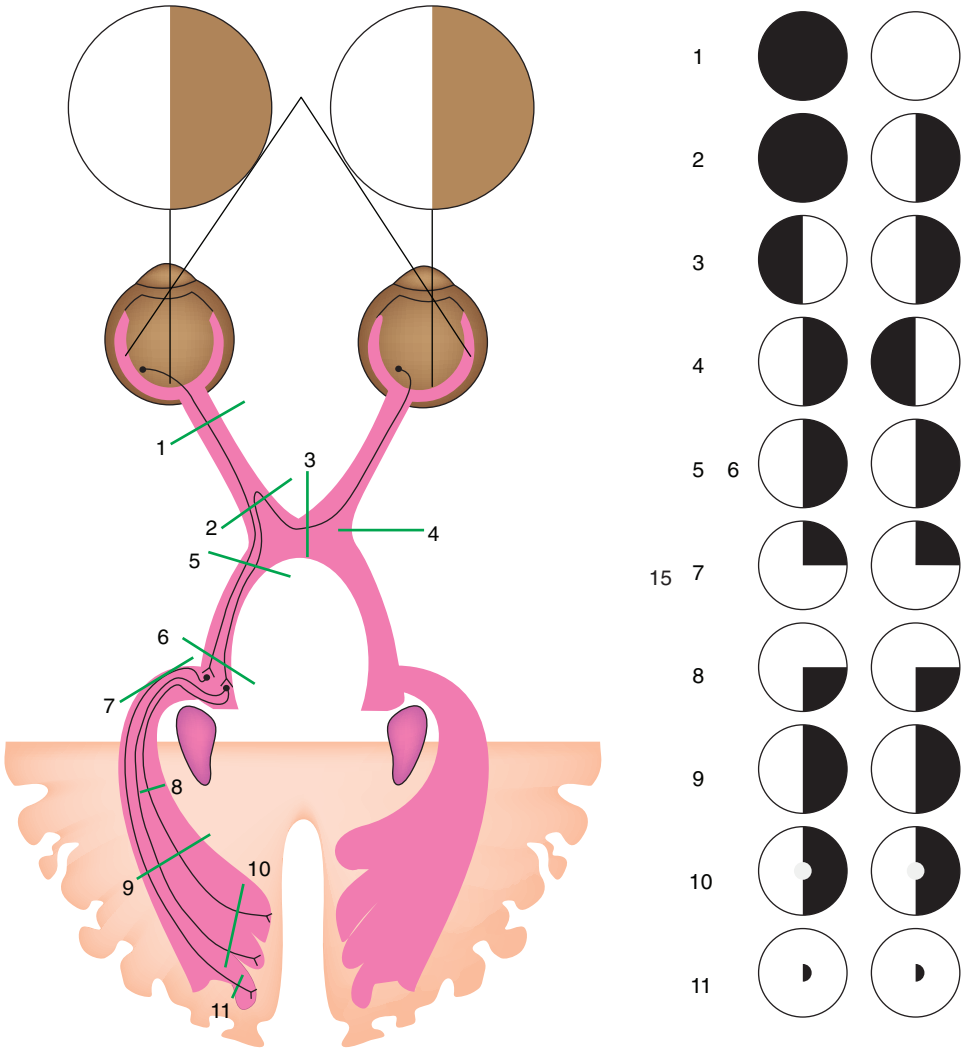


Fig. 11.1-53 Lesions of the visual pathways at the level of: 1, optic nerve; 2, proximal part of optic nerve; 3, central chiasma; 4, lateral chiasma (both sides); 5, optic tract; 6, geniculate body; 7, part of optic radiations in temporal lobe; 8, part of optic radiations in the parietal lobe; 9, optic radiations; 10, visual cortex sparing the macula and 11, visual cortex, only macula.

hemianopic paralysis of pupillary reflexes. *Common causes* of central chiasmal lesion are: suprasellar aneurysms and tumours of pituitary gland.

4. Lateral chiasmal lesions. Salient features of such lesions are binasal hemianopia associated with binasal hemianopic paralysis of the pupillary reflexes. *Common causes* of such lesions are distension of third ventricle causing pressure on each side of the chiasma.

5. Lesions of optic tract. These are characterized by homonymous hemianopia associated with contralateral hemianopic pupillary reaction (Wernicke's reaction).

6. Lesions of lateral geniculate body. These produce homonymous hemianopia with sparing of pupillary reflexes.

7. Lesions of optic radiations. Their features vary depending upon the site of lesion. Involvement of total optic radiations produces complete homonymous hemianopia (sometimes sparing the macula). Inferior quadrantic hemianopia (*pie on the floor*) occurs in lesions of parietal lobe (containing superior fibres of optic radiations). Superior quadrantic hemianopia (*pie in the sky*) may occur following lesions of the temporal lobe (containing inferior fibres of optic radiations).

Note. Pupillary reactions are normal as the fibres of the light reflex leave the optic tracts to synapse in the superior colliculi. *Common lesions* of the optic radiations include vascular occlusions.

8. Lesions of the visual cortex. Congruous homonymous hemianopia (usually sparing the macula) is a feature of occlusion of posterior cerebral artery supplying the anterior part of occipital cortex. Congruous homonymous macular defect occurs in lesions of the tip of the occipital cortex following head injury or gun shot injuries. Pupillary light reflexes are normal and optic atrophy does not occur following visual cortex lesions.

9. Lesions of visual area 18 and 19. The visual sensibility remains intact but there is disturbance in higher visual functions (visual agnosia).

BINOCULAR SINGLE VISION

When a normal individual fixes his visual attention on an object of regard, the image is formed on the fovea of both the eyes separately; but the individual perceives a single image. This state is called binocular single vision. The central part of visual fields of both eyes coincides (Fig. 11.1-54). The impulses setup in two retinae by light rays from an object fused at the cortical level to form a single image (fusion). In a state of normal binocular single vision, there exists a precise physiological relationship between the corresponding points of two retinae. Thus, the foveae of two eyes act as corresponding points and have the same visual direction. This adjustment is called *normal retinal correspondence*. It is a conditioned reflex which is not present since birth but is acquired during first 6 months and is completed during first few years.

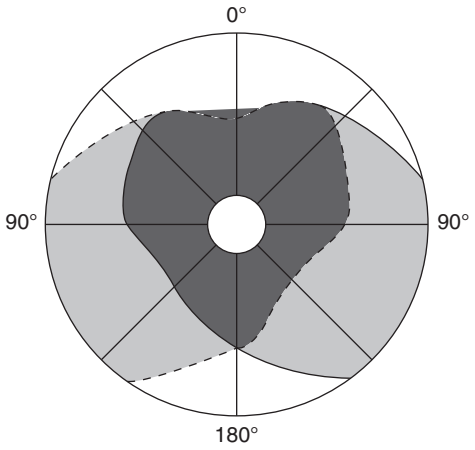


Fig. 11.1-54 Monocular and binocular visual fields. The dashed outline depicts visual field of left eye and solid line, that of right eye. The common area (clear zone) is viewed by binocular vision and coloured areas are viewed by monocular vision.

IMPORTANT NOTE

Binocular single vision has an important role in perception of depth.

Anomalies of binocular vision

Anomalies of binocular vision include suppression, amblyopia and diplopia.

1. Suppression. It is a temporary active cortical inhibition of the image of an object formed on the retina of the squinting eye. This phenomenon occurs only during binocular vision (with both eyes open). However, when the fixating eye is covered, the squinting eye fixes (i.e. suppression disappears).

2. Amblyopia. It is an impairment of vision in the absence of any organic disease of ocular media and visual pathway.

3. Diplopia occurs due to the formation of image on dissimilar points of the two retinae. It is the main symptom of paralysis of extraocular muscles.

PHYSIOLOGY OF OCULAR MOTILITY

EXTRAOCULAR MUSCLES

A set of six extraocular muscles (four recti and two obliques) control the movements of each eye (Fig. 11.1-55). Rectus muscles are superior (SR), inferior (IR), medial (MR) and lateral (LR). The oblique muscles include superior (SO) and inferior (IO).

Nerve supply

The extraocular muscles are supplied by third, fourth and sixth cranial nerves. The third cranial nerve (oculomotor) supplies the superior, medial and inferior recti and inferior oblique muscles. The fourth cranial nerve (trochlear) supplies the superior oblique and the sixth nerve (abducent) supplies the lateral rectus muscle.

Actions

The extraocular muscles rotate the eyeball around vertical, horizontal and anteroposterior axes. Medial and lateral rectus muscles are almost parallel to the optic axis of the eyeball; so they have got only the main action. While superior and inferior rectus muscles make an angle of 23° and reflected tendons of the superior and inferior oblique muscles of 51° with the optical axis in the primary position; so they have subsidiary actions in addition to the main action. Actions of each muscle (Fig. 11.1-56) are shown in Table 11.1-3.

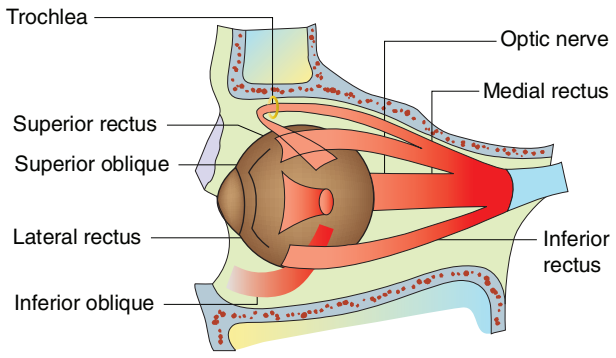


Fig. 11.1-55 Extraocular muscles.

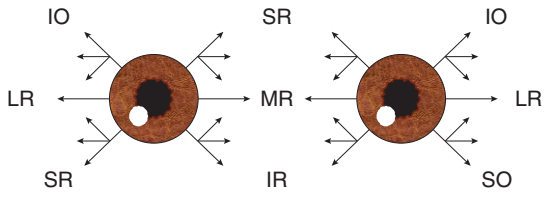


Fig. 11.1-56 Action of extraocular muscles. (SR = superior oblique; LR = lateral rectus; IO = inferior oblique; MR = medial rectus; IR = inferior rectus and SO = superior oblique.)

SUPRANUCLEAR CONTROL OF EYE MOVEMENTS

There exists a highly accurate, still not fully elucidated, supranuclear control of eye movements, which keeps the two eyes yoked together so that the image of the object of interest is simultaneously held on both fovea despite movement of the perceived object or the observer's head and/or body.

Following supranuclear eye movement systems have been recognized:

1. Saccadic system,
2. Smooth pursuit system,
3. Vergence system,
4. Vestibular system,
5. Optokinetic system and
6. Position maintenance system.

All these systems perform specific functions and each one is controlled by a different neural system but share the same final common path—the motor neurons that supply the extraocular muscles.

1. Saccadic system

Saccades are sudden, jerky conjugate eye movements that occur as the gaze shifts from one object to another. Thus they are performed to bring the image of an object quickly on the fovea. Though normally voluntary, saccades may be involuntary aroused by peripheral, visual or auditory stimuli.

2. Smooth pursuit eye movement system

Smooth pursuit movements are tracking movements of the eye as they follow moving objects. These occur voluntarily when the eyes track moving objects but take place involuntarily if a repetitive visual pattern is displayed continuously. When velocity of the moving object is more, the smooth pursuit movement is replaced by *small saccades* (*catch-up saccades*).

3. Vergence movement system

Vergence movements allow focussing of an object, which moves away from or towards the observer or when visual fixation shifts from one object to another at a different distance. Vergence movements are very slow (about 20°/s) disjugate movements. They have a latency of about 160 ms.

Table 11.1-3 Actions of extraocular muscles

Muscle	Primary action	Secondary action	Tertiary action
MR	Adduction	–	–
LR	Abduction	–	–
SR	Elevation	Intorsion	Adduction
IR	Depression	Extorsion	Adduction
SO	Intorsion	Depression	Abduction
IO	Extorsion	Elevation	Abduction

4. Vestibular eye movement system

Vestibular movements are usually effective in compensating for the effects of head movements in disturbing visual fixation. These movements operate through the vestibular system (see page 846).

5. Optokinetic system

This system helps to hold the images of the seen world steady on the retinae during sustained head rotation. This system becomes operative, when the vestibular reflex gets fatigued after 30 s. The optokinetic response is evoked by rotation of the visual field before the eyes. It consists of a movement following the moving scene, succeeded by a rapid saccade in the opposite direction.

6. Position maintenance system

This system helps to maintain a specific gaze position by means of rapid micromovements called *flicks* and slow micromovements called *drifts*. This system co-ordinates with other systems. **Neural pathway** for this system is believed to be the same as for saccades and smooth pursuits.

STRABISMUS AND NYSTAGMUS

Strabismus

Definition. Normally visual axes of the two eyes are parallel to each other in the 'primary position of gaze' and this alignment is maintained in all positions.

A misalignment of the visual axes of the two eyes is called squint or strabismus.

Nystagmus

It is defined as regular and rhythmic to and fro involuntary oscillatory movements of the eyes.

Aetiology. It occurs due to disturbance of the factors responsible for maintaining normal ocular posture. These include disorders of sensory visual pathway, vestibular apparatus, semicircular canals, mid brain and cerebellum. For details see page 846.

AQUEOUS HUMOUR AND INTRAOCULAR PRESSURE

AQUEOUS HUMOUR AND ITS PRODUCTION

Volume. The aqueous humour is a clear watery fluid filling the anterior chamber (0.25 mL) and posterior chamber (0.06 mL) of the eyeball.

Functions of aqueous humour are:

- It maintains a proper intraocular pressure.
- It plays an important metabolic role by providing substrates and by removing metabolites from the avascular cornea and lens.
- It maintains optical transparency.
- It takes the place of lymph that is absent within the eyeball.

Refractive index of aqueous humour is 1.336.

Composition. The composition of aqueous is similar to plasma except that it has:

- Proteins (colloid) content. The protein content of aqueous humour (5–16 mg/dL) is much less than that of plasma (6–7 g/dL) because of blood aqueous barrier.
- High concentrations of ascorbate, pyruvate and lactate.

Aqueous humour is derived from plasma within the capillary network of ciliary processes. The normal aqueous production rate is 2.3 mL/min. The three mechanisms: *diffusion, ultrafiltration and secretion* (active transport) play a part in its production.

DRAINAGE OF AQUEOUS HUMOUR

- Aqueous humour flows from the posterior chamber into the anterior chamber through the pupil against the slight physiologic resistance. From the anterior chamber, the aqueous is drained out by two routes:
- Trabecular (conventional outflow) and
- Uveoscleral (unconventional) outflow.

A. Trabecular (conventional) outflow

It is the main outlet for aqueous from the anterior chamber. Approximately 90% of the total aqueous is drained out via this route. The aqueous outflow system includes the trabecular meshwork, the Schlemm's canal, collector channels, aqueous veins and the episcleral veins (Fig. 11.1-57).

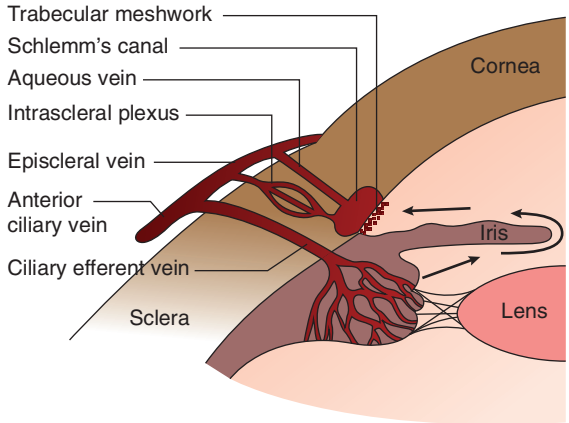


Fig. 11.1-57 The aqueous outflow system.

1. Trabecular meshwork. It is a sieve-like structure through which aqueous humour leaves the eye. It consists of three portions:

- Uveal meshwork,
- Corneoscleral meshwork and
- Juxtacanalicular (endothelial) meshwork.

2. Schlemm's canal. This is an endothelial lined oval channel present circumferentially in the scleral sulcus.

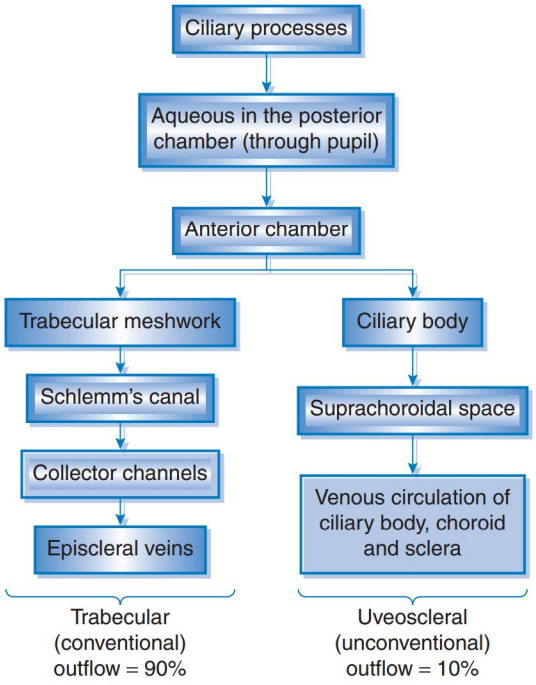
3. Collector channels. These, also called intrascleral aqueous vessels, are about 25–35 in number and leave the Schlemm's canal at oblique angles to terminate into the episcleral veins.

From the Schlemm's canal, the aqueous is transported via external collector channels into the episcleral veins by direct and indirect systems. A pressure gradient between intraocular pressure and intrascleral venous pressure (about 10 mm Hg) is responsible for unidirectional flow of aqueous.

B. Uveoscleral (unconventional) outflow

It is responsible for about 10% of the total aqueous outflow. Aqueous passes across the ciliary body into the suprachoroidal space and is drained by the venous circulation in the ciliary body, choroid and sclera.

The drainage of aqueous humour is summarized in the flowchart:



MAINTENANCE OF INTRAOCULAR PRESSURE

The intraocular pressure (IOP) refers to the pressure exerted by the intraocular fluids on the coats of the eyeball. The normal IOP varies between 10 and 21 mm Hg (mean 16±2.5 mm Hg). The normal level of IOP is essentially maintained by a dynamic equilibrium between the formation and outflow of the aqueous humour. Various factors influencing intraocular pressure are:

1. Age. The mean IOP increases after the age of 40 years, possibly due to reduced facility of aqueous outflow.

2. Diurnal variation of IOP. Usually, there is a tendency of higher IOP in the morning and lower in the evening. This has been related to the diurnal variation in the levels of plasma cortisol. Normal eyes have a smaller fluctuation (<5 mm Hg) than glaucomatous eyes (>8 mm Hg).

3. Blood pressure. As such it does not have long-term effect on IOP. However, prevalence of glaucoma is marginally more in hypertensives than the normotensives.

4. Osmotic pressure of blood. An increase in plasma osmolarity (as occurs after intravenous mannitol, oral glycerol or in patients with uraemia) is associated with a fall in IOP, while a reduction in plasma osmolarity (as occurs with water drinking provocative tests) is associated with a rise in IOP.

GLAUCOMA

Glaucoma is not a single disease process but a group of disorders in which intraocular pressure is raised above the tolerance limit of the affected eye, resulting in damage to the optic nerve head and irreversible visual field defects.

Types of glaucomas

1. Congenital and developmental glaucomas occur due to developmental anomalies of the angle of anterior chamber.

2. Primary glaucomas occur without any obvious systemic or ocular cause, usually after the age of 40 years which occurs due to sclerosis of trabecular meshwork or closure of the angle of anterior chamber.

3. Secondary glaucomas are characterized by the rise in intraocular pressure, secondary to some other ocular disease.

PHYSIOLOGY OF PUPIL

PUPILLARY REFLEXES

Light reflex

When light is shown in one eye, both the pupils constrict. Constriction of the pupil to which light is shown is called *direct light reflex* and that of the other pupil is called *consensual (indirect) light reflex.* Light reflex is initiated by rods and cones.

Pathway of light reflex (Fig. 11.1-58). The *afferent fibres* extend from retina to the pretectal nucleus in the mid brain. These travel along the optic nerve to the optic chiasma where fibres from the nasal retina decussate and travel along

the opposite optic tract to terminate in the contralateral pretectal nucleus. While the fibres from the temporal retina remain uncrossed and travel along the optic tract of the same side to terminate in the ipsilateral pretectal nucleus.

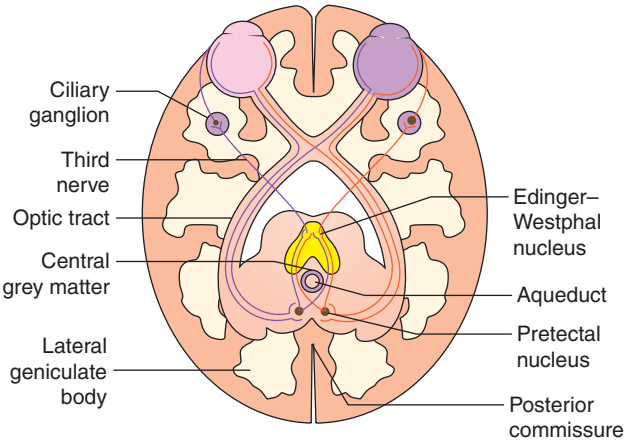


Fig. 11.1-58 Pathway of the light reflex.

Internuncial fibres connect each pretectal nucleus with the Edinger–Westphal nuclei of both sides. This connection forms the basis of consensual light reflex.

Efferent pathway consists of the parasympathetic fibres which arise from the Edinger–Westphal nucleus in the mid brain and travel along the third (oculomotor) cranial nerve. The pre-ganglionic fibres enter the inferior division of the third nerve and via the nerve to the inferior oblique reach the ciliary ganglion to relay. Post-ganglionic fibres travel along the short ciliary nerves to innervate the sphincter pupillae.

Accommodation reflex

During accommodation, when eyes are focused from distant to near object to achieve clear vision, three reactions occur; wiz.

- Changes in the radius of curvature of the lens (more convex) by contraction of ciliary muscles,

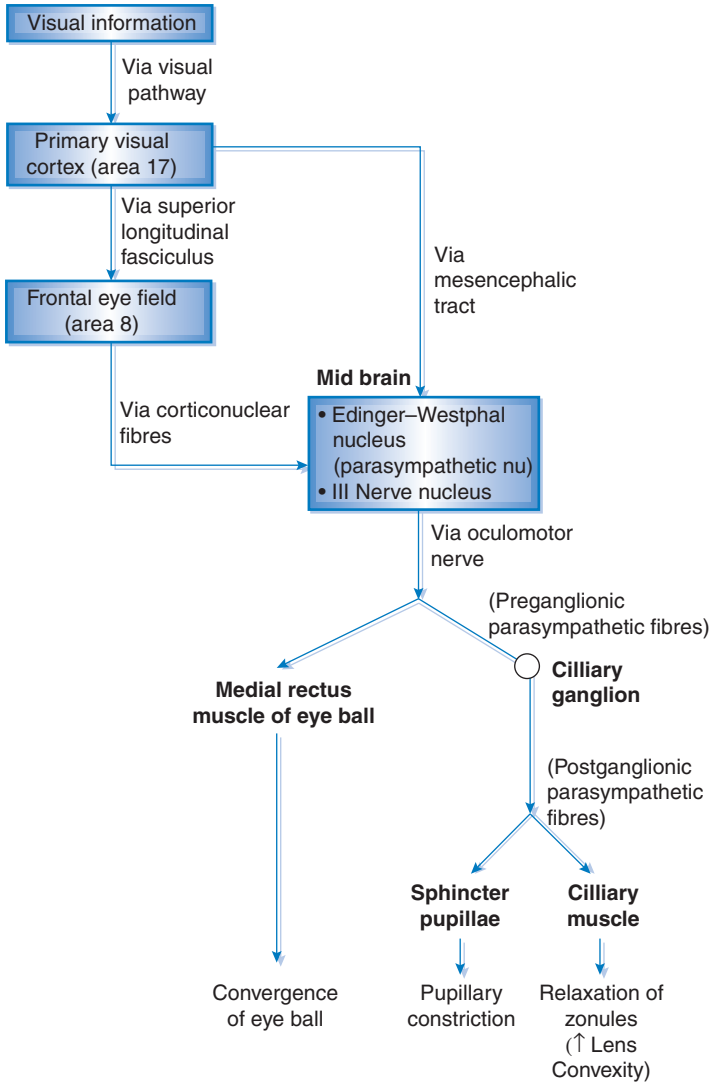


Fig. 11.1-59 Pathway of the near reflex.

- Pupillary constriction (meiosis) by contraction of sphincter pupillae and
- Convergence of eyes due to contraction of medial recti of eye balls.

Pathway of accommodation reflex (Fig. 11.1-59). The afferent impulses extend from the retina to the visual cortex via the optic nerve, chiasma, optic tract, lateral geniculate body, optic radiations and striate cortex. From the parastriate cortex, the impulses are relayed to the Edinger–Westphal nucleus of both sides via the occipitomesencephalic tract. From the Edinger–Westphal nucleus, the efferent impulses travel along the 3rd nerve and reach the sphincter pupillae and ciliary muscle after relaying in the ciliary ganglia.

For convergence reaction impulses from visual cortex reach the frontal eye field (area 8) through association fasciculus (superior longitudinal fasciculus). From frontal eye field the corticonuclear fibres project to the third nerve nuclei of both sides and through oculomotor nerves supply the medial recti of the eye balls.

Psychosensory reflex

It refers to the dilatation of the pupil in response to the sensory and psychic stimuli. It is very complex and its mechanism is still not elucidated.

ABNORMALITIES OF PUPILLARY REACTIONS

1. Efferent pathway defect. Absence of both direct and consensual light reflex on the affected side (say right eye) and presence of both direct and consensual light reflex on the normal side (i.e. left eye) indicates efferent pathway defect (sphincter paralysis). Near reflex is also absent on the affected side. Its causes include: effect of parasympatholytic drugs (e.g. atropine, homatropine), internal ophthalmoplegia and third nerve paralysis.

2. Wernicke's hemianopic pupil. It indicates lesion of the optic tract. In this condition, light reflex (ipsilateral direct and contralateral consensual) is absent when light is thrown on the temporal half of the retina of the affected side and nasal half of the opposite side; while it is present when the light is thrown on the nasal half of the affected side and temporal half of the opposite side.

3. Argyll Robertson pupil (ARP). Here the pupil is slightly small in size and reaction to near reflex is present but light reflex is absent, i.e. there is light near dissociation (to remember, the acronym ARP may stand for 'accommodation reflex present'). Both pupils are involved and dilate poorly with mydriatics. It is caused by a lesion (usually neurosyphilis) in the region of tectum.

Chapter 11.2

Sense of Hearing

FUNCTIONAL ANATOMY

- **The ear**
  - External ear
  - Middle ear
  - Internal ear
- **Auditory pathways**
  - Spiral ganglion
  - Superior olivary complex
  - Inferior colliculus
  - Medial geniculate body
  - Auditory cortex

PHYSIOLOGY OF HEARING

- **Stimuli or sound waves**
  - Definitions
  - Physical properties
- **Conduction of sound waves**
  - Role of external ear
  - Role of middle ear
    - Impedance matching
    - Phase differential between oval and round window
    - Natural resonance of external and middle ear
    - Attenuation reflex
- **Transduction of sound waves**
  - Vibration of basilar membrane
  - Stimulation of the hair cells
  - Membrane potential changes in the hair cells
- **Neural transmission of signals**
  - Salient features of auditory pathway
- **Neural processing of auditory information**
  - Encoding of sound frequency
  - Encoding of intensity
  - Feature detection
  - Localization of sound in space

APPLIED ASPECTS

- Noise and masking
- Hearing loss, deafness and tinnitus
- Hearing tests

FUNCTIONAL ANATOMY

THE EAR

The mechanism of hearing is closely associated with the mechanism of equilibrium; therefore, the inner ear acts as an organ of hearing and equilibrium. For hearing, the sound waves have to pass through the three subdivisions of the ear, which are (Fig. 11.2-1):

- External ear,
- Middle ear and
- Internal ear.

EXTERNAL EAR

The external ear consists of the pinna (auricle) and the external auditory meatus.

Pinna or auricle consists of a single convoluted plate of elastic cartilage covered by skin, which is tightly attached to the underlying perichondrium.

Functions. The pinna collects and reflects the sound waves into the external auditory canal.

- In lower animals, pinna is more important, in whom it can be moved by muscular action in the direction of sound source.
- In humans, the pinna is not moveable, but its peculiar shape aids in *discerning* the source of sound (e.g. in front of versus behind the head).

External auditory meatus extends from the pinna to the tympanic membrane. It consists of two distinct portions: external one-third, cartilaginous portion and internal two-thirds, bony portions. It is lined with skin, which in the cartilaginous part secretes wax (from the ceruminous glands) and oil (from the sebaceous glands).

MIDDLE EAR

Walls of middle ear

The middle ear or the tympanic cavity is a six-sided air-filled rectangular space in the petrous part of the temporal

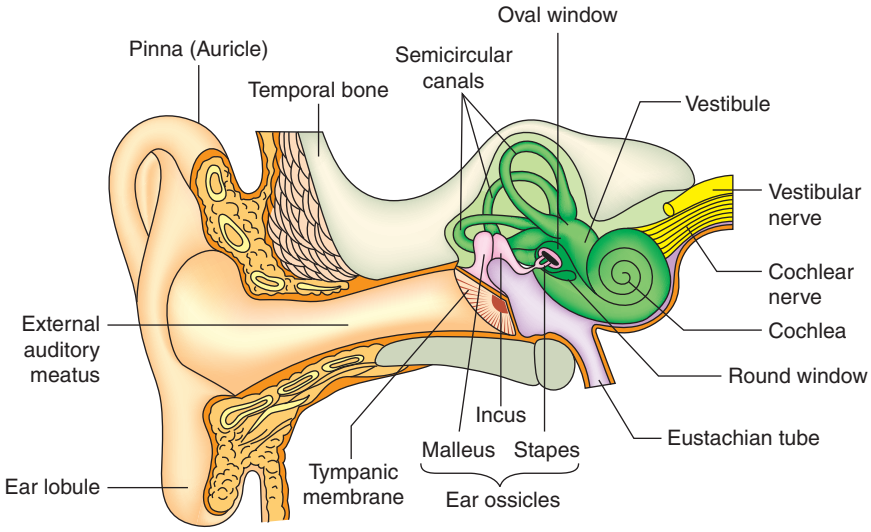


Fig. 11.2-1 Structure of three subdivisions of the ear.

bone with a roof, a floor, and an anterior, a posterior, a medial and a lateral wall.

Lateral wall is formed by a tympanic membrane, which shuts the medial end of external auditory meatus. *The tympanic membrane* is a cave-shaped structure with concavity directed towards the external auditory meatus. Point of maximum convexity is called umbo. It consists of a connective tissue covered with skin on the outside and mucous membrane on the inside.

Anterior wall contains two canals; upper one lodges tensor tympani muscle and lower one lodges the eustachian tube.

Eustachian tube or the pharyngotympanic tube connects the middle ear cavity with pharynx. Air can pass through this tube into the middle ear. Therefore, it serves the function of equalization of pressure on two sides of *the tympanic membrane*. When its pharyngeal opening is blocked, e.g. in common cold, the air cannot pass into the tympanic cavity. The air present in the middle ear gets absorbed by the mucous membrane and the tympanic membrane is retracted inwards. As a result, the vibrations of the tympanic membrane get decreased or abolished, causing discomfort and loss of hearing.

Posterior wall of the middle ear communicates with the air cavities in the mastoid process.

Medial wall (labyrinthine wall). It contains two windows:

- *Oval window* (fenestra vestibuli) is present above, in which foot plate (face plate or stapes) is attached. It leads to the vestibule of the internal ear and transmits the sound vibrations of the ossicles to the perilymph of scala vestibuli.
- *Round window* (fenestra cochlea) is present in the lower part and is closed by a thin membrane called *secondary tympanic membrane*. It accommodates the pressure waves transmitted to the perilymph of the scala tympani.

Roof. It is formed by a thin bone called the tegmen tympani of the petrous temporal bone and separates the middle ear from the middle cranial fossa.

Floor. It is formed by a convex plate of bone, which separates the middle ear from jugular fossa, which lodges the superior bulb of the internal jugular vein.

Ear ossicles

The three ear ossicles (auditory ossicles) include malleus, incus and stapes. They are attached to each other by ligaments and form a chain (Fig. 11.2-2).

Malleus. It resembles a mallet (hammer) and consists of a head, neck and three processes: the handle or manubrium, the lateral and anterior processes.

- *Manubrium* (handle) of the malleus is connected to the inner surface of the tympanic membrane.
- *Head* articulates with incus posteriorly.

Incus is the middle ossicle that resembles an anvil in shape. It consists of a body and two processes. The body of incus articulates with the head of malleus.

Stapes. It resembles a stirrup. Its head articulates with the incus and the oval footplate contacts the membrane of the oval window of the cochlea.

Muscles of the middle ear

The middle ear contains two muscles: the tensor tympani and stapedius.

Tensor tympani. It arises from the wall of the semicanal for tensor tympani and is inserted into the handle of the malleus. It is innervated by a branch of mandibular division of fifth cranial nerve. It constantly pulls the handle of malleus

inwards and thus keeps the tympanic membrane tensed. Due to this, vibrations on any portion of the tympanic membrane are transmitted to the malleus.

Stapedius. It arises from the posterior wall of the middle ear and is inserted on the neck of stapes. It is innervated by a branch from the facial nerve and on contraction it pulls the footplate of stapes out from the oval window.

Function. Both muscles of the middle ear act simultaneously and reflexly in response to loud sound and attenuate the sound (see also page 932).

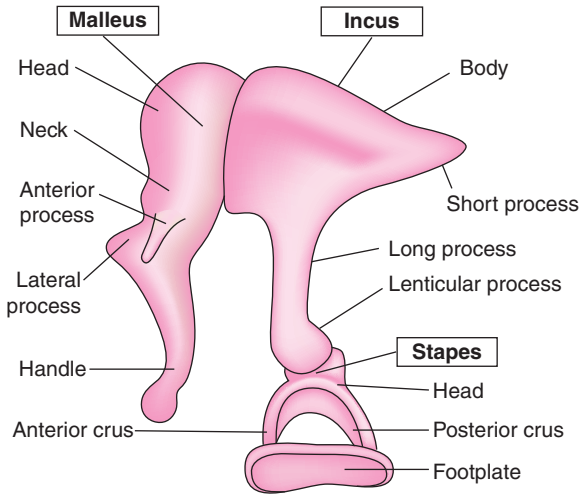


Fig. 11.2-2 Ear ossicles and their parts.

INTERNAL EAR

The internal ear or labyrinth is situated in the petrous part of the temporal bone. It consists of a bony labyrinth and a membranous labyrinth (Fig. 11.2-3).

Bony labyrinth consists of three parts: vestibule, semicircular canals and the cochlea.

Membranous labyrinth is lodged within the bony labyrinth. It is filled with endolymph (which resembles intracellular fluid) and is surrounded by perilymph (which resembles extracellular fluid in its composition. The inner ear can be divided into two main parts:

1. *Vestibular receptor apparatus.* It consists of (Fig. 11.2-3):

- *Utricle and saccule*, which are lodged in the bony vestibule and are collectively called *otolith organs.*
- *Semicircular ducts*, which lie within the body of semicircular canals.

Vestibular apparatus is concerned with equilibrium and is described on page 842.

2. *Auditory receptor apparatus* is formed by the *duct of cochlea,* which lies within the bony cochlea.

Auditory apparatus

The bony cochlea containing membranous cochlear duct (which houses the organ of Corti) forms the so-called auditory apparatus (Fig. 11.2-3).

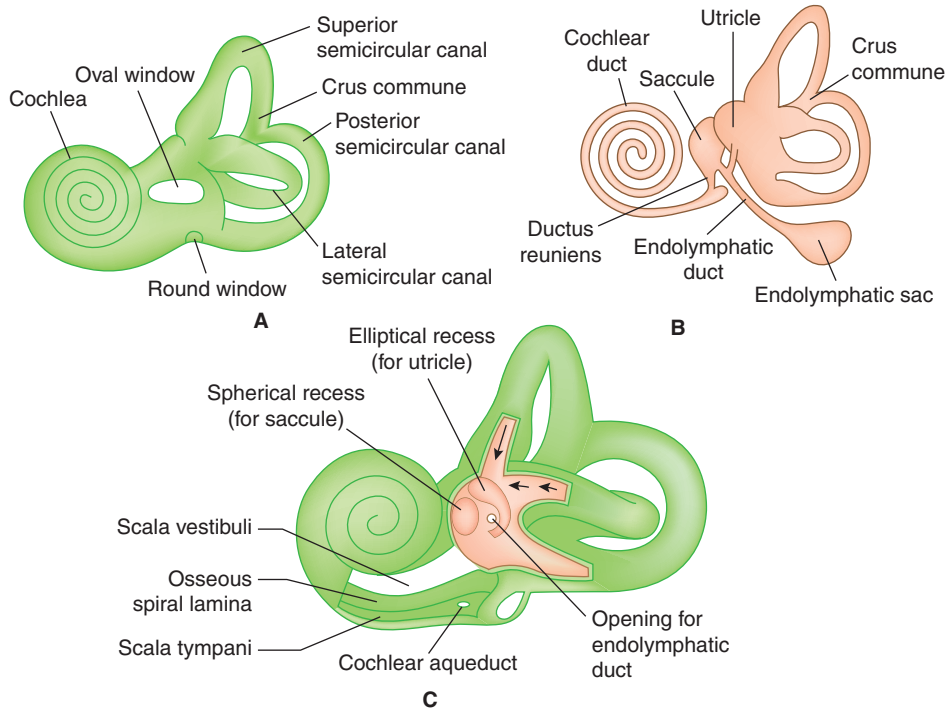


Fig. 11.2-3 Structure of inner ear: A, left bony labyrinth; B, left membranous labyrinth and C, cut section of bony labyrinth.

Bony cochlea is a spiral tube, which in humans has a two and three-fourth turns around a central bone called the *modiolus*. The base of the modiolus is directed towards internal acoustic meatus and transmits vessels and nerves to the cochlea. Around the modiolus and winding spirally, like the thread of a screw, is a thin plate of bone called *osseous spiral lamina*. It divides the bony cochlea incompletely and gives attachment to the basilar membrane. Two membranes (basilar membrane and Reissner's membrane) divide the bony cochlea into three compartments (Fig. 11.2-4):

- Scala vestibuli,
- Scala media (membranous cochlear duct) and
- Scala tympani.

Scala vestibuli and scala tympani are filled with perilymph and communicate with each other at the apex of cochlea through an opening called *helicotrema*.

- *Scala vestibuli* is separated from the scala media by *Reissner's membrane* and is closed by the footplate of stapes, which separates it from the air-filled middle ear (Figs 11.2-4 and 11.2-5).
- *Scala tympani* is separated from the scala media by the *basilar membrane* and is closed by *secondary tympanic membrane*. It is also connected with a subarachnoid space through the aqueduct of cochlea (Figs 11.2-4 and 11.2-5).
- *Scala media* or *cochlear duct* or *membranous cochlea* appears triangular on cross-section. Its three walls are formed by (Fig. 11.2-4):
- *Basilar membrane*, which is attached medially to the osseous spiral lamina and laterally to the fibrous spiral ligament (which lines the bony cochlea), forms the inferior wall of the cochlear duct. The basilar membrane supports the *organ of Corti*.

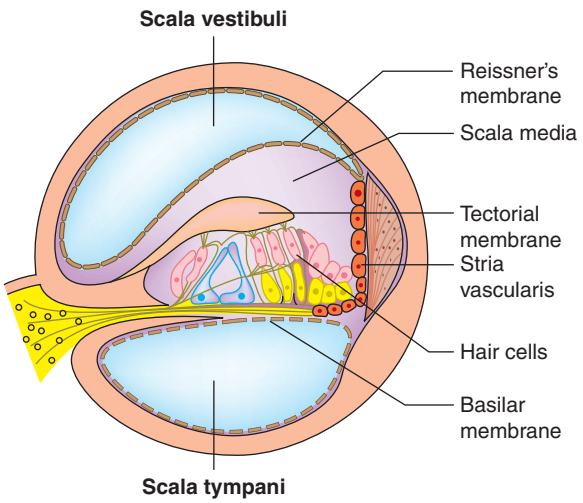


Fig. 11.2-4 Vertical section through cochlea showing scala vestibuli, scala media (cochlear duct) and scala tympani. Note the location of organ of Corti over the basilar membrane in the cavity of cochlear duct.

- *Reissner's membrane*, which is attached medially to the wall of limbus and laterally to the upper margin of stria vascularis, forms the superior wall of the cochlear duct.
- *Stria vascularis* forms the lateral wall of the cochlear duct. It consists of vascular epithelium and is concerned with the secretion of *endolymph*.

Perilymph and endolymph

Perilymph is the fluid present in the scala tympani and scala vestibuli compartments of the cochlea. Its composition is similar to the extracellular fluid in that it is high in Na^+ and low in K^+.

Endolymph is the fluid present within the scala media or the membranous cochlea. Its composition is similar to the intracellular fluid in that it is high in K^+ and low in Na^+. It is secreted by the stria vascularis, which forms the lateral wall of the scala media.

Organ of Corti

The organ of Corti, the sense organ of hearing, is situated on the top of the basilar membrane in the scala media (Fig. 11.2-4). It contains the auditory receptors or the peripheral receptors of sense of hearing. Important components of the organ of Corti are (Fig. 11.2-6):

1. Rods of Corti. These are two projections (inner and outer rods) from the basilar membrane into the scala media. In between the two rods is the *tunnel of Corti,* which contains a fluid called cortilymph. The exact function of the rods and cortilymph is not known.

2. Hair cells. Hair cells are the receptor cells that transduce sound energy into electrical energy. Two groups of hair cells lie on the basilar membrane (Fig. 11.2-6).

- *Inner hair cells*. These form a single row of cells internal (i.e. medial) to the inner rod. These are about 3500

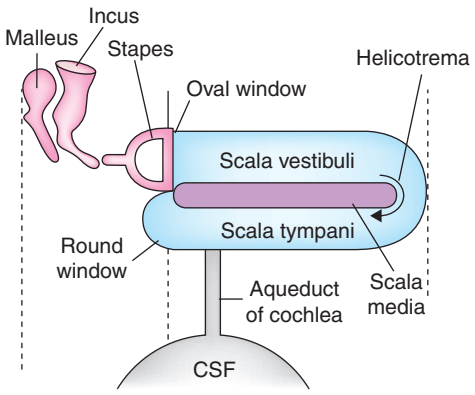


Fig. 11.2-5 Diagrammatic depiction of arrangement of perilymphatic system, three compartments of cochlea (scala vestibuli, scala media, and scala tympani), and ear ossicles of the middle ear. Note the CSF passes into scala tympani through aqueduct of cochlea.

in number. These cells are probably more important in the transmission of the auditory impulses. These are responsible for fine auditory transmission.

- *Outer hair cells*. These are about 20,000 in number and are arranged in three or four layers external (i.e. lateral) to the outer rod. These are responsible for *detecting the presence of sound*. They mainly receive efferent innervation from the olivary complex and are concerned with modulating the function of inner hair cells.

Structure of a hair cell. The inner hair cells are flask shaped while the outer hair cells are cylindrical. On the upper surface of the hair cells are present tiny cilia (*stereo-cilia*) which protrude into the overlying tectorial membrane.

3. Supporting cells. of following type are known (Fig. 11.2-6):

- Inner phalangeal cells support the inner hair cells.
- Deiter's cells (outer phalangeal cells) are situated between the outer hair cells and provide support to the latter.
- Hensen's cells lie outside the Deiter's cells.
- Claudius cells lie outside the Hensen's cells.

4. Tectorial membrane. It consists of gelatinous matrix with delicate fibres. It is a thin but stiff membrane made of glycoprotein material. This membrane is attached to the upper surface of spiral lamina and its free edge extends just beyond the outermost neuroepithelial cells.

The shearing force between the hair cells and tectorial membrane produces the stimulus to hair cells.

Nerve supply of hair cells. *Afferent fibres* supplying the hair cells constitute the cochlear division of eighth cranial nerve. Cell bodies of these fibres are located in the spiral ganglion.

- *Inner hair cells* receive 90–95% of the afferent fibres and so are more important in the transmission of auditory impulses.
- *Outer hair cells* receive only 5–10% of the afferent fibres.

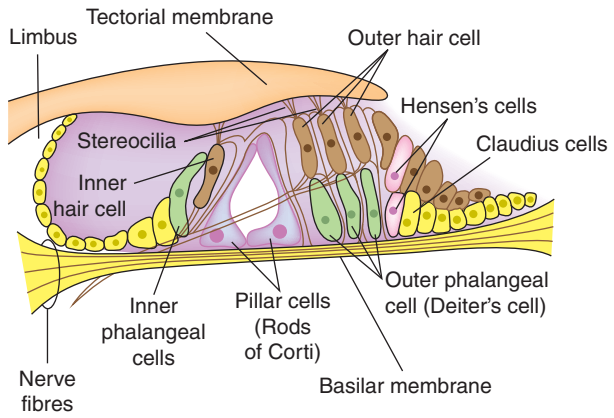


Fig. 11.2-6 Structure of organ of Corti. Note the connections between tectorial membrane and cilia of hair cells.

Note. There are about 30,000 fibres in each auditory nerve, so there is no net convergence of receptors on the first-order neurons. Most of the afferent fibres, however, supply more than one hair cell and conversely, most of hair cells are supplied by more than one fibre.

Efferent fibres to the hair cells come from both the ipsilateral and contralateral sides via the olivocochlear bundle. Their cell bodies are situated in the superior olivary complex. These fibres descend to join the eighth nerve. Outer hair cells receive most of the efferent fibres, while the inner hair cells receive only a few efferent fibres. The efferent fibres are cholinergic and cause inhibition of the afferent fibres by liberating a hyperpolarizing mediator, which is probably acetylcholine (Ach). Thus, the outer hair cells are mainly concerned with modulation of the function of inner hair cells.

AUDITORY PATHWAYS

Auditory pathways comprise following relay stations (Fig. 11.2-7):

- Spiral ganglion,
- Superior olivary nucleus complex, trapezoid nucleus and nucleus of lateral lemniscus,
- Inferior colliculus,
- Medial geniculate body and
- Auditory cortex.

Spiral ganglion

First-order neurons are the bipolar cells of the *spiral ganglion,* which is situated in the Rosenthal's canal (canal running along osseous spiral lamina).

- *Dendrites* of these bipolar cells constitute the afferent fibres innervating the hair cells.
- *Axons* of these bipolar cells form the cochlear division of eighth cranial nerve. The cochlear nerve ends in the cochlear nuclei in the medulla.

Cochlear nuclei. ***Second-order neurons*** have their cell bodies in the cochlear nuclei, which are situated in the rostral part of the medulla. There are two cochlear nuclei: dorsal and ventral.

The axons of second-order neurons from the cochlear nuclei pass medially in the dorsal part of pons. Most of them cross to the opposite side, but some remain uncrossed.

- The crossing fibres of two sides form a conspicuous mass of fibres called the *trapezoid body*.
- Some crossing fibres run separately in the dorsal part of pons and do not form part of the trapezoid body.

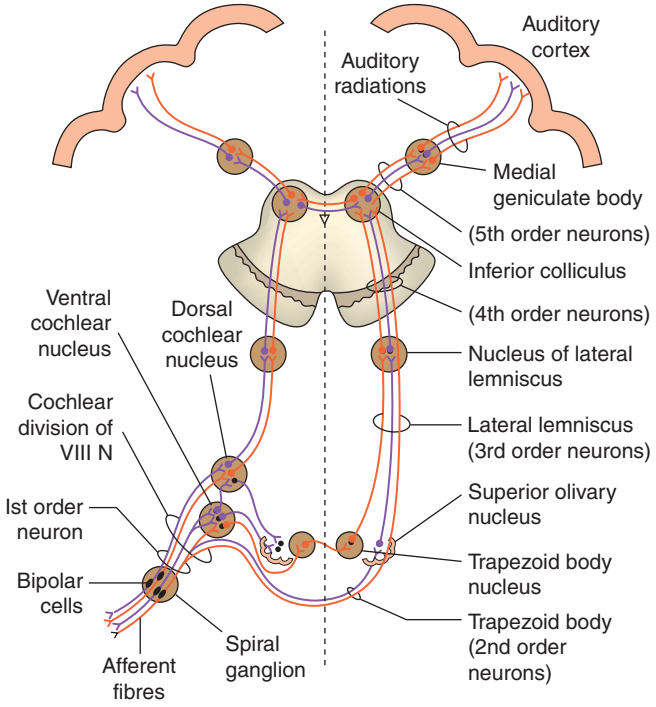


Fig. 11.2-7 Central auditory pathways.

Superior olivary nucleus complex, trapezoid nucleus and nucleus of lateral lemniscus

Third-order neurons have their cell bodies mainly in the *superior olivary complex* (made up of a number of nuclei) and also in trapezoid nucleus and nucleus of lateral lemniscus.

- *Superior olivary nuclear complex* receive the large majority of lateral lemniscus fibres from the cochlear nuclei. Axons arising from the superior olivary complex form an important ascending bundle called the *lateral lemniscus.*
- *Trapezoid nucleus.* Some cochlear fibres that do not relay in the superior olivary nucleus join the lateral lemniscus after relaying in the scattered groups of cells lying within the trapezoid body (which constitute the *trapezoid nucleus*).
- *Nucleus of lateral lemniscus* refers to the collection of cells that lie within the lemniscus itself. Some cochlear fibres relay in these cells.

The fibres of *lateral lemniscus* ascend to the mid brain and terminate in the inferior colliculus.

Inferior colliculus

Fourth-order neurons have their cell bodies in the *inferior colliculus,* where the fibres of lateral lemniscus terminate. Fibres arising in the inferior colliculus enter the inferior brachium to reach the medial geniculate body.

Medial geniculate body

Fifth-order neurons have their cell bodies in the medial geniculate body where most of the fibres arising in inferior colliculus terminate. Some fibres from the lateral lemniscus reach this body without relay in the inferior colliculus. Fibres arising in the medial geniculate body form the *acoustic radiation,* which ends in the acoustic area of the cerebral cortex.

Auditory cortex

Major areas constituting auditory cortex present in the temporal lobe are:

- Primary auditory cortex (areas 41 and 42) and
- Auditory association areas (areas 22, 21 and 20), for details see page 757.

PHYSIOLOGY OF HEARING

Hearing, i.e. detection of sound waves, may serve to warn of impending danger or localize friends. But most importantly, audition allows social communication. Physiology of audition can be discussed under following headings:

- Stimuli or sound waves,
- Conduction of sound waves,
- Transduction of sound waves,
- Neural transmission of signals and
- Encoding of signals.

STIMULI OR SOUND WAVES

Definition

Stimuli for the receptors of hearing are sound waves. Sound is a form of energy produced by a vibrating object. A sound wave consists of alternating phases of compression and rarefaction of molecules of the medium (air, liquid, or solid) in which it travels.

Physical properties of sound

Physical properties of sound and certain terms, which are frequently used in audiology and acoustics are (Fig. 11.2-8):

1. Speed of sound. Speed or velocity of the sound waves is different in different media:

- *In the air*, at 0°C, at sea level, sound travels at a rate of approximately 330 m/s (1100 ft/s), while at 20°C it travels at a rate of 349 m/s (1150 ft/s).
- *In the water*, at 20°C, sound travels at much faster speed of 1450–1500 m/s. The speed is faster in salt water as compared to the fresh water. Further, the speed of sound in water slightly increases with temperature and altitude.

2. Frequency of sound refers to the number of waves per second.

- The *unit* of frequency is hertz (Hz).
- Range of human hearing is approximately 20–20,000 Hz.
- Range of average speaking voice is approximately 2000–5000 Hz.

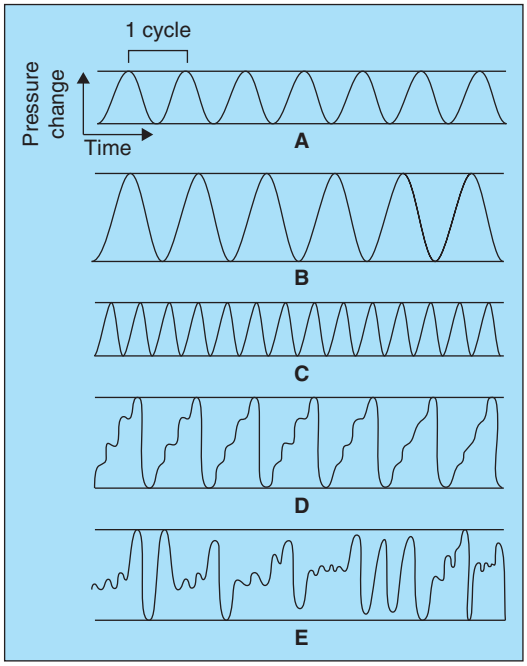


Fig. 11.2-8 Characteristics of sound wave: A, pure tones; B, increase in amplitude (intensity) of sound wave thus the sound is louder; C, increase in frequency (pitch); D, complex waveform due to mixture of pure tones and overtones determines the quality of sound (timbre) and E, aperiodic irregular waveform (noise).

3. Amplitude (intensity) of sound is the strength which determines its loudness. The intensity of sound is measured in terms of maximum pressure change at the tympanic membrane, which is more commonly expressed as *sound pressure level* (SPL). The unit of SPL is decibel (dB), which is expressed as:

$$dB = 20.\log \frac{P_S\,(\text{Pressure of the stimulus sound})}{P_R\,(\text{Pressure of the reference sound})}$$

Reference (standard) pressure is 0.0002 dynes/cm^2. It is defined as a sound pressure that is just detectable, i.e. auditory *threshold*.

Thus, sound intensities are measured on a ratio scale using a subjective intensity (i.e. the threshold), rather than an arbitrary intensity, as a base.

At a distance of 1 m, intensity of some common sounds is:

- Whisper : 30 dB
- Normal conversation : 60 dB
- Rock music : 90 dB
- Discomfort of the ear is produced by sounds of : 120 dB
- Pain in the ear is produced by sound of above (10^7 times threshold) : 140 dB

4. Pure tone. A single frequency sound is called a pure tone, e.g. a sound of 25–50 or 100 Hz.

5. Complex sound refers to that with more than one frequency. For example, human voice is a complex sound.

6. Pitch. It is the subjective sensation produced by the frequency of sound. Higher the frequency greater is the pitch.

- Pitch of average male voice is 120 Hz and that of female is 250 Hz.
- An average individual can distinguish about 2000 different pitches.

7. Overtones. A complex sound is a mixture of pure tones. The lowest frequency at which a source vibrates is called *fundamental (or primary) frequency*. All other frequencies, which are multiples of the fundamental frequency are called *overtones* or *harmonics*. The overtones determine the quality or the timbre of sound.

Variations in the quality (timbre) permit us to identify the sounds of various musical instruments (e.g. guitar, piano, *tabla*, *sarangi* etc.) even though they are playing notes of the same pitch.

CONDUCTION OF SOUND WAVES

Role of external ear

External ear captures the sound waves.

Pinna, collects and reflects the sound waves into the external auditory meatus. Its peculiar shape, in humans, aids in discerning the source of sound (e.g. in front versus behind the head).

External auditory meatus conducts the sound waves to the tympanic membrane. It is S-shaped course:

- Helps in amplifying the sound waves.
- Prevents mechanical injury to the tympanic membrane and
- Helps in maintaining favourable temperature and humidity for normal functioning of the tympanic membrane.

Role of middle ear

Conduction of sound stimulus by the tympanic membrane to ear ossicles

The sound waves that pass through the pinna and external auditory meatus strike the tympanic membrane. The vibrating tympanic membrane causes the ear ossicles to vibrate. Thus, the tympanic membrane acts as:

- *Pressure receiver*, i.e. it is extremely sensitive to pressure changes produced by the sound waves.
- *Resonator*, i.e. starts vibrating with pressure changes produced by the sound waves.
- *Critically dampens*, i.e. the vibrations of tympanic membrane cease immediately after the end of sound.

Conduction of sound waves mechanically from the middle ear to the inner ear

Various mechanisms which play role during conduction of sound waves mechanically from the middle ear to the inner ear are:

- Impedance matching mechanism,
- Phase differential between oval and round window,
- Natural resonance of external ear and middle ear and
- Attenuation reflex.

Impedance matching mechanism. The air-filled middle ear conducts sound waves mechanically to the fluid-filled internal ear through the ossicular system. Effective transfer of sound energy from an air to a fluid medium is difficult because most of the sound is reflected as a result of the different mechanical properties of the two media, i.e. *impedance mismatching*. This fact can be appreciated by the observation that a person under water cannot hear any sound made in the air. This happens because 99.9% of the sound energy is reflected away from the surface of water because of the impedance offered by it. Exactly, a similar situation exists in the ear when the air-filled middle ear has to conduct the sound to the fluid-filled inner ear. Nature has compensated for it by providing impedance matching mechanism to the middle ear.

- The middle ear functions as an impedance matching device, primarily by amplifying the sound pressure. It is accomplished by following three mechanisms (Fig. 11.2-9):

1. Lever action of ossicles. Handle of malleus is 1.3 times longer than the long process of incus, this provides a mechanical leverage advantage, due to which the middle ear ossicles increase the force of movement by 1.3 times.

2. Hydraulic action of the tympanic membrane is exerted because the effective vibratory area of the tympanic membrane (about 45 mm^2) is much greater than the stapes—oval window surface area (about 3.2 mm^2). This size difference, means the force produced by the sound is concentrated over a smaller area, thus amplifying the pressure exerted on the oval window (14 folds).

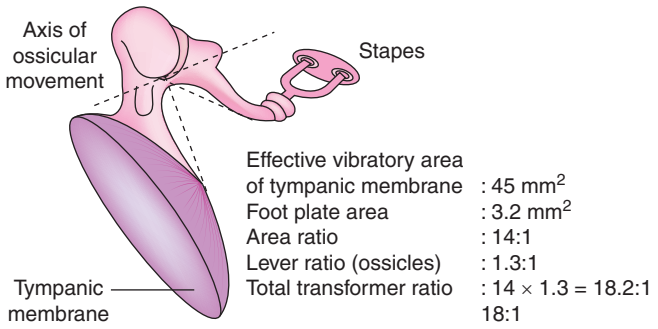


Fig. 11.2-9 Amplification of sound pressure by the combined hydraulic effect of tympanic membrane and leverage effect of ear ossicles.

3. Curved membrane effect. Movements of the tympanic membrane are more at the periphery than at the centre where malleus handle is attached. This too provides some leverage.

Thus, the above three mechanisms together increase the sound pressure 18 folds (i.e. 14 × 1.3). In this way, the impedance mismatching between the air-filled middle ear and fluid-filled inner ear is mostly compensated. Therefore, when the tympanic membrane and the ossicles are removed, and the sound waves strike the oval window directly, even very loud sounds are heard as whispers.

Minimum audibility curve. It is important to note that:

- Amplification of sound intensity is greatest between 1000 and 3000 Hz.
- Sounds below 16 Hz or above 20,000 Hz are not amplified at all.
- Because of the above, the human ear can perceive pitch of sound between 16 and 20,000 Hz, but maximum sensitivity is between 1000 and 3000 Hz. This effect is the basis of so-called minimum audibility curve (Fig. 11.2-10).

Phase differential between oval and round window. Sound waves striking the tympanic membrane do not reach the oval and round windows simultaneously. There is a preferential pathway to the oval window because of the ossicular chain. Thus, when oval window is receiving wave of compression, the round window is at the phase rarefaction. If the sound waves were to strike both the windows simultaneously, they would cancel each other's effect with no movement of perilymph and no hearing. This acoustic separation

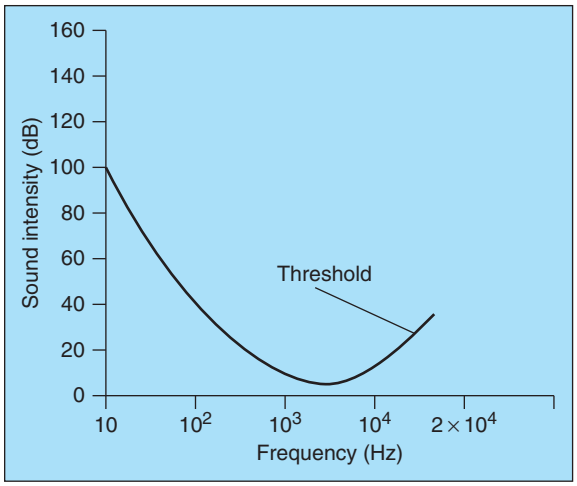


Fig. 11.2-10 Minimum audibility curve.

of windows is achieved by the presence of intact tympanic membrane and a cushion of air in the middle ear around the round window.

Natural resonance of external and middle ear. The external ear and middle ear, due to the inherent anatomic and physiologic properties, allow certain frequencies of sound to pass more easily to the inner ear. The natural resonance of different structures is:

- External auditory canal : 3000 Hz
- Tympanic membrane : 800–1600 Hz
- Middle ear : 800 Hz
- Ossicular chain : 500–2000 Hz

Thus (from the above) the greatest sensitivity of the sound transmission is between 500 and 3000 Hz, and these are the frequencies most important to human in day-to-day conversation.

Attenuation reflex. Attenuation reflex, also called tympanic reflex or acoustic reflex, is a preventive reflex which reduces sound pressure amplitude by affecting the mobility and transmission properties of the auditory ossicles.

Stimulus for this reflex is loud sound.
Latent period is 40–80 ms.
Reflex activity. The two muscles of the middle ear (tensor tympani and stapedius) contract reflexively in response to the intense sound.

Contraction of tensor tympani muscle pulls the malleus inwards whereas contraction of stapedius muscle pulls stapes outwards. These two opposing forces make the ossicular system very rigid and therefore it fails to vibrate with the sound waves. Thus, sound is not allowed to enter inner ear (i.e. is attenuated or intensity is reduced by 30–40 decibel).

Advantages of attenuation reflex are:

- It prevents occurrence of damage to the cochlea from the intense sounds like that of loud music, of jet aircraft, etc.
- It attenuates and masks all the low frequency environmental sounds and allows the person to concentrate on the sound above 1000 Hz, where most of the prominent information in voice communication is transmitted.
- It occurs just prior to vocalization and chewing, which suggests that the middle ear muscles may act to reduce the intensity of the sounds produced by these activities.

Note. Because the latent period of attenuation reflex is 40–80 ms, sudden, brief, extremely loud sound, such as due to bomb explosion or gun shot is likely to cause deafness due to damage to the cochlea.

TRANSDUCTION OF SOUND WAVES

Transduction of mechanical sound wave into electrical signal occurs in the organ of Corti of inner ear. Steps involved in the process of transduction are:

Vibration of basilar membrane

Sound waves from the middle ear are passed on to the inner ear through the oval window by in-and-out motion of the stapes (Fig. 11.2-11):

- Sound waves entering the inner ear from the oval window spread along the scala vestibuli as a travelling wave.
- Most of the sound energy is transferred directly from the scala vestibuli to the scala tympani. Very little of the sound wave ever reaches the helicotrema at the apex of cochlea.
- As the sound energy passes from the scala vestibuli to the scala tympani, it causes the basilar membrane to vibrate. It is important to note that the part of the cochlea where height of pressure wave reaches its maximum varies with the frequency of sound [travelling wave theory of Von Bekesy (see page 936)].

Stimulation of the hair cells

The up-and-down movements of the basilar membrane in turn cause the organ of Corti to vibrate up and down. The tops of the hair cells in the organ of Corti are held rigid by the reticular lamina and the hair of the outer hair cells are embedded in the tectorial membrane (Fig. 11.2-12). Because the tectorial and basilar membranes are attached at different points on the limbus (Fig. 11.2-12A), they slide past each other as they vibrate up and down.

Owing to the shear forces set up by the relative displacement of the basilar membrane and the tectorial membrane, the stereocilia of hair cells bend back and forth as:

- When the organ of Corti moves up, the tectorial membrane slides forward relative to the basilar membrane,

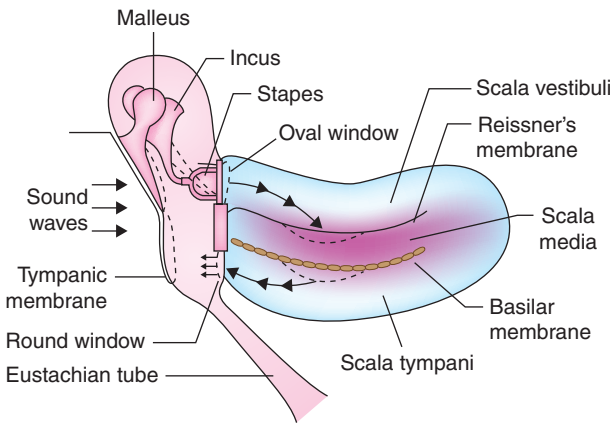


Fig. 11.2-11 Diagrammatic depiction of the mechanism of vibrations of basilar membrane produced by in and out motion of stapes (for details see text).

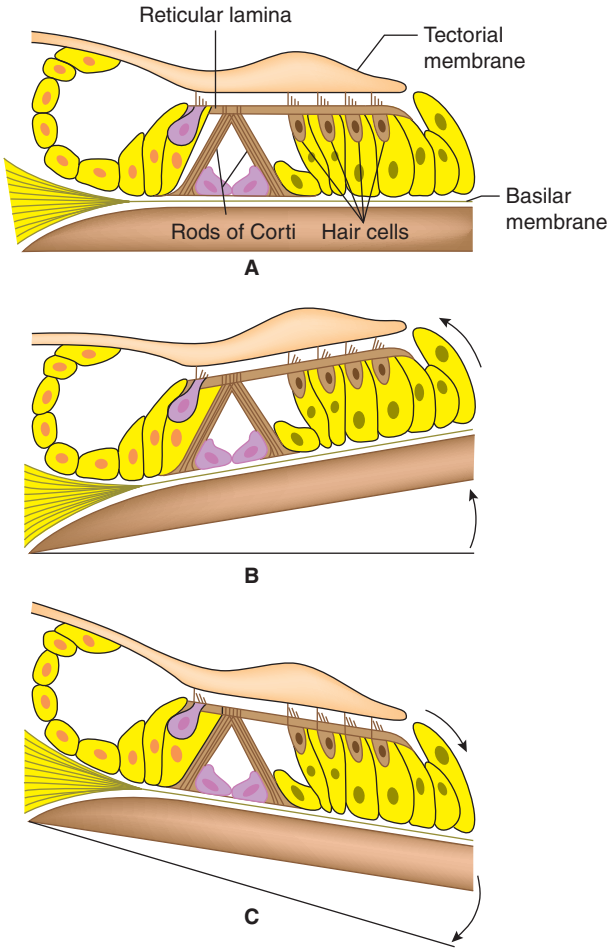


Fig. 11.2-12 Demonstration of bending of stereocilia with movement of organ of Corti: A, the tectorial membrane and basilar membrane are attached to limbus at different points; B, upward movement of organ of Corti causes bending of cilia away from limbus and C, downward movement of organ of Corti causes cilia to bend towards limbus.

bending the stereocilia away from the limbus (Fig. 11.2-12B).

- When the organ of Corti moves down, the tectorial membrane slides backwards relative to basilar membrane and bends the stereocilia towards the limbus (Fig. 11.2-12C).

The bending of stereocilia stimulates (excites) the hair cells.

- *Depolarization occurs* when the stereocilia bend away from the limbus and
- *Hyper-polarization* occurs when the stereocilia bend towards the limbus.

Membrane potential changes in the hair cells

The bending of the stereocilia produces a change in the membrane potential of the hair cells proportionate to the degree of displacement (generator potential). The electrical activity of the inner ear can be considered as under:

- Resting condition and
- During stimulation of ear.

Potentials recorded from the ear under resting condition

Under resting condition (when the ear is not stimulated with sound), two different potentials are recorded:

- Endocochlear potential and
- Resting potential of the hair cells.

Endocochlear potential. The endolymph contains a high concentration of K^+ (135 mEq/L) and is electrically positive in comparison to perilymph. The +80 mV electrical potential which exists between endolymph of scala media and perilymph of scala vestibuli and scala tympani is called *endolymphatic* or *endocochlear potential*. *Source* of endolymphatic potential is stria vascularis, which covers the lateral wall of the scala media. The characteristic features of the cells of stria vascularis, which contribute to high K^+ concentration of endolymph are:

- High concentration of Na^+–K^+–ATPase and
- Presence of a unique electrogenic K^+ pump.

Resting potential of the hair cells. Each hair cell has a negative resting membrane potential. Therefore, intracellular fluid is at a potential of –70 mV with respect to the perilymph of scala tympani. At the upper end of hair cell, the potential difference between intracellular fluid and endolymph is, therefore, –150 mV [–70 – (+80)]. However, there is not much difference between K^+ concentration of endolymph and intracellular fluid. The large negative potential and lack of K^+ concentration difference between the inside and outside of the hair cells make these cells very sensitive. Therefore, slightest movement of the hair stimulates the cells.

Potentials recorded from the ear on stimulation

When the ear is stimulated by sound, two types of potentials can be recorded:

- Cochlear microphonic potential and
- Action potential in the auditory nerve.

Cochlear microphonic potential. When stimulated by the sound wave, the changes in membrane potential of the hair cells result from changes in cation conductance at their apical ends. The gating of K^+ channels is controlled by bending of stereocilia as:

- When the stereocilia bend away from the limbus, they cause K^+ channels to open; K^+ then flows into the cell and the *hair cell depolarizes*.
- When the stereocilia bend toward the limbus, they cause K^+ channels to close and the *hair cell hyperpolarizes*.

The sum of receptor potentials of a number of hair cells when recorded extracellularly is called *cochlear microphonic potential*. It is an oscillatory event that can be recorded by placing one electrode in the scala media and other electrode in the scala tympani.

Cochlear microphonic potentials are similar to generator potential, because:

- These have no latency or refractory period,
- These do not obey all or none law and
- These are resistant to ischaemia and anaesthesia.

Relationship between intensity of sound and cochlear *microphonic potential.* The cochlear microphonic potentials recorded have the same form and polarity as that of the acoustic stimulus. The excitatory phase of cochlear microphonic potential (i.e. increasing negativity in scala media) is associated with a current flow outwards across the membrane of the nerve fibres (Fig. 11.2-13). As shown in Fig. 11.2-13, the base of cochlea responds to all frequencies of sound, while the apex responds to only low frequencies of sound.

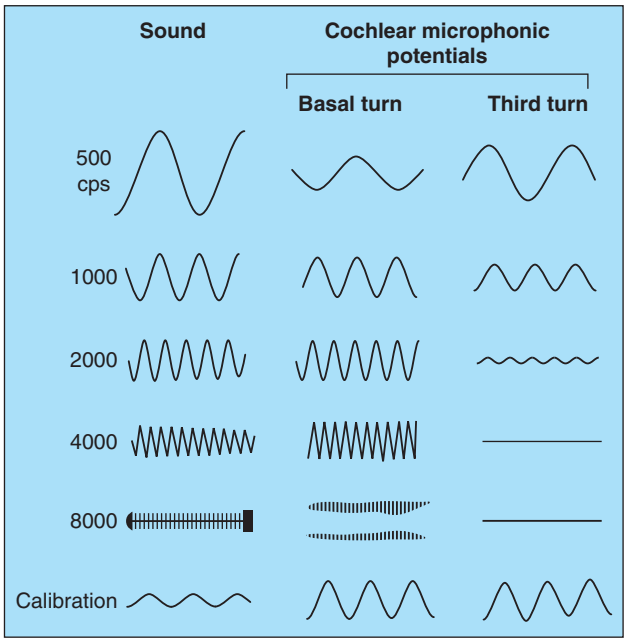


Fig. 11.2-13 Relationship between the intensity (loudness) of sound and cochlear microphonic potentials recorded through basal turn and third turn of cochlea.

Note. When organ of Corti is damaged by prolonged exposure to a loud tone, the cochlear microphonic potential produced by this particular band of frequency is abolished.

Genesis of action potential in afferent nerve fibres. The stereocilium of hair cells of organ of Corti are linked to the site of neighbouring hair cell by a very fine process called tip link. The arrangement is such that tip link tie the tip of stereocilium to the side of its higher neighbouring one stereocilium (Fig. 11.2-14). At the junction, mechanosensitive cation channels are present at the higher process. The events of genesis of action potential are as follows:

- When shorter stereocilia are pushed towards the higher neighbouring ones, the channels get open up and K^+ and Ca^{2+} influx causes depolarization.
- The molecular motors (myosin based) present in the higher neighbouring process next moves the channel towards the base and thereby releasing tension in the tip link. This causes closure of the mechanosensitive cation channel and permits restoration of the resting state.
- The depolarization of hair cells causes release of neurotransmitter (glutamate), which initiates depolarization of neighbouring afferent neurons and causes generation of action potential.
- The K^+ that enters into the hair cells through mechanosensitive channels is recycled (Fig. 11.2-14). It enters through tight junctions into the neighbouring supporting cells and reaches into the stria vascularis and secreted back into the endolymph, completing the cycle.

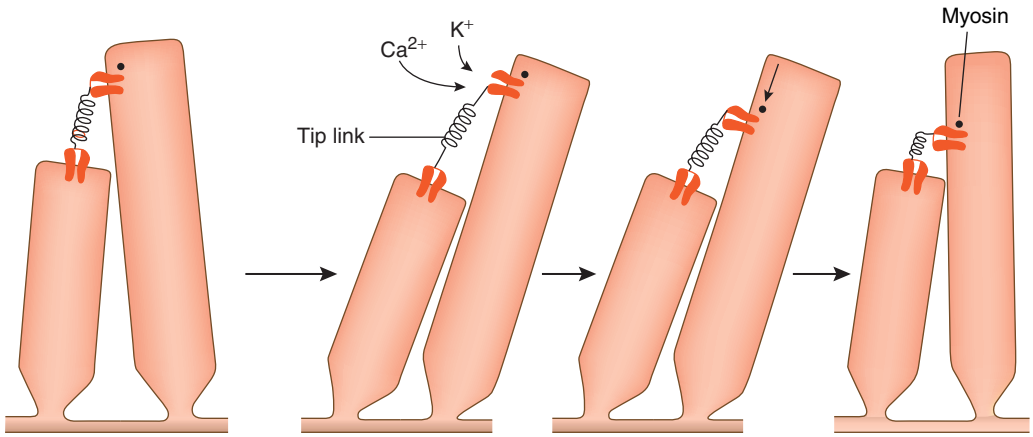


Fig. 11.2-14 Schematic representation of the role of tip links in the responses of hair cells.

Action potential of the auditory nerve. Action potentials in the auditory nerve fibre also show refractory period and obey all-and-none law. *Loudness of the sound stimuli determines the frequency of action potentials in a single auditory nerve fibre* (Fig. 11.2-15):

- *At low sound intensities*, each axon discharges to sounds of only one frequency called the characteristic frequency, which varies from axon to axon depending upon the part of cochlea from where the fibre originates.
- *At higher sound intensities* the individual axon responds to an increasingly wide range of sound frequencies.

Refractory period of auditory nerve fibre is 1 ms. Therefore, maximum rate of discharge through fibre can be only 1000 impulses/s. At a very low frequency (20–200 cycles/s) there is synchronization between the sound frequency and the rate of discharge.

NEURAL TRANSMISSION OF SIGNALS

The electrical signals which emanate from transduction of the sound waves in the hair cells are transmitted through a complex auditory pathway, which consists of following relay stations (Fig. 11.2-7):

- Spiral ganglion,
- Cochlear nuclei,
- Superior olivary nuclear complex, trapezoid nucleus and nucleus of lateral lemniscus,
- Inferior colliculus,
- Medial geniculate body and
- Auditory cortex.

Salient features of auditory pathway

Some salient features of auditory pathway which need special emphasis are:

1. Bilateral representation. From medulla onwards each ear is bilaterally represented in the auditory pathway with only slight preponderance in the contralateral pathway. Because of the bilateral representation, lesion beyond medulla has a slight effect on the auditory acuity.

2. Descending pathway. There is not only an ascending auditory pathway, but also a significant descending pathway forming feed-forward and feed-backward loops.

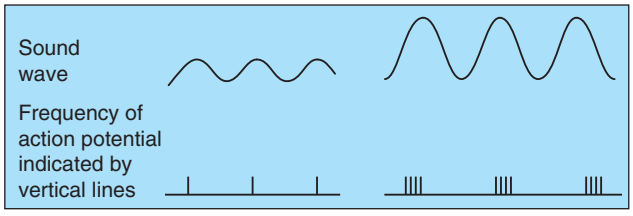


Fig. 11.2-15 Effect of intensity of sound wave on action potential recorded from cochlear nerve fibres.

3. Role in brain stem and spiral acoustic reflexes. Auditory pathway is also involved in the brain stem and spiral acoustic reflexes and brain stem mechanism for auditory visual reflexes. The integration of visual and auditory information occurs due to interconnection of the superior and inferior colliculi.

4. Role in general arousal. The auditory pathways in the brain stem give collaterals to the reticular formation and the cerebellum and thus play a role in general arousal.

5. Spatial organization. The different parts of organ of corti respond to tones of different frequencies from the basilar to the apical part of cochlea. Neurons receiving fibres from different parts of the spiral ganglion are arranged in a definite sequence in the cochlear nuclei. The tonotopic organization which is prominent in the cochlear nuclei is maintained in the superior olivary nucleus, inferior colliculus, medial geniculate body and auditory cortex. This *tonotopic organization* resembles the retinotopic organization of the visual pathway and somatotopic organization of the somatosensory system.

6. Features of auditory cortex. The auditory cortex exhibits following characteristic features:

(i) *Tonotopic organization,* as described above and
(ii) *Column organization.*

In addition to the tonotopic organization, the auditory cortex also exhibits feature extractions. For example, some neurons are selected for the direction of frequency modulation. Neurons in the primary auditory cortex form the so-called isofrequency, summation and suppression columns.

- *Isofrequency columns.* Neurons arranged in these columns have the same characteristic frequency.
- *Summation columns.* Neurons in these columns are more responsive to the binaural than to the monaural input.
- *Suppression columns.* Neurons in these columns are less responsive to the binaural than to the monaural stimulation, and accordingly the response to one ear is dominant.

7. Features of other cortical areas concerned with audition are:

(i) *Hemispheric specialization.* Brodmann's area 22 is concerned with the processing of auditory signals related to speech. During language processing, it is much more active on the left side than on the right side. Area 22 on the right side is more concerned with melody, pitch and sound intensities.

(ii) *Plasticity of auditory pathways.* There exists a great plasticity in the auditory pathways, i.e. they are modified

by experience. Examples of auditory plasticity in humans include the following observations:

- Individuals who become deaf before language skills are fully developed, viewing sign language activates auditory association areas.
- Conversely, individuals who become blind early in life are demonstrably better at localizing sound than individuals with normal eyesight.
- Musicians have an increase in the size of auditory area activated by the musical tones. They also have larger cerebellum than non-musicians, presumably because of learned precise finger movements.

NEURAL PROCESSING OF AUDITORY INFORMATION

Neural processing of auditory information involves:

- Encoding of frequency (pitch determination),
- Encoding of intensity (determination of loudness),
- Feature detection and
- Localization of sound in space.

Encoding of sound frequency

The human auditory mechanism has a remarkable power to discriminate between the sounds in the 60–20,000 Hz range. Cochlear nerve fibres encode frequency of sound stimulus. *Duplex theory*, which includes both, place theory and frequency theory, is required to explain the frequency coding of sound.

Place theory or Bekesy travelling wave theory

This theory can explain the discrimination between sound frequencies above 2000 Hz and up to 20,000 Hz.

Salient features of this theory are:

Basilar membrane is a mechanical analyser of source frequency. The basic pattern of movement of the basilar membrane is that of a travelling wave (Fig. 11.2-16). The high frequency sound waves produce waves of maximum height near the oval window, whereas low frequency sounds produce waves of maximum height near the helicotrema.

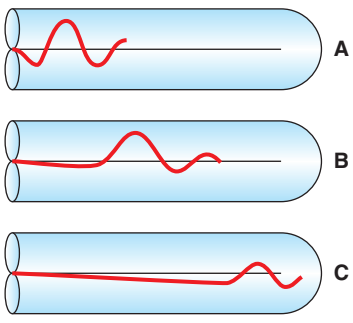


Fig. 11.2-16 Travelling wave along the basilar membrane for high A, medium B, and low C, frequency sounds.

Correspondingly, the basilar membrane near the oval window vibrates in response to high frequency sounds. As the distance of the basilar membrane from the oval window increases there is a gradual decrease in the frequency of sounds to which the membrane responds. Near the helicotrema the basilar membrane responds to a very low frequency sound (Fig. 11.2-17). This differential response to different frequencies of sound is possible because of a systematic variation in the mechanical properties and along the basilar membrane. The basilar membrane is narrowest and stiffest at the base of the cochlea (near the oval and round windows) and widest and most compliant at the apex of the cochlea (near the helicotrema).

Thus, as per the travelling wave theory of Von Bekesy, the higher frequencies are represented in the basal turn of cochlea and the progressively lower ones towards the apex. *Different hair cells respond to different frequencies of sound depending upon their location on the basilar membrane.* The auditory nerve fibre activated by a particular sound frequency is similarly dependent upon the location of hair cell it innervates. There are about 30,000 nerve fibres in the auditory nerve and each gets maximally stimulated by a particular frequency called the *characteristic frequency*. As described above, there is a spatial organization of the auditory pathways all the way from the hair cells to the auditory cortex. With each sound frequency therefore specific neurons are activated.

Frequency theory

Frequency theory or *volley principle* accounts for the coding of low frequencies of sound up to 2000 Hz. For very low frequencies of sound, there is a synchronization

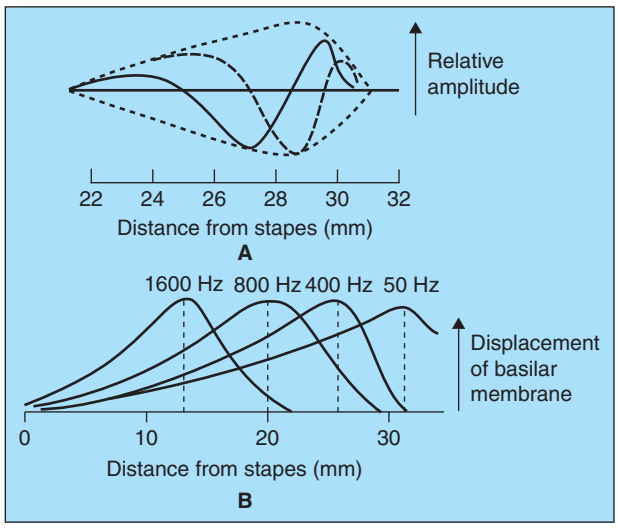


Fig. 11.2-17 Amplitude pattern of vibration of basilar membrane for a medium frequency sound wave A, and displacement of the basilar membrane by the waves generated by stapes vibrations of frequencies shown at the top of each curve, B.

between frequency of sound and rate of discharge through cochlear nerve. This is called volley principle of frequency discrimination.

The importance of the volley principle is limited. The frequency of action potentials in a given auditory nerve fibre determines principally the loudness rather than the pitch of a sound.

Other factors affecting pitch of sound. Pitch is the subjective sensation produced by the *frequency of sound*. Therefore, higher the frequency greater is the pitch. However, discrimination of pitch also depends on some other factors which are:

- *Loudness* of sound also plays a part, low tones (below 500 Hz) seem lower and high tones (above 4000 Hz) seem higher as their loudness increases.
- *Duration* of sound also affects pitch to a minor degree. The pitch of a tone cannot be perceived unless it lasts for more than 0.01 s and with durations between 0.01 and 0.1 s, pitch rises as duration increases.

Encoding of intensity

Encoding of sound intensity (loudness) occurs at the level of cochlear nerve fibres by following mechanisms:

- *Increase in frequency of firing of an auditory nerve fibre.* With the increase in intensity (loudness) of sound wave, the amplitude of vibration of the basilar membrane increases, which in turn increases the frequency of firing in an auditory nerve fibre.
- *Increase in number of nerve fibres stimulation.* As the amplitude of vibration increases, a larger portion of the basilar membrane is vibrated and thus more and more hair cells are stimulated. This increases the number of auditory nerve fibres which are activated.
- *Stimulation of inner hair cells.* Certain hair cells (inner hair cells) are not stimulated unless the sound is very loud. Stimulation of these cells, therefore, apprises the nervous system that intensity of sound is high.

Feature detection

Higher auditory centres respond to particular features of sound stimuli. For example, cortical neurons may respond specifically to a shift from high-to-low-frequency notes, which is why lesions of the auditory cortex may not impair the ability to discriminate frequency. Instead, lesions of auditory cortex cause a loss of ability to recognize a patterned sequence of sounds.

Location of sound in space

A human can distinguish sounds originating from the sources separated by as little as 1°. Binaural receptive fields (which is a feature of most auditory neurons above the level of cochlear nuclei) contribute to sound localization. In other words, relay nuclei in the brain stem (especially the superior olivary nuclei complex) mediate localization of sound sources. The auditory system uses following clues to judge the origin of sound:

- *Time lag between the entry of sound in two ears.* The detectable interaural time differences even 20 μs is more important clue, especially for relatively low-frequency sounds (below 3000 Hz).

Sound localization is markedly disrupted by lesions of the auditory cortex.

APPLIED ASPECTS

- Noise and masking
- Hearing loss and deafness
- Hearing tests.

NOISE AND MASKING

Noise

Noise is defined as an aperiodic complex sound. There are three types of noise:

White noise. It is a broadband noise which contains all frequencies in an audible spectrum. It is analogous to the white light, which contains all the colours of the visible spectrum. It is used for masking.

Narrow-band noise. It is a white noise, out of which certain frequencies above and below the given noise have been filtered out. Thus, its frequency range is smaller than the broadband white noise. It is used to mask the tested frequency in pure tone audiometry.

Speech noise. It is a noise having all frequencies in the speech range (300–3000 Hz). All other frequencies are filtered out.

Masking

Masking refers to a phenomenon in which the presence of one type of sound decreases the ability of the ear to hear another type of sound. In other words, masking represents the inability of the auditory mechanism to separate the simultaneous stimulation into separate components. Masking is more effective for sounds with similar frequencies than with sounds for widely different frequencies. Low-frequency tones mask high-frequency tones more easily than the reverse. Example of masking observed is the difficulty in conversation in noisy surroundings.

Clinical applications. In clinical audiometry, one ear is kept busy by a sound while the other is being tested. Masking of non-test ear is essential in all bone conduction tests.

HEARING LOSS, DEAFNESS AND TINNITUS

Hearing loss refers to the impairment of hearing and its severity may vary from mild to profound.

Deafness is labelled when there is little or no hearing at all.

Degree of hearing loss. WHO, 1980, recommended the following classification (Table 11.2-1) on the basis of pure tone audiogram taking the average of the thresholds of hearing for frequencies of 500, 1000, and 2000 Hz.

Types of hearing loss. Hearing loss can be of three types:
- Conductive hearing loss,
- Sensorineural hearing loss and
- Mixed hearing loss.

1. Conductive hearing loss

Any disease process which interferes with the conduction of sound from the external ear to cochlea causes conductive hearing loss.

Causes. The causes of conduction hearing loss may lie in the:

- External ear: any obstruction in the ear canal, e.g. by wax, tumours, atresia etc.
- Tympanic membrane, e.g. perforation.
- Middle ear cavity, e.g. fluid in the middle ear (as in otitis media).
- Ear ossicles, e.g. disruption of ear ossicles and fixation of ear ossicles (otosclerosis).
- Eustachian tube obstruction as in retracted tympanic membrane.

Characteristic features
- Characteristically, hearing loss is partial and never complete because skull bones themselves conduct sound to the cochlea (bone conduction) and the basilar membrane can be set into vibrations.

Table 11.2-1 WHO classification of hearing loss

Degree of hearing loss		Hearing threshold in better ear (average of 500, 1000 and 2000 Hz)
0	Not significant	0–25 dB
1	Mild	26–40 dB
2	Moderate	41–55 dB
3	Moderate severe	56–70 dB
4	Severe	71–91 dB
5	Profound	Above 91 dB
6	Total	

2. Sensorineural hearing loss

Sensorineural (SN) hearing loss results from lesions of cochlea (*sensory type*) or eighth cranial nerve and its central connections (neural type).

Causes. SN hearing loss can be congenital or acquired.

Congenital SN hearing loss is present at birth. It may be due to anomalies of the inner ear or damage to the hearing apparatus by prenatal or perinatal factors.

Acquired SN hearing loss appears later in life. Cause may be genetic (delayed onset) or non-genetic. Causes of non-genetic acquired SN deafness are:

- *Infection* of labyrinth (viral, bacterial or spirochaetal).
- *Acoustic trauma*, i.e. injury to labyrinth or eighth nerve.
- *Noise trauma or noise*-induced hearing loss occurs due to prolonged exposure to industrial noise.
- *Ototoxicity*. Certain drugs cause damage to inner ear, e.g. streptomycin, neomycin, quinine, chloroquine, etc.
- *Neoplasms*, e.g. acoustic neuroma.
- *Systemic disorders*, e.g. diabetes mellitus, hypertension etc.

Characteristic features
- Usually loss of hearing is complete.
- Speech discrimination is poor.
- Hearing loss may exceed 60 dB.

3. Mixed hearing loss

Both conductive and sensorineural hearing loss is present in the same ear.

Characterized by:

- Air-bone gap indicating conductive hearing loss and
- Impairment of bone conduction indicating sensorineural hearing loss.

Tinnitus

Tinnitus refers to ringing sensation in the ear. It is caused by irritative stimulation of either the inner ear or the vestibulocochlear nerve.

IMPORTANT NOTE

Presbycusis. The gradual hearing loss associated with aging in called presbycusis. It occurs due to the gradual loss of hair cells and neurons.

HEARING TESTS

A. Clinical tests of hearing

1. Finger friction test. It is a rough and quick method for screening. In it thumb and index finger are rubbed near the ear and patient is asked to appreciate with eyes closed.

2. Watch test. A clicking watch is brought close to the ear and the distance at which it is heard is noted.

3. Tuning fork tests. These tests are performed with tuning forks of different frequencies (commonly used are 256 and 512 Hz). These are quite useful in distinguishing conductive deafness from the sensorineural deafness. Commonly used tests are:

(i) Rinne test. In this test, air conduction (AC) of the ear is compared with bone conduction (BC). Base of a vibrating tuning fork is placed on the mastoid bone (Fig. 11.2-18A), and when he stops hearing, it is brought besides the meatus (Fig. 11.2-18B). If he still hears, AC is more than BC and Rinne test is positive.

- In normal subjects, Rinne test is positive.
- In conductive deafness, Rinne test is negative.
- In partial nerve deafness, Rinne test is positive.
- In complete nerve deafness, both bone conducted and air conducted sounds are not perceived.

(ii) Weber's test In this test, base of the vibrating tuning fork is placed in the middle of the forehead (Fig. 11.2-18C) or vertex and patient is asked in which ear the sound is heard better.

- *Normally,* the sound is heard equally in both ears.
- In *conductive hearing loss,* the sound is lateralized (better heard) towards affected ear. This is because the masking effect of environmental noise is absent in the affected ear.
- In *sensorineural hearing loss,* the sound is lateralized towards better ear because the sound is reaching the normal cochlea through bone.

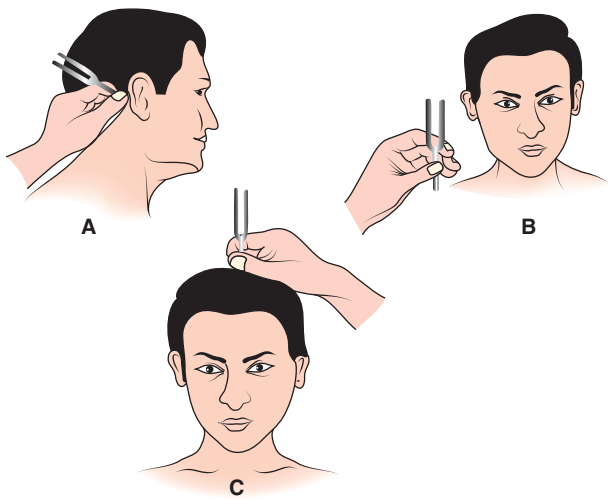


Fig. 11.2-18 Tuning fork test: A, test for bone conduction; B, test for air conduction and C, Weber's test.

(iii) Schwabach's test. In this test, patients' bone conduction is compared with that of the examiner (presuming that the examiner has normal hearing).

- *Normally,* both the examiner and the subject hear the sound equally well.
- In *conductive deafness,* the patient hears the fork for longer period than the examiner (because there is no masking effect of environmental noise).
- In *sensorineural deafness,* the examiner hears the tuning fork for a longer duration than the patient.

Table 11.2-2 summarizes the interpretation of tuning fork tests.

B. Audiometric tests

Audiometer is the device used to perform audiometry. Audiometry refers to the measurement of auditory acuity (sharpness of hearing) using the audiometer. An audiometer consists of following main parts:

- Electronic oscillator. It can generate pure tones of frequencies ranging from low to high.
- Intensity dial. It helps to adjust the threshold intensity of hearing for each tone.
- Headphone. It helps to deliver the pure tones of various frequencies to each ear separately.

Pure tone audiometry. It is performed in a sound-proof room. Each ear is tested separately. Usually AC thresholds are measured for tones of 125, 250, 500, 1000, 2000, 4000 and 8000 Hz and BC thresholds for 250, 500, 1000, 2000 and 4000 Hz. The results are interpreted as:

- Audiometer is so calibrated that the hearing of a normal person, both for air and bone conduction is at 0 dB and there is no A–B gap, while tuning fork tests normally show AC > BC.

Table 11.2-2 Tuning fork tests and their interpretation

Test	Normal	Conductive hearing loss	Sensorineural hearing loss
Rinne test	AC > BC Rinne +ve	BC > AC Rinne –ve	AC > BC in partial deafness
Weber's test	Not lateralized	Lateralized towards affected ear	Lateralized towards healthy ear
Schwabach test	Equal	Better conduction in patient	Better conduction in examiner

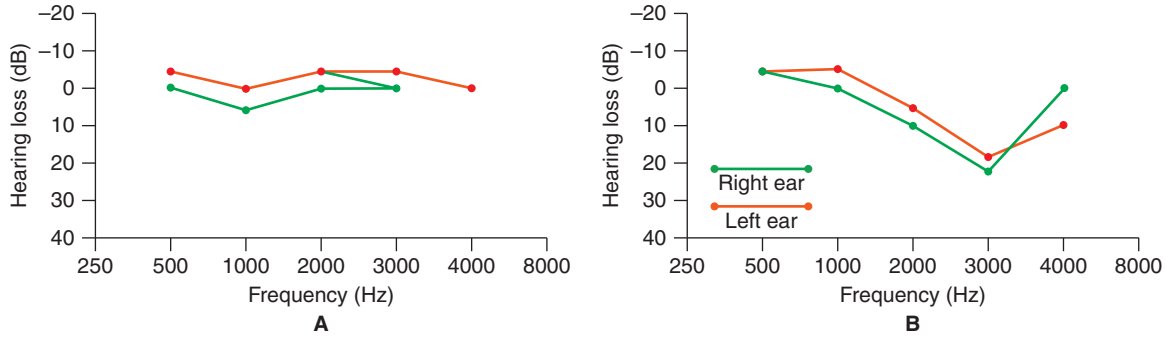


Fig. 11.2-19 Audiogram showing: A, normal hearing; B, loss of 20 dB hearing for 3000 Hz frequency in both ears.

- The amount of intensity that has to be raised above the normal level is a measure of the degree of hearing loss at that frequency. It is charted in the form of a graph called *audiogram* (Fig. 11.2-19).
- Threshold of BC is a measure of cochlear function.
- Difference in the thresholds of air and bone conduction (A–B gap) is a measure of degree of conductive deafness.

C. Special tests for hearing

Certain special tests have been devised to elucidate different aspects of hearing loss. These tests include:

- Evoked response audiometry which includes:
  - Electrocochleography and
  - Auditory brainstem responses (see page 859).

Chapter 11.3

Chemical Senses: Smell and Taste

SENSE OF SMELL

- Site of olfaction
  - Olfactory mucosa
  - Vomeronasal organ
- Olfactory pathways
  - Olfactory nerves
  - Olfactory bulb
  - Olfactory tracts
  - Olfactory cortex
- Physiology of olfaction
  - Odoriferous stimuli
  - Olfactory receptors
  - Steps in transduction in the olfactory receptor neurons
  - Processing of olfactory sensations in the olfactory bulb
  - Transmission of odorant information to the olfactory cortex and neocortex
  - Factors affecting olfaction
- Applied aspects
  - Abnormalities of olfaction
  - Measurement of sense of smell

SENSE OF TASTE

- Site of taste
  - Papillae
  - Taste buds
- Taste pathways
  - First-order neurons
  - Second-order neurons
  - Third-order neurons
- Physiology of taste
  - Gustatory stimuli
  - Transduction of gustatory stimuli
  - Transmission of information about taste to the cortex
  - Encoding of taste information
  - Taste thresholds and intensity discrimination
  - Sensation of flavours
  - Phenomenon of variation and after effects in taste sensations
  - Factors influencing taste sensation
  - Abnormalities of taste sensations

THE SENSE OF SMELL

The sense of smell or olfaction is well developed in animals like dog and rabbit to give warning of the environmental dangers. Such animals are called *macrosmatics*. In humans, apes and monkeys (primates), the sense of smell is comparatively less developed, but still it is important for pleasure and for enjoying the taste of food. Therefore, the humans and primates are called *microsmatics*.

SITE OF OLFACTION

The olfactory stimuli are detected by the specialized receptors located on the free nerve endings of the olfactory nerves, which are located in the:

- Olfactory mucosa of nose in human beings and
- Vomeronasal organ in reptiles and certain mammals.

Olfactory mucosa

In humans, the olfactory mucosa is confined to upper one-third of nasal cavity. It includes the roof of nasal cavity and the adjoining areas on the medial wall (septum) and superior nasal concha on the lateral wall (Fig. 11.3-1). The olfactory

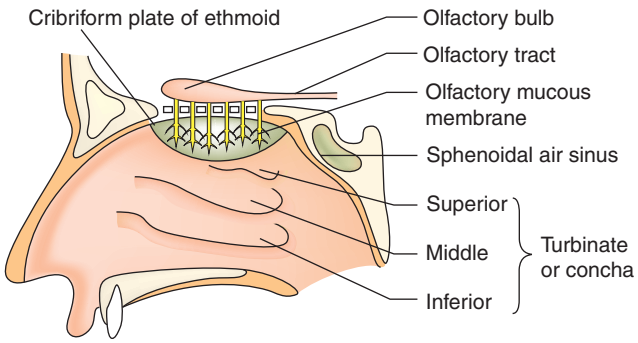


Fig. 11.3-1 Location of olfactory mucosa.

neuroepithelium is a patch of thin and dull yellow mucosa about 5.0 cm^2 in area. A mucous layer covers the entire epithelium.

Histological structure

Histologically, the olfactory mucosa consists of three types of cells (Fig. 11.3-2):

1. Receptor cells. About 10–20 million receptor cells are present in the olfactory mucosa. These cells are bipolar neurons, which lie between the supporting (sustentacular) cells. The dendrites of the receptor cells terminate in a rod from which 10–20 fine cilia project and form a dense mat into the mucous layer of the olfactory mucosa. The cilia are about 2 µm in length and 0.1 µm in diameter. Their axons are fine unmyelinated fibres, which form the olfactory nerves.

Characteristic features of olfactory receptor cells, which differentiate it from other sensory neurons, are:

- These are the only sensory neurons whose cell bodies are closest to the external environment.
- These cells have a short life span of about 60 days and get replaced by the proliferation of basal cells. This natural turnover is a unique feature of these sensory neurons.

Note. Bone morphogenic protein (BMP) inhibits the renewal turnover. BMP is a growth factor that promotes bone growth but also acts on other tissues.

2. Supporting cells, also known as sustentacular cells, are columnar in shape. Microvilli extend from the surface of these cells into the mucous layer covering the olfactory mucosa. These cells secrete mucus.

The *Bowman's glands* lying just under the basement membrane also secrete mucus.

3. Basal cells are stem cells from which new receptor cells are formed. As mentioned above, there is a continuous replacement of receptor cells by mitosis of the basal cells.

Distinguishing features of olfactory mucosa from the surrounding respiratory mucosa of nasal cavity are:

- Presence of receptor cells,
- Presence of Bowman's glands,
- Absence of rhythmic ciliary beating (which is a characteristic feature of respiratory mucosa) and
- Presence of a distinctive yellow-brown pigment.

Nerve supply of olfactory mucosa

- *Special sensory nerves* innervating the olfactory mucosa are 15–20 bundles of olfactory nerve fibres (first cranial nerve) which convey sense of smell.
- *General sensory nerves* supplying the olfactory mucosa are branches of trigeminal nerve (fifth cranial nerve). The irritative character of some odorants results from stimulation of free nerve endings of the trigeminal nerve.

IMPORTANT NOTE

Vapours of ammonia are never used to test the sense of smell as they stimulate the fibres of trigeminal nerve and cause irritation in the nose rather than stimulating the olfactory receptors.

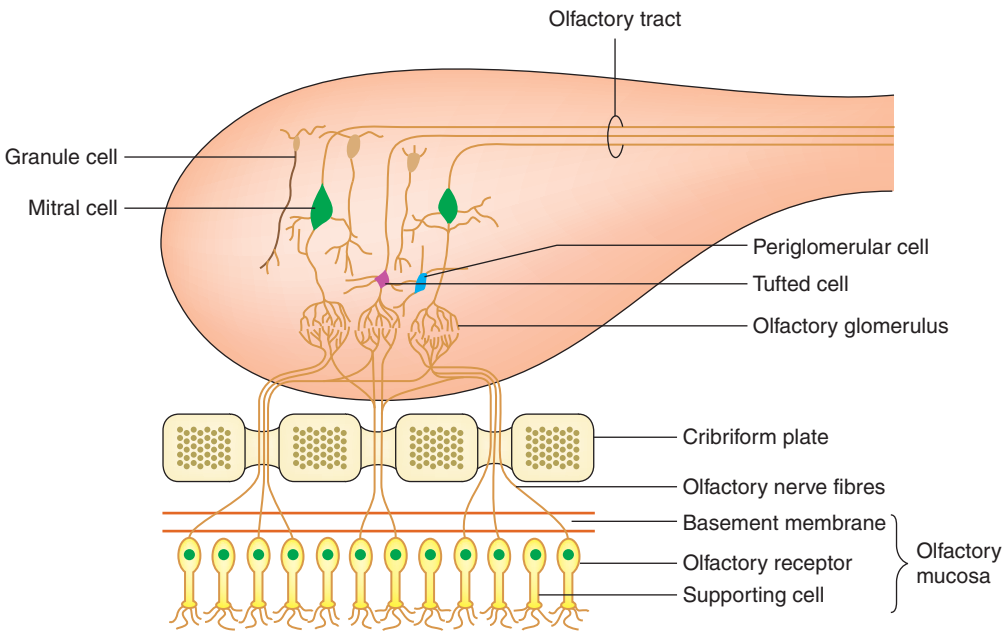


Fig. 11.3-2 Histological features of olfactory mucosa and diagrammatic depiction of the synapses between the axons of olfactory neurons (first-order neurons) with the dendrites of mitral and tufted cells (second-order neurons) to form the olfactory glomeruli, which lie in the olfactory bulb. The inhibitory neurons present in the olfactory bulb are granule cells and periglomerular cells.

Vomeronasal organ

It is a pouch-like structure found along the nasal septum in the nose of some animals (rodents and various other animals). Receptors present in the vomeronasal organ are concerned with the perception of odour that emanates from the *pheromones* and foodstuffs and are thus related to food and sex behaviour of the animals. Nerve fibres emerging from the vomeronasal organ project to the *accessory olfactory bulb* and from those primarily to areas in the amygdala and hypothalamus that are concerned with reproduction and eating behaviour.

Vomeronasal organ is not well developed in humans, but there is anatomically separate and biochemically unique area of olfactory mucous membrane in a pit in the anterior third of the nasal septum, which appears to be homologous structure.

Pheromones. These are hormone-like substances, which emit specific odour and produce hormonal, behavioural or other physiological changes in another animal of the same species. Usually a pheromone is secreted by an animal during mating season only. The smell of pheromones often is the cause of sex, which an animal follows to find out its mating partner which may be waiting at a distance. It is being assumed that pheromone also exists in humans and that there is close relationship between smell and sexual function.

OLFACTORY PATHWAYS

Olfactory pathways comprise:

1. Olfactory nerves. About 15–20 olfactory nerve filaments which consist of the axons of the bipolar olfactory neurons, which pierce the cribriform plate on either side to reach the olfactory bulb.

2. Olfactory bulb (Fig. 11.3-2). It is an oval flattened strip of grey matter lying on the cribriform plate, which receives the olfactory nerve filaments. There is point-to-point representation of olfactory mucosa in the olfactory bulb. The upper part of the mucosa is represented in the anterior part of bulb while the lower part is represented posteriorly. The olfactory bulb contains three types of cells: mitral cells, tufted cells and interneurons (granule cells and periglomerular cells). The mitral and tufted cells constitute second-order neurons.

- *Dendrites of mitral and tufted cells* branch and form synapses with the axon terminals of olfactory neurons (first-order neurons) to constitute globular masses called olfactory glomeruli. Olfactory axons converge extensively onto the mitral cell dendrites, as many as axons from 1000 olfactory neurons synapse on the dendrites of a single mitral cell.
- *Granule and periglomerular cells* are inhibitory neurons. They form dendro-dendritic reciprocal synapses with the dendrites of the mitral cells. The periglomerular cells also participate in the formation of olfactory glomeruli.
- *Axons of the mitral and tufted cells* leave the olfactory bulb and run in the olfactory tract.

3. Olfactory tract. It lies in the olfactory sulcus on the orbital surface of the frontal lobe and proceeds backwards from each olfactory bulb to the region of anterior perforated substance on the base of brain where it divides into lateral, intermediate and medial olfactory striae (Fig. 11.3-3). *Olfactory trigone* refers to the flattened part of the olfactory tract, near the anterior perforated substance before it divides into the striae.

Anterior olfactory nucleus. It is made up of scattered neurons within the olfactory tract. Neurons in this structure receive synaptic connections from neurons of the olfactory bulb and send axons through the anterior commissure to excite inhibitory neurons on the contralateral olfactory bulb (Fig. 11.3-4).

Olfactory striae. Three striae are derived from each olfactory tract (Fig. 11.3-4):

- *Lateral olfactory stria*. Axons of the lateral olfactory stria synapse in the primary olfactory receiving area, which includes the *prepiriform cortex* (and in many animals the piriform lobe).
- *Medial olfactory stria* includes projections to the *amygdaloid nucleus*, as well as to the part of the cortex of the basal forebrain.
- *Intermediate olfactory stria* terminates in the *olfactory tubercle*, an area of the cortex rostral to the anterior perforated substance.

4. Olfactory cortex. It includes the anterior olfactory nucleus, prepiriform cortex, olfactory tubercle and amygdala. All these are the parts of limbic system.

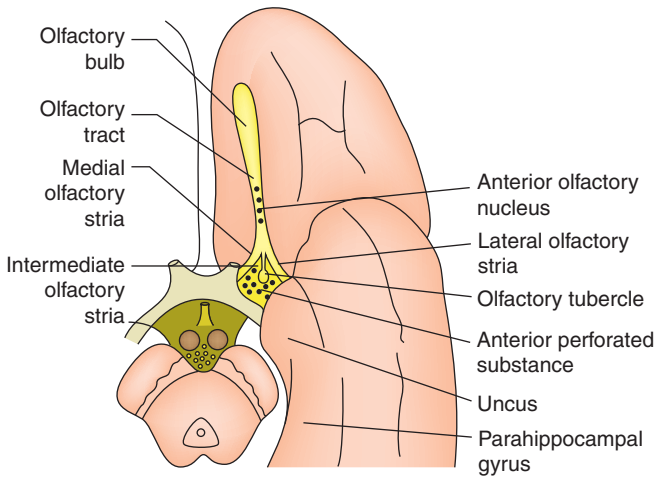


Fig. 11.3-3 Olfactory bulb and olfactory tract.

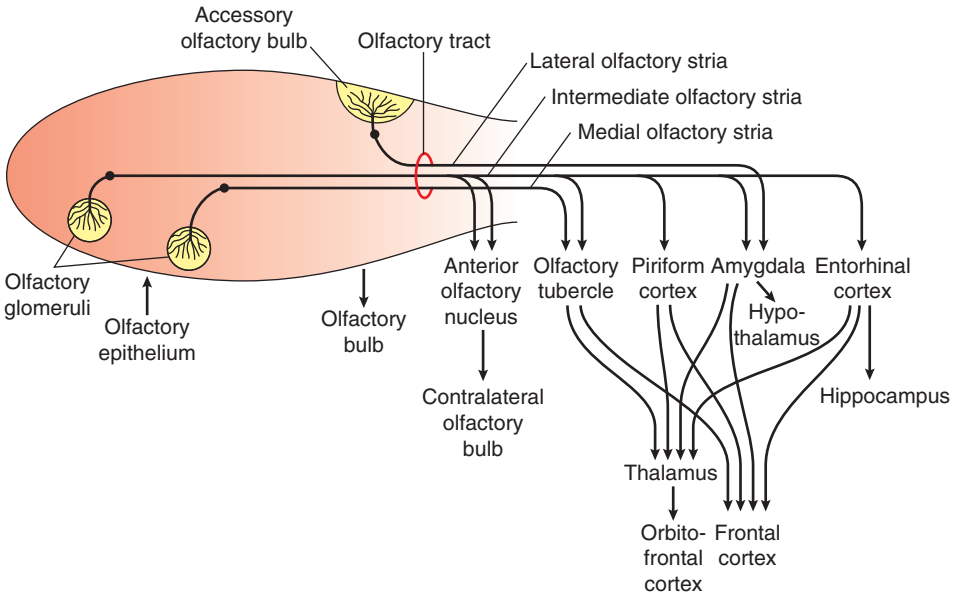


Fig. 11.3-4 Olfactory pathways.

PHYSIOLOGY OF OLFACTION

Odoriferous stimuli

The odoriferous (smell producing molecules) stimuli enter the nasal cavity while breathing. During quiet breathing, the air passes through the lower parts of the nasal cavity. Through eddy currents, however, some air does reach the olfactory epithelium. The amount of air reaching the olfactory mucosa can be increased by sniffing, which causes turbulence in the airflow in nasal cavity. Sniffing is an act of deep breathing (semi-reflex response), which occurs when a new odour is encountered. The odorant molecules must dissolve in the mucous layer (lining the olfactory mucosa) before they can come in contact with the olfactory receptors.

Characteristic features of odorant molecules. To be effective an odorant molecule must be:

- *Volatile*, because the olfactory receptors respond to chemicals transported by the air into the nose.
- *Water soluble* (to some extent) to penetrate the watery mucous layer (lining the nasal epithelium) to reach receptor cell membrane.
- *Lipid soluble* (to some degree) to penetrate the cell membranes of the olfactory receptor cells to stimulate those cells.

Types of odorant stimuli. There seems to be over 50 *primary smell sensations* (in contrast to three primary sensations of colour and four primary sensations of taste). Although the olfactory capability of humans is somewhat limited, compared with that of macrosmatic animals; nevertheless, humans are able to perceive more than 10,000 different odorous molecules.

Common odours encountered are named as:

- *Aromatic or resinous odours*, e.g. camphor, lavender and cloves.
- *Fragrant odours*, e.g. perfumes and flowers.
- *Ethereal odours*, e.g. ether, chloroform.
- *Garlic odours*, e.g. garlic, onion and sulphur compounds.
- *Burning odours*, e.g. tobacco, burning of feathers, meat and bones.
- *Nauseating odours*, e.g. excreta, decomposed meat and vegetables.
- *Goat odours*, e.g. sweat, ripe cheese.
- *Repulsive odours*, e.g. odour of the bed bug.
- *Musky odours*, e.g. musk.

Olfactory receptors

The cilia of the olfactory neurons are specialized for odour detection. They have specific receptors for odorants as well as the transduction machinery needed to amplify sensory signals and generate action potentials in the neuron's axon. Some important points about olfactory receptors are:

- A large family of odorant receptors permits discrimination of a wide variety of odorants.
- A large multigene family appears to code for as many as 1000 different types of odorant receptors.
- The odorant receptors belong to a large superfamily of structurally related receptor proteins that transduce signals by interactions with G-proteins.

Steps in transduction in the olfactory receptor neurons

1. Binding of odorant molecule to receptors. As mentioned earlier, the odorant molecules entering the nasal cavity

dissolved in the mucous layer covering the olfactory mucosa. The cilia of olfactory neurons are projected into this mucous layer.

Role of odorant binding proteins. It has been suggested that the mucous layer covering the olfactory mucosa contains one or more odorant binding protein that concentrate and transfer the odorant molecules to the receptors present on the cilia of olfactory neurons.

2. Activation of receptor. The interaction of an odorant with its receptor induces an interaction between the receptor and a heterotrimetric G-protein. This interaction causes release of the G-proteins GTP-coupled α subunit, which then stimulates adenylyl cyclase to produce cAMP.

3. Depolarization receptor potential. The increased intracellular cAMP opens cyclic nucleotide-gated (CNG) cation channels, leading to cation (Na^+ and Ca^{2+}) influx and a change in membrane potential in the cilium membrane, i.e. produces a depolarizing receptor potential.

4. Action potentials. The receptor potential depolarizes the initial segment of the axon to threshold leading to the generation of action potentials in the sensory axon and the *transmission of signal to olfactory bulb*.

Note. A specific olfactory receptor does not respond to a particular compound or category of compounds. Instead, an individual receptor responds to many odours. Furthermore, no two receptor cells have identical responses to a series of stimuli. Sensory perception, therefore, is based on the pattern of receptors activated by the stimulus.

Processing of olfactory sensation in the olfactory bulb

Odorant information is encoded spatially in the olfactory bulb. In the olfactory glomeruli, there is lateral inhibition mediated by periglomerular cells and granule cells (Fig. 11.3-2). Another potential source of signal refinement, or adjustment, is the multiple inputs to the olfactory bulb from the olfactory areas of the cortex as well as the basal forebrain and mid brain. Thus, sensory information is extensively processed, and perhaps refined in the olfactory bulb before it is sent to the olfactory cortex.

Transmission of odorant information to the olfactory cortex and neocortex

- From the olfactory bulb, the odorant information is first transmitted to the *olfactory cortex,* which includes piriform cortex, parts of amygdala, the olfactory tubercle and parts of entorhinal cortex.
- From the olfactory cortex, information is relayed to the frontal cortex (directly) and orbitofrontal cortex (via thalamus) (Fig. 11.3-4).

Note. The olfactory pathway is the only sensory system that does not have an obligatory synaptic relay in the thalamus. The olfactory tracts project directly to the olfactory cortex, while all other sensations are first processed in the thalamus before projection to the cerebral cortex.

Role played by different regions of cerebral cortex involved in processing of olfactory information is summarized as:

- *Piriform cortex* is activated by sniffing in humans.
- *Amygdala* and hypothalamus are probably involved with the emotional and motivational responses to olfactory stimuli as well as many of the behavioural and physiological effects of odour. In animals, effects of pheromones are also thought to be mediated by signals from the main and accessory olfactory bulbs to the amygdala and hypothalamus.
- *Entorhinal cortex* is concerned with olfactory memories.
- *Neocortex (orbitofrontal and frontal cortex)* is thought to be concerned with conscious discrimination of odours. People with lesions of orbitofrontal cortex are unable to discriminate odours.

Factors influencing olfactory function

1. Threshold of olfactory receptors. The threshold of olfactory receptors varies from substance to substance. For example, methyl mercaptan, a substance which gives garlic its characteristic odour, has extremely low threshold. It can be smelled at a concentration of less than 500 pg/L of air (Table 11.3-1).

2. Intensity/concentration of the odour. The concentration of an odoriferous substance must be changed by about 30% before a difference can be detected.

3. Adaptation. Olfactory sensation adapts very rapidly with continued exposure to an odour. When one is continuously exposed to even the most disagreeable odour, perception of the odour decreases and eventually ceases. However, a brief exposure to fresh air allows one to smell the unpleasant odour again.

Table 11.3-1 Threshold concentration of some odoriferous substances

Substance	Concentration (mg/L) of air
Ethyl ether	5.83
Chloroform	3.30
Pyridine	0.03
Oil of peppermint	0.02
Iodoform	0.02
Butyric acid	0.009
Propyl mercaptan	0.006
Artificial musk	0.00004
Methyl mercaptan	0.0000004

APPLIED ASPECTS

Abnormalities of olfaction

1. Anosmia and hyposmia. Anosmia is total loss of sense of smell while hyposmia refers to a diminished olfactory sensitivity.

Causes of anosmia or hyposmia are:

- *Injuries* to olfactory nerves or olfactory bulb in fractures of anterior cranial fossa.
- *Intracranial lesions* like abscess, tumour or meningitis, which may cause pressure on the olfactory tracts.
- *Nasal obstruction* due to nasal polyp, enlarged turbinates or marked oedema of nasal mucosa in allergic or vasomotor stimuli.
- *Atrophic rhinitis*, a degenerative disease of nasal mucosa, also causes anosmia.
- *Old age*. Olfactory thresholds increase with advancing age and more than 25% of humans over the age of 80 have an impaired ability to identify smells.
- *Kallmann's syndrome*. In this condition, anosmia is associated with hypogonadism.
- *Absence or disrupted function of receptors* is responsible for several dozen of different types of anosmias seen in humans.

2. Parosmia or dysosmia. It refers to a distortion or perversion of smell. In it, person interprets the odours incorrectly. Often these persons complain of disgusting odours.

Measurement of sense of smell

1. Qualitative testing. Qualitatively, sense of smell can be tested by asking the patient to smell common odours, such as onion, peppermint, rose, garlic or cloves from each side of the nose separately, with eyes closed.

2. Olfactometry is the method of quantitative estimation of sense of smell with the help of an instrument called olfactometer.

SENSE OF TASTE

Sense of taste (gustation) is a chemical sense that is stimulated by food and drink. It contributes considerably to the quality of life and is important stimulant for digestion. Taste must be distinguished from flavour, which includes the olfactory, tactile and thermal attributes of food in addition to taste.

SITE OF TASTE

The taste (gustatory) stimuli are detected by specialized chemoreceptors called taste receptors or taste cells. The taste receptors are clustered in the taste buds located on the tongue, palate, pharynx, epiglottis and upper third of oesophagus.

Tongue, the main site of taste detection, contains numerous taste buds on its dorsal surface. The mucous membrane of the dorsal surface of tongue exhibits numerous papillae, which increase the surface area of the mucosa available for taste receptors. The taste buds are located in the walls of these papillae.

Papillae

Papillae, present on the tongue, are of four types (Fig. 11.3-5):

1. Circumvallate papillae. These are large (2–4 mm in diameter) papillae, about 10–12 in numbers, forming a single row in front of the sulcus terminalis. Sulcus terminalis is V-shaped groove (with apex posteriorly), which separates the anterior two-thirds of the dorsum of tongue from the posterior one-third. Each circumvallate papilla is surrounded by a groove. About 200 taste buds are located along the sides of each circumvallate papilla.

2. Fungiform papillae. These are bright red, flat dot-like structures (each of about 1 mm in diameter) located in the anterior two-thirds of tongue along the edges, dorsum and tip. There are 8–10 taste buds on each papilla.

3. Foliate papillae. These are transverse mucosal folds, found on the posterolateral surfaces of the tongue anterior to the circumvallate papillae. Each foliate papilla has numerous taste buds.

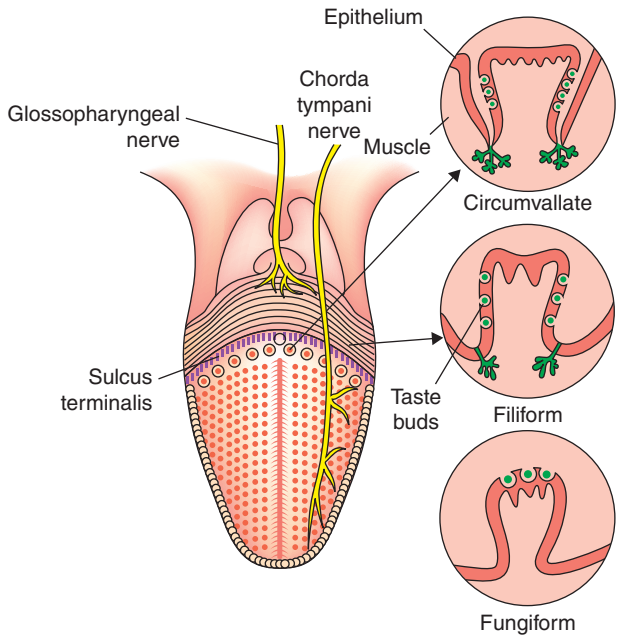


Fig. 11.3-5 Structure and distribution of papillae on the tongue and arrangement of taste buds in the three types of papillae. Innervation by the cranial nerves is also indicated.

4. Filiform papillae. These are small conical projections, covering the entire remaining surface of the dorsum of the anterior two-thirds of tongue, giving it a velvety appearance. They are arranged in rows parallel to the sulcus terminalis. They are not gustatory structures, i.e. *do not contain taste buds*. However, they may play a role in breaking up food particles and also called *mechanical papillae* in contrast to the other three forms, which are called gustatory papillae.

Taste buds

Structure. Each taste bud is barrel shaped. Cluster of cells with a small opening (taste pore) in the surface that allows substances to reach the interior of the taste bud. Each taste bud measures about 50–70 µm in diameter, and consists of following cells (Fig. 11.3-6):

1. Receptor cells. Each taste bud has about 100 receptor cells (modified epithelial cells) which have following characteristics:

- The receptor cells are elongate, bipolar shaped and extend from the epithelial opening of the taste bud to its base.
- The taste cells have a short life (about 10 days) and are continuously replaced by new taste cells differentiating from the basal cells.
- Through the taste pore, microvilli (cilia) of all the taste cells protrude into the oral cavity and come in contact with the saliva.
- The taste cells are innervated by sensory neurons (primary gustatory afferent fibres) at its basal pole. Although taste cells are non-neuronal epithelial cells, the contacts between these cells and sensory cells have morphological characteristics of chemical synapses. Each taste nerve fibre innervates taste cells in several taste buds and conversely, each taste bud is innervated by approximately 50 nerve fibres.

2. Basal replacement cells. These are small round cells present at the bottom of taste bud (Fig. 11.3-6). They are thought to be stem cells, which are continuously being differentiated into taste cells.

3. Supporting cells. In addition to the taste cells and basal cells, the taste buds contain supporting or sustentacular cells.

Innervation. The special sensory nerve fibres innervating the taste cells come from the branches of the facial, glossopharyngeal and vagus nerve (seventh, ninth and tenth cranial nerve, respectively). Each taste bud is innervated by approximately 50 nerve fibres and each nerve fibre in turn receives inputs from five taste buds. Further details of taste nerve fibres are described in the taste pathways. The tactile and temperature receptors of the mouth, tongue and pharynx are innervated by the trigeminal nerve (fifth cranial nerve).

TASTE PATHWAYS

The taste pathways consist of three orders of neurons (Fig. 11.3-7):

First-order neurons. The cell bodies of the first-order neurons innervating the taste cells in taste buds are located in different ganglia of the seventh, ninth and tenth cranial nerves as:

- *From the taste buds located on anterior two-thirds of tongue,* the taste fibres run in lingual nerve which branches from the chorda tympani nerve, which is a branch of facial nerve. The cell bodies are located in the geniculate ganglion.
- *From the taste buds located on the posterior one-third of tongue*, the taste fibres run in glossopharyngeal nerve. The cell bodies lie in the *superior and inferior ganglia* of this nerve.
- *From the taste buds located on pharyngeal aspect of tongue, epiglottis, hard and soft palate*, the taste fibres

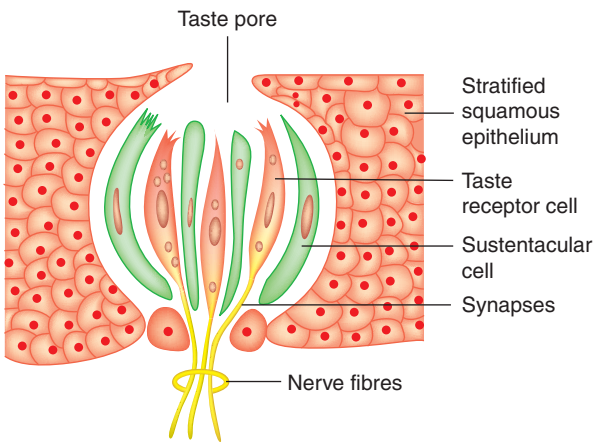


Fig. 11.3-6 Structure of a taste bud.

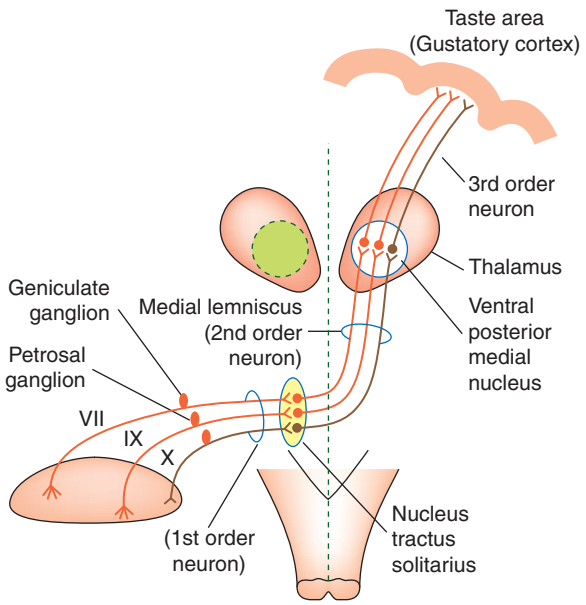


Fig. 11.3-7 The taste pathways.

run in the vagus nerve. The cell bodies are located in the superior and inferior ganglia of the vagus nerve.

- *Termination of first-order neurons*. Ultimately, all the taste fibres, travelling in different cranial nerves join the tractus solitarius to terminate in the nucleus of tractus solitarius (Fig. 11.3-7).

Second-order neurons. The cell bodies of second-order neurons of taste pathways are located in the *nucleus of tractus solitarius* in the medulla. Axons of the second-order neurons cross the midline to join the medial lemniscus and terminate with fifth cranial nerve fibres (carrying pain, touch and temperature fibres) in the ventral posterior medial nucleus of thalamus.

Third-order neurons. The cell bodies of third-order neurons are located in the *ventral posterior medial nucleus of thalamus*. Axons of third-order neurons proceed to terminate in the inferior part of the post-central gyrus, i.e. the part of sensory cortex called taste cortex (Fig. 11.3-7).

PHYSIOLOGY OF TASTE

Gustatory stimuli

Types of stimuli and most sensitive areas of tongue

There are about 10,000 taste buds, which after the age of 45 years start decreasing in number, resulting in blunting of taste sensations in old age.

Conventionally, four basic types of taste sensations have been described: *sweet*, *salt*, *sour* and *bitter*. Recently, a fifth stimulus type called *umami* has also been considered in the list of basic tastes. All other taste sensations (hundreds in humans) are assumed to result from various combinations of these five primary (basic) taste sensations. In addition to the above, associated sensations of olfaction, heat, cold and texture contribute for different flavours. Earlier, it was believed that there are special areas on the surface of tongue for each of the four conventional basic types of tastes, i.e. the sweet tastes are detected best at the tip of tongue, salty and sour tastes originate from the sides and bitter tastes are sensed best at the base. However, it is now clear that all tastes are sensed from all parts of the tongue and adjacent structures containing taste buds.

Substances producing primary (basic) taste sensations

Primary (basic) taste sensations are produced by following (rapid taste producing) substances:

1. Sweet sensation is produced by a number of organic molecules including sugars, glycols, alcohols, aldehydes, esters, etc. Saccharin is a chemical 600 times as sweet as sucrose. Being non-calorigenic, it is often used as a sweetening agent for diabetic patients.

2. Salty sensation is produced by the anions of ionizable salts especially the sodium chloride.

3. Sour sensation. It is produced by acids and the intensity of this sensation relates, to some degree, to the pH of stimulus solutions.

4. Bitter sensation is produced by alkaloids, such as quinine, caffeine, nicotine and strychnine. Many alkaloids are harmful when swallowed. Perhaps, the highly bitter taste has been given by the nature to these substances to prevent their ingestion by humans and animals.

5. Umami sensation. It has been recently added to the four basic taste sensations. It is produced by glutamate particularly by monosodium glutamate used extensively in Asian cooking. This taste is pleasant and sweet but differs from the standard sweet taste.

Transduction of gustatory stimuli

Transduction of gustatory stimuli into electrical signals is initiated at the level of receptors. The taste receptors are chemoreceptors, which are stimulated by the substances dissolved in the mouth by saliva. The dissolved substances act on the microvilli of taste receptors exposed in the taste pore of taste buds. This interaction typically depolarizes the cell either directly or via the action of second messengers. This causes the development of *receptor potential* in the receptor cell, which in turn generates action potential in the sensory nerves.

The mechanisms involved in the transduction of five types of basic taste stimuli into electrical signals are different.

Transmission of information about taste to the cortex

The detection of tastants is transduced into a receptor potential that induces action potentials in the taste cell and the release of neurotransmitter at synapses formed between the taste cell and sensory fibres. Each sensory fibre contacts a number of taste cells and each taste cell synapses with numerous sensory fibres. Thus the electrical activity recorded from a single sensory fibre represents the input of many taste cells. As described in the taste pathway (page 947, Fig. 11.3-7), the signals carried by sensory fibres that innervate the taste buds travel through several different nerves to the gustatory area of the nucleus of the solitary tract, which relay information to the thalamus. The thalamus transmits taste information to the gustatory cortex.

Encoding of taste information

As described above, each sensory fibre carries information derived from a variety of taste stimuli. However, each fibre responds best to one of the five primary taste qualities. Thus, the encoding of a gustatory sensation is not a simple,

labelled-line, chemical sensory system, instead, the identity of a taste stimulus appears to be encoded by a unique pattern of inputs from many separate fibres that provide components of the patterns for different stimuli. In this respect, the processing of taste information involves a comparison of the activity of different cells that respond preferentially, but not exclusively, to certain features of sensory stimuli.

Taste thresholds and intensity discriminations

Taste threshold. To be recognized as salty a substance need only be 0.01 M, whereas for quinine to be perceived as bitter, its concentration need only be 0.000008 M. This correlates with the notion that bitter serves a protective function against dangerous alkaloids, thus its intensity is high. The threshold concentration of some substances to which the taste buds respond is shown in Table 11.3-2.

Intensity discrimination. The ability of humans to discriminate differences in the intensity of tastes, like intensity discrimination in olfaction, is crude. A 50% change in the concentration of the substances being tasted is necessary before an intensity difference can be detected. Women are more sensitive to sweet and salt and less sensitive to sour.

Sensation of flavours

The multitude of different sensation of flavours that one experiences results from a combination of gustatory, olfactory and somatosensory inputs.

Gustatory inputs. The almost infinite varieties of tastes are synthesized from the five basic taste components described above.

Olfactory inputs are responsible for much of what we think as the flavour of foods. Volatile molecules released from foods or beverages in the mouth are pumped into the back of nasal cavity (retronasally) by the tongue, cheek and throat movements that accompany chewing and swallowing. Although the olfactory epithelium of the nose clearly makes a major contribution to sensations of taste, we experience taste as being in the mouth, not in the nose. It is thought that the somatosensory system is involved in this localization and that the coincidence between somatosensory stimulation of the tongue and retronasal passage of odorants into the nose causes the odorants to be perceived as flavours in the mouth.

Somatosensory input frequently contributes to the sensation of flavour. This component includes the texture (consistency) and temperature of foods as well as pain sensations evoked by spicy and minty foods and by carbonation.

Table 11.3-2 Threshold concentration of some taste-producing substances

Substance	Taste	Threshold concentration (μmol/L)
Hydrochloric acid	Sour	100
Sodium chloride	Salt	2000
Strychnine hydrochloride	Bitter	1.6
Quinine	Bitter	8
Glucose	Sweet	80,000
Sucrose	Sweet	10,000
Saccharine	Sweet	23

Phenomenon of variation and after effects in taste sensations

It has been reported that the taste sensations exhibit after reactions and contrast phenomena. These are similar in some way to visual afterimages and contrasts. Some of these occur due to chemical tricks, while others are considered to be the result of a true *central phenomenon*.

Factors influencing taste sensation

1. Area of stimulation. The perception of sense of taste is directly proportional to the area of taste buds stimulated. Therefore, stimulation of a small area of the tongue by one drop of solution produces weaker sensation than the same solution by the whole mouth.

2. Temperature of the tastant. An optimal response to taste-producing substances is obtained when their temperature is between 30 and 40°C.

3. Age of the person. After the age of 45 years, the number of taste buds starts decreasing resulting in blunting of sensation of taste.

4. Sex. In general, women are more sensitive to sweet and salt and less sensitive to sour.

5. Adaptation. Taste sensation adapts rapidly when taste-producing substance is kept for a long time in one place in the mouth. The adaptation is peripheral. Further adaptation to one acid produces adaptation to other acids, because H^+ is the stimulus in all cases.

6. Interaction between taste-producing substances also affects taste sensation. For example, the reduction of sour taste of fruits by sucrose is a well known phenomenon.

7. Effect of taste modifying proteins. A taste-modifier protein, *miraculin*, has been discovered in a west African plant. When applied to tongue, this protein makes acids taste sweet.

8. Abnormalities of taste sensations obviously affect the various taste sensations.

Abnormalities of taste sensations

1. Ageusia. Ageusia refers to the absence of taste sensation. Causes of ageusia are:

- *Lesions of mandibular* division of trigeminal nerve (through lingual branch of which the chorda tympani nerve reaches the tongue) cause loss of taste sensations in the anterior two-thirds of tongue.
- *Lesions of facial nerve* also lead to loss of taste sensations in the anterior two-thirds of tongue.
- *Lesions of glossopharyngeal nerve* are associated with the absence of taste sensations from the posterior one-third of the tongue.
- *Drugs* like captopril and penicillamine, which contain sulphydryl groups cause temporary loss of taste sensation. The reason for this effect of sulphydryl compounds is not known.
- *Familial dysautonomia*. It is a congenital widespread sensory disorder characterized by the absence of taste sensations associated with other abnormalities, such as postural hypotension, lacrimation, hyporeflexia and insensitivity to temperature and noxious stimuli.

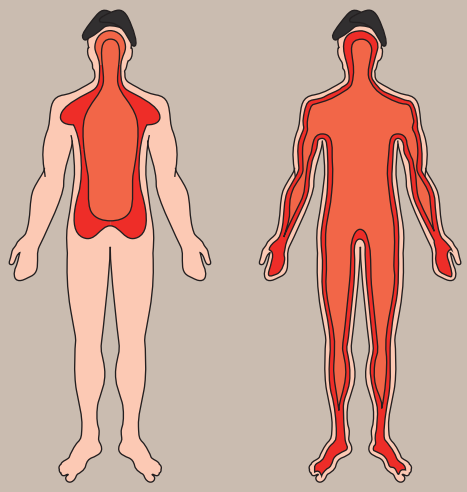
2. Hypogeusia. Hypogeusia refers to a diminished taste sensitivity. In it, the taste sensations are not completely lost but there occurs an increase in the threshold for different taste sensations. Many different diseases can produce hypogeusia.

3. Dysgeusia. Dysgeusia refers to a disturbed sense of taste. It is a feature of temporal lobe syndrome.

4. Selective taste blindness. Selective taste blindness is an inherited autosomal recessive trait characterized by markedly elevated threshold for phenyl thiocarbamide, i.e. PTC (a chemical substance with very bitter taste). Such individuals are called *non-taster for PTC*. The defect is highly selective; probably, there is a particular receptor protein, which is not synthesized in these individuals.

Specialised Integrative Physiology

SECTION 12



This section includes chapters on miscellaneous topics and not on different physiological aspects related to each other as is in the case of systemic physiology. In systemic physiology, each section, say for example section on 'Respiratory System', includes various chapters dealing with the different aspects of respiration. The only similarity between the various physiological aspects discussed in different chapters of this section is that each involves integrated role of two or more than two systems to perform a highly specialized physiological activity. For instance, '*Physiology of exercise*' (see page) includes integrated role of physiological activities related to skeletal muscle system, respiratory system, cardiovascular system, endocrinal and nervous system. Similarly, the chapter on '*Physiology of Body Temperature Regulation*' highlights the nicely integrated role of cardiovascular system, respiratory system, skin and the nervous system to perform the highly specialized task of regulating the body temperature. In other words, this section is perfect demonstration of unity in diversity. Various diverse aspects of physiology discussed in this section are enumerated above.

“This page intentionally left blank"

Chapter 12.1

Physiology of Body Temperature Regulation

BODY TEMPERATURE

HOMEOTHERMIC VERSUS POIKILOTHERMIC ANIMALS

Homeothermic animals, also called warm-blooded animals, are able to maintain their body temperature within a normal narrow range in spite of wide variations in the environmental temperature. Birds and mammals, including humans, belong to this category.

Poikilothermic animals, also called cold-blooded animals, do not have an efficient temperature regulating system; therefore, their body temperature fluctuates with the fluctuations in the environmental temperature. The reptiles, amphibians and fishes belong to this category.

NORMAL BODY TEMPERATURE

There is a slight variation in normal temperature at different parts of the body:

- *Oral temperature*, when measured with the help of a clinical thermometer, varies from 36.0 to 37.5°C (97.5–99°F) with an average of 37°C (98.6°F). Oral temperature is affected by hot and cold drinks and food, smoking, chewing gums and mouth breathing.
- *Axillary temperature* is slightly lower (about 0.5°C) than the oral temperature.
- *Rectal and oesophageal temperatures* are slightly higher (about 0.5°C) than the oral temperature.
- *Superficial skin or surface temperature* varies to some extent with the environmental temperature (see below, shell temperature).
- *Extremities* are generally cooler than rest of the body.
- *Scrotal temperature* is carefully regulated at 32°C (89.6 °F).

Concept of core versus shell temperature. The body is hypothetically divided into core and shell (Fig. 12.1-1).

- *Shell temperature,* i.e. temperature of the limbs and the surface layer of trunk, i.e. skin and underlying structure exhibits variations of the temperature with the change in the external temperature.

  In cold weather, the temperature of the shell may be several degrees lower than the core temperature (Fig. 12.1-1A). This decreases the loss of body heat to the environment by conductional radiation. In hot environment, the shell

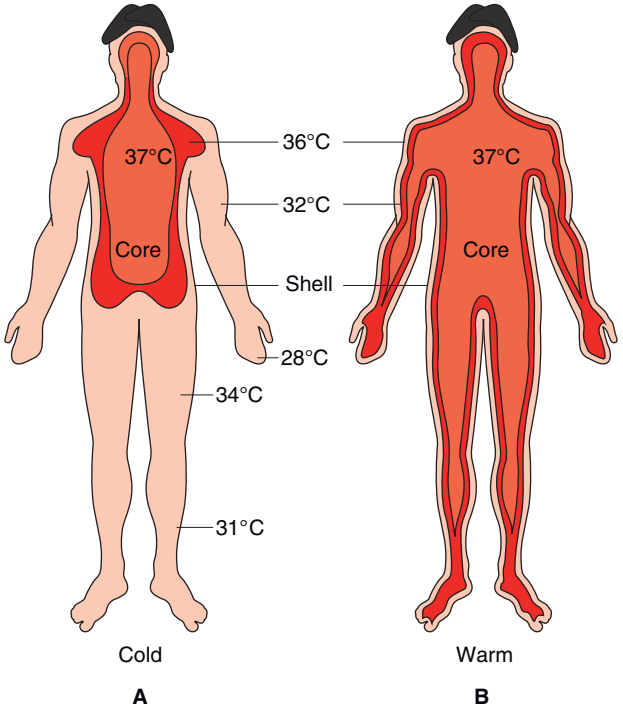


Fig. 12.1-1 Concept of body core and shell temperature. Red areas represent body core and superficial light areas represent body shell.

temperature approaches the core temperature (Fig. 12.1-1B), and this helps in heat loss by conduction and radiation.

- *Core temperature,* i.e. temperature of deeper body structures (e.g. temperature of intra-abdominal, intrathoracic and intracranial structures) is maintained strictly constant. It has been widely assumed to be an accurate index of the temperature of blood to which hypothalamic thermoregulatory receptors are exposed. The core temperature is always slightly more than the oral temperature (about 37.8°C or 100°F). Rectal, vaginal and oesophageal temperatures represent the core temperature.

Lower versus upper lethal core temperature. As shown in Fig. 12.1-2, the *lower lethal core temperature* is about 26°C, at which cardiac arrhythmias occur and lead to death due to cardiac failure. *Upper lethal core temperature* is 43.5°C, which leads to death due to heat stroke. Core temperature of 41°C for prolonged periods produces irreversible brain damage.

FACTORS AFFECTING BODY TEMPERATURE

I. Physiological variations

1. Diurnal variation. Body temperature is *highest* in the evening (after day's labour—between 5 and 7 PM) and *lowest* in early hour of morning (after night's rest between 2 and 4 AM). Difference between the two values may be 1°C.

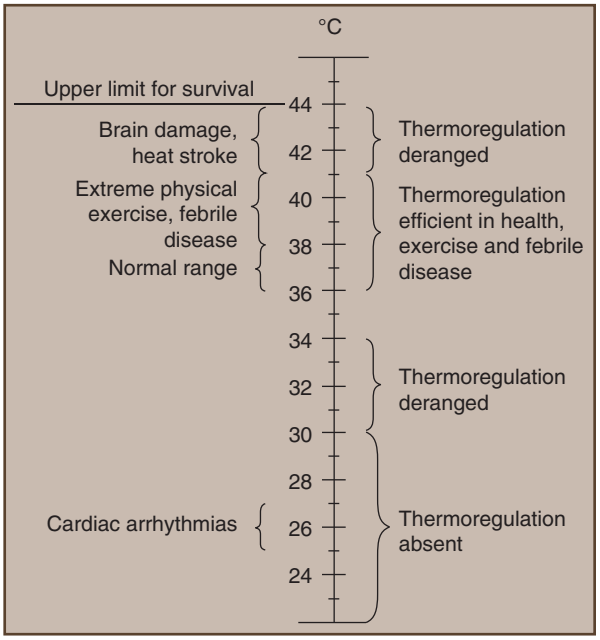


Fig. 12.1-2 Effects of alteration of body temperature.

In the night workers, the rhythm is reversed. This diurnal variation is related to exercise and specific dynamic action of food. Fasting and absolute bed rest abolish this variation.

2. Age. *Infants* have an imperfect regulation of temperature. Hence range of variation is wider. A fit of crying may raise and a cold bath may lower the body temperature.

In old age, the body temperature tends to be subnormal due to decreased activity and decreased basal metabolic rate (BMR). In addition, due to compromised circulatory system, older individuals cannot tolerate extremes of environmental temperature.

3. Sex. *Females* have a slightly low body temperature due to relatively low BMR and thick layer of subcutaneous fat (non-conductor). Further, due to thermogenic effect of progesterone, the body temperature is higher in the *post-ovulatory phase* of menstrual cycle than in the pre-ovulatory phase.

4. Size. Heat production and heat loss depends upon the ratio of mass to body surface area. In a mouse, heat production is 450 K calories/kg body weight/24 h, whereas in a horse it is only 14.5 K calories/kg body weight/24 h.

5. Food. Protein food, due to high specific dynamic action may raise body temperature. The act of ingestion of food may also raise body temperature.

6. Exercise increases temperature. Only 25% of muscular energy is converted into mechanical work, the rest comes out as heat. Inability of the heat dissipating mechanisms to handle the greatly increased amount of heat produced increases body temperature.

7. Sleep. Because of muscular inactivity, sleep results in a slight fall of body temperature.

8. Emotions. Body temperature may rise due to emotional disturbances. The rise of temperature may be as high as 2°C due to tensing of muscles.

II. Pathological variations

1. Hyperthermia or fever refers to the pathologically raised body temperature (see page 960).

2. Hypothermia refers to the lowered body temperature (below normal) due to some pathological causes, such as:

- Hypothyroidism,
- Hypopituitarism,
- Lesions in hypothalamus and
- Haemorrhage in certain parts of the brain, particularly pons.

HEAT BALANCE

Heat balance refers to the balance between the mechanisms of net heat gains by the body and mechanisms of heat losses from the body (Fig. 12.1-3). The heat balance in the body is maintained by adjusting the heat production in accordance to heat loss and vice versa. In turn, the heat balance maintains the body temperature at a constant level.

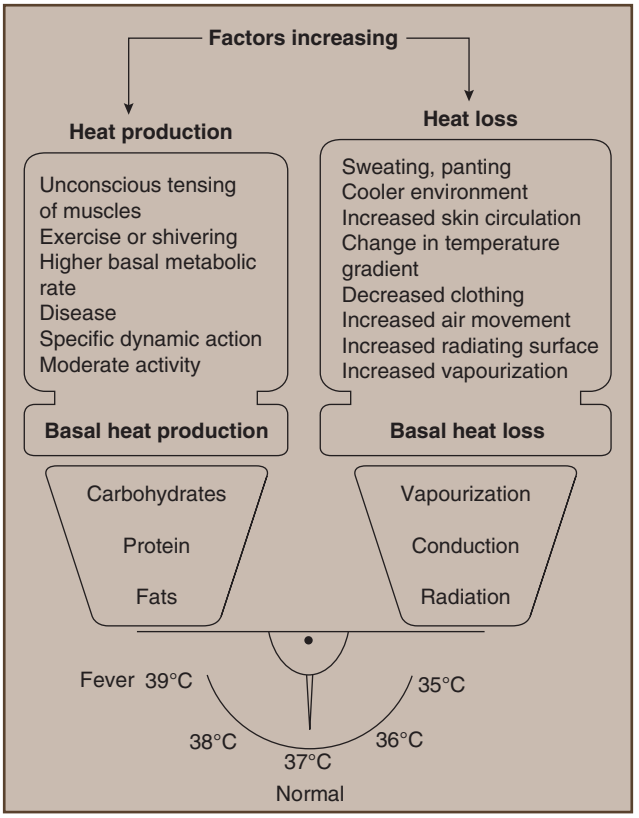


Fig. 12.1-3 Balance between factors contributing to heat gain and heat loss from the body.

MECHANISMS OF HEAT GAIN

The main mechanisms responsible for heat gain by the body are:

- Heat production in the body and
- Heat gain from the environment.

HEAT PRODUCTION OR THERMOGENESIS

Thermogenesis refers to the heat production in the body by various physiological/metabolic processes which include:

1. Basal metabolic activity. The main mechanism responsible for the heat production in the body is physiological oxidation of food materials, i.e. combustion of carbohydrates, proteins and fats. 1 g of each yields about 4, 4, and 9 calories, respectively. This is called *heat of metabolism*. Of all the organs, the liver contributes the highest amount of heat of metabolism. Heat produced by the liver and heart is relatively constant.

2. Muscular activity. Though heat produced by the skeletal muscles is variable and depends upon the physiological activity; yet skeletal muscles are a major source of heat. The heat produced during muscular activity is called *heat of activity*. Muscular activities contributing to heat production are:

(i) *Muscle tone* and unconscious tensing of muscles produce heat even when the individual is resting.

(ii) *During exercise,* a great deal of heat is produced by the skeletal muscles.

(iii) *Respiratory muscles activity* produces about 38% of activity heat.

(iv) *Shivering* refers to the muscle response to cold. It is characterized by oscillating rhythmic muscle tremors occurring at a rate of 10–20/s. As no work is performed during shivering, all the energy liberated by muscles appears as an internal heat (*shivering thermogenesis*).

3. Specific dynamic action of food is the obligatory energy expenditure that occurs during assimilation of food. Maximum heat production is seen after ingestion of protein. During digestion, the peristaltic action of intestines and the activity of various digestive glands produce heat.

4. Non-shivering thermogenesis refers to the heat production due to an increase in the metabolic rate resulting from the increased secretion of epinephrine and to certain extent thyroid hormone.

HEAT GAIN FROM THE ENVIRONMENT

Heat is gained from the objects in the environment, which are hotter than the body by following mechanisms:

1. Radiation. The body gains heat by *direct radiation* from the sun and heated ground and by *reflected radiation* from

the sky. This type of heat gain is independent of the temperature of air. The amount of heat gained by the radiation can be reduced by wearing garments, which reflect the radiations or by making use of any available shade. For example, in the desert, the body takes up more heat when naked than when covered by thin white clothes.

2. Conduction. The body surface takes up heat when immersed in hot water or when the temperature of the surrounding air exceeds than that of skin.

3. Ingestion of hot fluids and food can add a small amount of heat to the body.

4. Ventilation also adds to body heat in hot climates when air is heated.

MECHANISMS OF HEAT LOSS

Heat is lost from the body by the following routes:

- Heat loss from the skin,
- Heat loss from the lungs and
- Heat loss in the excreta.

I. Heat loss from the skin

Mechanisms of heat loss from the skin surface include (Table 12.1-1):

1. Radiation refers to the transfer of heat from an object to another object with which it is not in contact. The magnitude of heat loss by radiation depends on the size of the body surface and the average temperature difference between the skin and surrounding objects. About 50% of the total heat loss from the body occurs by radiation. The colour of clothing may play a part but the colour of human skin has no effect on the radiation.

2. Conduction refers to the heat exchange between objects at different temperatures that are in contact with one another.

Table 12.1-1 Mechanisms of heat loss from the body

Mechanism	Amount of heat loss in calories	Percentage (%)
Radiation	1500	50 } 70
Conduction and convection	600	20 }
Evaporation of water from:		
• Skin	690 } 1000	20 } 27
• Lungs	210 }	7 }
Warming inspired air	60	2
Excreta (urine and faeces)	30	1
Total	3000	100

The amount of heat transferred by conduction is proportionate to the temperature difference between two objects.

3. Convection refers to the movement of molecules of a gas or liquid at one temperature to another location that is at a different temperature. Thus the heat loss through this process depends upon the temperature of the surrounding atmosphere. Thus heat loss through convection depends upon the relative density and temperature of air and wind velocity.

Note. About 20% of heat is lost from the body by conduction and convection.

4. Evaporation. About 27% of the heat is lost by evaporation from the skin, mucous membranes and respiratory passages. Vaporization of 1 g (approximately 1 mL) of water resumes about 0.6 K cal of heat. Evaporation from skin only accounts for loss of 600 cal (20%) per day, which occurs in two forms:

(i) *Insensible water loss (perspiration).* Perspiration occurs due to continuous diffusion of fluid through the epidermis (in absence of sweating). It occurs over whole of the body surface at a uniform rate and is largely independent of environmental conditions. Perspiration amounts to about 60 mL/day and is equivalent to heat loss by evaporation of approximately 400 K cal/day. But this heat loss is not under control and, therefore, cannot be changed as required.

(ii) *Evaporation of sweat.* The eccrine sweat glands play a very important role in thermoregulation of the body. *Thermal sweating* from the eccrine sweat glands increases when the external or internal body temperature rise (details are given). Sweat is vaporized from the skin, which decreases its temperature. Evaporation decreases to a great extent if the humidity of the atmosphere is high, and thus body temperature regulation becomes seriously affected.

Factors affecting cutaneous heat loss

1. Gradient between the temperature of the skin and the environmental temperature is the most important factor determining the cutaneous heat loss especially by the radiation, conduction and convection mechanisms. The temperature of skin that depends upon the amount of heat reaching the surface from the deeper tissues can be varied by changing the blood flow to the skin depending upon the requirement. This is accomplished by the radiator system of the body, which is formed by the cutaneous circulation. For details see page 273.

2. Insulator system. The subcutaneous fat acts as the heat insulator for the body. Fat conducts heat only one-third as readily as other tissues. This insulation is important in maintaining the core temperature even though the skin temperature varies with that of surroundings. Women have a thicker layer of subcutaneous fat than men; this is partly

the reason why in winter, males feel more cold than the females. This is in spite of the fact that males produce more heat because of higher BMR.

3. Piloerector muscle contracts in response to cold and causes erection of the hair. The layer of air entrapped between the hair acts as an insulator and thus reduces the cutaneous heat loss.

4. Clothing. Woollen clothes offer better protection against cold than cotton clothes because of larger amount of air entrapped in the former. In animals, the fur and feathers serve the same purpose more effectively.

II. Heat loss from the lungs

Heat loss from the lungs occurs by three processes:

1. *Evaporation of water* in expired air causes heat loss. On an average, the water loss from the lungs is approximately 300 mL/day equivalent to heat loss of 200 Kcal. It is the main mechanism through which heat is lost in dogs and sheeps.
2. *Warming inspired air* to the body temperature accounts for 2% of heat loss in man.
3. *Panting.* Some mammals lose heat by panting. Panting refers to the rapid shallow breathing, which greatly increases the amount of water to be evaporated in the mouth and respiratory passages and thereby results in the heat loss.

III. Heat loss in the excreta

About 1% of the total body heat loss occurs in the excreta (urine and faeces).

REGULATION OF BODY TEMPERATURE

Like other homeothermic animals, humans have been provided with a *temperature control system,* which maintains the internal (core) body temperature constant within the range of ±1°F of the normal temperature. Normal body temperature is the 'set point' in the system of temperature regulation. In humans, the set point of the temperature control system is approximately 98.6°F (37°C), although it normally varies somewhat diurnally, decreasing to a minimum during sleep. The set point can be altered by pathological status, for example, by the action of pyrogens, which induce fever (see page 960).

Thermoneutral zone

Before discussing the temperature control system, it will be appropriate to know about the thermoneutral zone (TNZ). The TNZ refers to the range of ambient temperature within which the metabolic rate is at a minimum, i.e. at which the O_2 consumption at rest or when asleep is minimal and temperature regulation is achieved by non-evaporative physical processes alone (Fig. 12.1-4). The values of thermoneutral zone in naked humans are (Fig. 12.1-5):

- Adults: 26–28°C,
- Newborn infants: 32–34°C and
- Premature infants: 35°C.

Certain other terms which need mention are:

Critical temperature. It is the lower limit of thermoneutral zone. Below it the metabolic heat production of a resting thermoregulating animal increases to maintain thermal balance (Fig. 12.1-4).

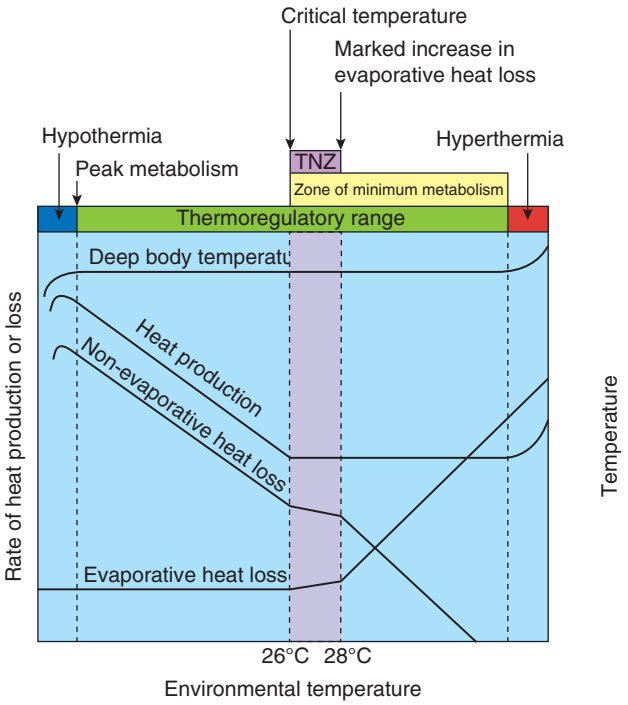


Fig. 12.1-4 Thermoneutral zone and rate of heat production and heat loss in relation to environmental temperature.

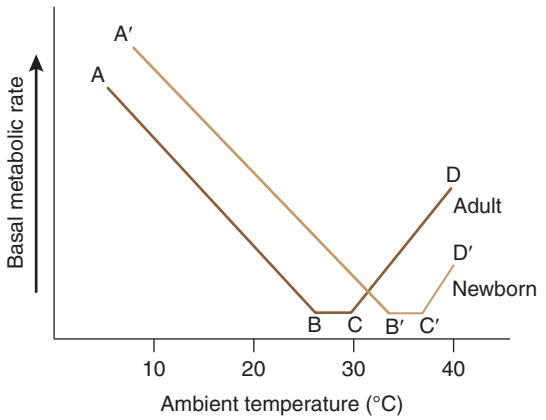


Fig. 12.1-5 Thermoneutral zone of the newborn (B′–C′) as compared to that of an adult human (B–C).

Preferred ambient temperature (PAT). It is the range of ambient temperature associated with thermal comfort. It is not the same as TNZ. It is important to note that:

- When humidity is high, the PAT would be lower than the TNZ and
- When air movement is brisk, the PAT would be higher than the TNZ.

Thermal comfort. It is maximum when the skin temperature is about 33°C and also depends upon:

- Level of humidity,
- Amount of air movement,
- Level of body activity and
- Amount of clothing.

Temperature control system

The temperature control system comprises the hypothalamus integrated autonomic, endocrine and skeletomotor responses. The components of temperature control system are:

- Thermoreceptors,
- Hypothalamus, the thermostat and integrator of temperature control system and
- The thermoregulatory effector mechanisms.

THERMORECEPTORS

Thermoreceptors or temperature receptors, which give information about the body temperature to the temperature control centre in the hypothalamus, are of two types: peripheral thermoreceptors and central thermoreceptors.

1. Peripheral thermoreceptors are present throughout the body in the skin and mucous membrane (and probably other organs, such as muscle and viscera).

- *Cutaneous thermoreceptors* sense the ambient temperature; 90% of them are cold receptors.
- *Deep receptors* present in the viscera sense the core temperature unlike cutaneous receptors that sense surface temperature. However, like cutaneous receptors, the deep receptors also mainly detect cold than warmth. Probably, both the cutaneous and deep receptors are concerned with preventing hypothermia.

2. Central thermoreceptors are mainly present in the hypothalamus. The hypothalamic receptors are probably neurons whose firing rate is highly dependent on the local temperature, which in turn is importantly affected by the temperature of blood.

HYPOTHALAMUS: INTEGRATOR OF TEMPERATURE CONTROL SYSTEM

The integrator and many controlling elements for temperature regulation appear to be located in the hypothalamus. The system acts as a *servo-mechanism* (a control system that uses negative feedback to operate another system) with a *set point* at normal body temperature (98.6 °F or 37°C). The hypothalamic neurons involved in temperature regulation can be divided into three types, depending upon the function subserved by them:

- Sensing neurons or feedback detectors,
- Heat-loss centre or antirise centre and
- Heat production and conservation centre or antidrop centre.

Sensing neurons or feedback detectors. These neurons, located in the anterior hypothalamus, collect information about temperature from both the central as well as the peripheral thermoreceptors.

Two types of neurons have been identified in this area, the *warm sensitive neurons*, which respond to information from warmth receptors and cold sensitive neurons, which respond to cold receptors.

The central and peripheral thermoreceptors rarely provide identical information about body temperature. In dealing with the body temperature, therefore, the hypothalamus calculates an integrated temperature from the feedback received.

Heat-loss centre or antirise centre is composed of neurons in the pre-optic region and anterior or rostral hypothalamus. It organizes the heat loss responses as illustrated:

- *Electrical stimulation* of this area produces heat loss by cutaneous vasodilatation, sweating, panting and decreases heat production by inhibiting shivering.
- *Lesions of heat-loss centre,* conversely, prevent sweating and cutaneous vasodilatation and they cause hyperthermia (neurogenic fever) when the individual is placed in a warm environment.

Heat production and conservation centre (heat-gain centre or antidrop centre) is formed by the neurons in the area of posterior or caudal hypothalamus, which is dorsolateral to the mammillary body. These neurons organize the heat production and conservation responses as illustrated:

- *Stimulation* of this area conserves heat by cutaneous vasoconstriction and activates heat production by evoking shivering and increasing the metabolic rate through the release of thyroid-stimulating hormone (TSH).
- *Lesions* in this area interfere with heat production and conservation, and they can cause hypothermia when the subject is in cold environment.

THERMOREGULATORY EFFECTOR MECHANISMS

The effector mechanisms of thermoregulation are integrated by a thermostat located in the hypothalamus, and

that the *hypothalamic thermostat* has a set point which is normally at 98.6°F (37°C). Error signals which represent a deviation from the set point, either due to raised or lowered body temperature, evoke responses that tend to restore body temperature towards the set point. These responses are mediated by autonomic, somatic and endocrine systems. The thermoregulatory effector mechanisms activated by hypothalamus can be grouped as:

A. Mechanisms activated by heat

I. Mechanisms increasing heat loss

1. Cutaneous vasodilatation. As described on page 274, the cutaneous vessels form the radiator system of the body. Cutaneous vasodilatation occurring on the exposure to heat stresses increases cutaneous blood flow from 4–5 mL/100 g skin weight/min to as high as 150 mL/100 g of skin weight/min. By this, warm blood from the deeper tissues is brought to the surface, and heat loss by conduction, radiation and convection is facilitated as described above.

Mechanisms of cutaneous vasodilatation. Cutaneous vasodilatation is produced through a local effect, spinal reflex as well as through hypothalamus. For details see page 274.

2. Sweating. As mentioned earlier, the evaporation of the sweat is most important mechanism of heat loss, especially when a person is exposed to the environmental temperature greater than the body temperature.

Mechanism of heat induced sweating. When the environmental temperature rises above the *thermal comfort level* (about 33°C), the sweating is induced by three mechanisms: a local response, spinal reflex as well as through hypothalamus influence. The impulses from the heat-loss centre of the anterior hypothalamus increase the impulse discharge in the sympathetic cholinergic fibres to the sweat glands (see page 276). As a result the sweating starts suddenly, and the rate of sweating progressively increases with the increase in environmental temperature.

Rate of sweating, which is practically zero in cold weather, may reach to maximum of 700 mL/h in bad weather. In very hot climate, sweat secretion may be over 10 L/day. Such a heavy sweating causes a marked loss of body water and NaCl. This happens because body homeostasis mechanisms give priority to temperature regulation over water and electrolyte regulation. Therefore, in acute heat stress death may occur due to severe dehydration and salt loss leading to circulatory failure.

Acclimatization of sweating mechanism is an adaptation which occurs following prolonged exposure to high environmental temperature. The acclimatization is very useful in conserving body NaCl.

Mechanism of acclimatization. Acclimatization occurs due to an increased secretion of aldosterone (stimulated due to slight decrease in NaCl levels of body fluids because of excessive sweating before acclimatization). Aldosterone increases absorption of Na^+ and Cl^- from the renal tubules. This reduces excretion of salt to 3–5 g/day as compared to 15–30 g/day before acclimatization.

3. Panting. Panting refers to the rapid shallow breathing, which increases heat loss by increasing water vaporization in the mouth and respiratory passages. In some animals, like dogs, panting is an effective means of heat loss. Because the breathing is shallow, it produces little disturbance in the arterial pCO_2 or pH.

II. Mechanisms decreasing heat production and heat gain from environment

1. Anorexia and lethargy. A rise in ambient temperature produces anorexia and lethargy. Anorexia results in decreased food intake, which decreases heat production because of decrease in specific dynamic action of food. Lethargy decreases muscular activity, which decreases heat of activity.

2. Behavioural responses include shelter in shade or a cooler place and preference for cold food and drinks. These acts decrease heat gain from the environment.

B. Mechanisms activated by cold

I. Mechanisms conserving heat

1. Cutaneous vasoconstriction. As mentioned earlier (see page 274), the immediate reflex response to cold is cutaneous vasoconstriction, which reduces the heat loss from the body core to the surface and thus conserves heat.

Mechanism. Like vasodilatation, the cutaneous vasoconstriction is produced through:

- Direct local effect of cold,
- Local spinal reflex, evoked through peripheral thermoreceptors and
- Hypothalamus controlled cutaneous vasoconstriction. Impulses from 'heat production and conservation centre' (which is located in the posterior hypothalamus) increase the sympathetic discharge to the cutaneous vessels causing extreme vasoconstriction. As a result, the practically blood less skin prevents heat loss by becoming an insulating barrier between the warm core of the body and the cold environment.

2. Piloerection, i.e. cold induced erection of the body hair, as mentioned earlier, entraps a layer of air in the hair, which acts as an insulator and thus reduces the cutaneous heat loss.

3. Behavioural responses, which help heat conservation, are:

- Curling up while sleeping. It reduces the body surface area in contact with the environment.
- To put warmer clothes also prevent heat loss.

II. Mechanisms increasing heat production

1. Shivering thermogenesis. When the heat loss prevented by the cutaneous vasoconstriction is not sufficient to cope up with the environmental cold, heat production is increased by shivering (for details see page 745). Shivering is evoked in response to the impulses from the 'heat production and conservation centre' located in the posterior or caudal hypothalamus.

2. Chemical (non-shivering) thermogenesis. It refers to an increased heat production due to increased cellular metabolism. Chemical thermogenesis occurs because of following effects:

(i) *Increased sympathetic stimulation and increased secretion of catecholamines* (epinephrine and norepinephrine), which occurs as a part of response to cold.

Role of brown fat. The amount of chemical thermogenesis occurring in response to increased catecholamine secretion is proportional to the amount of brown fat in the tissues, because the brown fat contains a large number of special mitochondria where the uncoupled oxidation occurs. Adults do not have brown fat; therefore, heat production by this mechanism is increased by only 10–15%. However, in infants (who have brown fat in interscapular region, around the neck, behind the sternum and around the kidneys) increased secretion of catecholamines may increase heat production by 100%.

(ii) *Increased secretion of thyroxine* also promotes chemical thermogenesis. After several week of exposure to severe cold, hyperplasia of thyroid gland and increased secretion of thyroxine can be demonstrated. This occurs because the cold temperature stimulates hypothalamic release of TRH, which in turn stimulates the secretion of TSH and thyroxine. Consequently, thyroxine secretion in winters is somewhat greater than in the summer. In infants, even short exposure to cold increases thyroxine secretion.

3. Behavioural responses associated with increased heat production are:

(i) *Hyperphagia* helps in increased heat production because of specific dynamic action of food.

(ii) *Hyperactivity,* e.g. in the form of rubbing of palms also increases heat production.

(iii) *Seeking out sources of heat,* e.g. standing out in the sun, or heat fire and also consumption of hot food and hot drinks, help in gaining heat.

ABNORMALITIES OF BODY TEMPERATURE REGULATION

FEVER

Fever, also known as *pyrexia*, refers to an increase in the body temperature above the normal range. It is the most common symptom/sign of the ill health.

Causes. Common causes of fever are:

1. *Infections* caused by bacteria, viruses, protozoa (e.g. malaria) and other infecting agents are usually associated with fever.
2. *Tissue destruction,* as in myocardial infarction, uninfected neoplasms, serum sickness and rheumatism, etc. is also associated with fever.
3. *Pyrexia of unknown origin.* This term is used when cause of fever cannot be ascertained.

Pathogenesis. Fever develops when the hypothalamic set point is reset at a higher temperature by the pyrogens as explained (Fig. 12.1-6):

Role of pyrogens. Toxins liberated from the infecting organism and tissue destruction act on the phagocytic cells

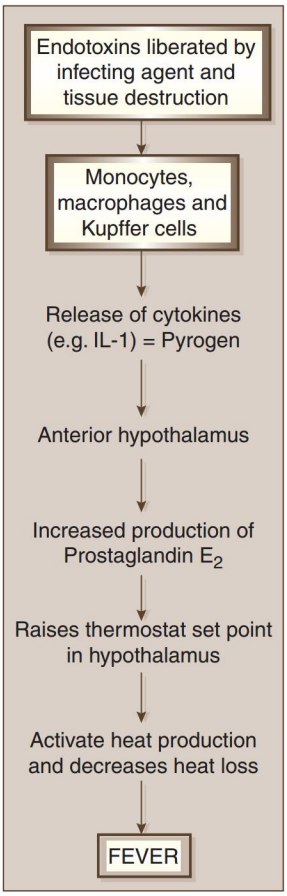


Fig. 12.1-6 Pathogenesis of fever.

(monocytes, macrophages and Kupffer cells) to produce cytokines that act as *endogenous pyrogens.* The pyrogens are polypeptides and include interleukin-I (IL-I) and other cytokines, which act on the anterior hypothalamus to increase the production of prostaglandin E_2. Prostaglandin E_2 acts on the hypothalamus to increase the thermostat 'set point'.

IMPORTANT NOTE

The drugs like aspirin, which prevent the formation of prostaglandin E_2 from arachidonic acid, act as antipyretics (which lower the temperature).

Production of fever. Once the thermostat set point is raised by the pyrogens, the heat producing mechanisms and heat conserving mechanisms of the body are activated till the body temperature equals the elevated hypothalamic thermostat set point, i.e. till fever is produced. Because of these mechanisms, during production of fever there occurs:

- Shivering (which produces heat),
- Skin vessels are constricted to minimize heat loss,
- Rate of metabolism is increased which increases further heat production and
- Chills are felt in fever when the heat generating and heat conserving mechanisms are active.

Termination of fever. When the causes producing pyrogens are removed, the set-point of hypothalamic thermostat is reset back to normal. At this juncture, since the body temperature is higher than the set point of thermostat, the heat production is decreased and mechanisms of heat loss are activated. Because of these mechanisms during termination of fever there occurs:

- Cutaneous vasodilatation and
- Profuse sweating.

This sudden change in the febrile condition associated with profuse sweating and red and hot skin is called *crisis* or *flush.*

Beneficial effects of fever include:

- *Inhibition of growth of bacteria, viruses* and other infecting organisms occurs at a high body temperature. Because of this effect, artificial flue therapy was used to treat infections before the advent of antibiotics.
- *Antibody production* is increased when body temperature is raised.
- *Growth of some tumours* is slowed down by the increased body temperature.

Harmful effects of fever include:

- Dehydration, negative nitrogen balance, loss of NaCl and alkalosis (because of hyperventilation).
- Permanent damage to the brain, kidney and liver may occur when core temperature is more than 41°C (hyperpyrexia) for prolonged period.
- Death may occur due to heat stroke when temperature rises above 43°C.

HEAT EXHAUSTION AND HEAT STROKE

Heat exhaustion refers to a condition of circulatory failure caused by excessive sweating following prolonged exposure to heat. It is characterized by dehydration, salt loss, decreased blood volume, decreased arterial pressure and syncope (fainting).

Heat stroke usually occurs when heavy physical work is performed in hot and humid environment. In this condition normal response to increased ambient temperature (sweating) is impaired and core temperature increases to the point of tissue damage. Convulsion, loss of consciousness and even death may occur when body temperature exceeds 41°C.

MALIGNANT HYPERTHERMIA

Malignant hyperthermia is caused in susceptible individuals by inhalation anaesthesia. It is characterized by a massive increase in oxygen consumption and heat production by the skeletal muscle, which causes a rapid rise in body temperature. In susceptive individuals, a defective ryanodine receptor due to mutation of gene coding leads to excess Ca^{2+} release during muscle contraction triggered by stress.

HYPOTHERMIA

Hypothermia results when the ambient temperature is so low that the body's heat generating mechanisms (e.g. shivering and metabolism) cannot adequately maintain core temperature near the set point. Infants and old people develop hypothermia more easily than the adults.

It has been observed that:

- At rectal temperature of 28°C, the body's ability to spontaneously return the temperature is lost.
- Humans can tolerate body temperature of 21–24°C without permanent ill effects, i.e. if rewarmed with external heat, returns to a normal state.

Effects of hypothermia on body include:

- Slowing of metabolic and physiologic processes,
- Retardation of glucose metabolism,
- Slowing of respiration and heart rate,
- Lowering of blood pressure,
- Slowing of reflexes and occurrence of muscular rigidity,
- Loss of consciousness and
- Death may occur when temperature remains below 25°C for some time.

Accidental hypothermia occurs after due to prolonged exposure to cold air or cold water, e.g. after ship wreck or accidents in high mountain. It is a serious condition and requires careful monitoring and prompt rewarming.

Induced hypothermia. The fact that human body can tolerate hypothermia (of 21–24°C) for quite some time without ill effect has been explained for use in heart and brain surgery. The induction of hypothermia during surgery is made easier with the use of anaesthesia and muscle relaxants, both of which abolish shivering.

POIKILOTHERMIA

Poikilothermia refers to a condition in which the individual is not able to maintain core body temperature during fluctuations in the environmental temperature. That is, with increase in the environmental temperature the body temperature increases and with the decrease in environmental temperature, the body temperature decreases. Such a condition of impaired thermoregulation occurs in the hypothalamic lesions and brain stem lesions that interrupt descending hypothalamic fibres to the spinal cord.

Chapter 12.2

Physiology of Growth and Behavioural Development

GROWTH AND DEVELOPMENT
- Growth curves
- Factors affecting growth
- Growth factors

BEHAVIOURAL DEVELOPMENT
- Behaviour pattern
- Developmental and intelligent quotient
- Milestones

GROWTH AND DEVELOPMENT

The terms growth and development are intimately interdependent and interacting with each other. Growth per se refers to an increase in the physical size, as the child grows to adulthood, while development refers to maturity, i.e. improvement in the capability of the tissue. They are, therefore, termed together to signify a process of maturation, both in quality and quantity.

GROWTH CURVES

Growth of different parts of the body does not follow a uniform pattern. The patterns of growth of different parts are described in the form of different growth curves (Fig. 12.2-1):

1. General growth curve

General growth curve shows two growth spurts: one in infancy and another around puberty (Fig. 12.2-1A). General growth refers to the increase in height, weight and growth of the skeletal muscle, blood volume, respiratory system, cardiovascular system, gastrointestinal tract and excretory system.

Infancy growth spurt. Weight and height are generally considered a good index of the child's growth potential and a delicate measure of the individual's health.

First spurt of growth occurring in infancy is characterized by an increase in birth weight to two times by 6 months of age, three times at 1 year of age, four times at 2 years of age. After this the weight increases by about 2 kg/year till about 12 years of age.

Height increases in increments of 2–2.5 cm/month in the first year of post-natal life. Thereafter it slows down (Table 12.2-1).

Adolescent growth spurt. It is characterized by a rapid increase in weight gain, about 3.5 kg/year between 12 and 18 years of age (Table 12.2-1). During this period, the rapid growth of bones is brought about by the various endocrinal influences. After this, the rate of growth again slows down. *Height* increases during adolescent growth of spurt varies between 4 and 7 cm/year depending on the genetic potential and endocrinal factors (Table 12.2-1).

2. Neural growth curve

Neural growth curve shows that brain, spinal cord and visual apparatus grow very rapidly after birth (Fig. 12.2-1B). At the end of first year of post-natal life, the brain has already achieved 2/3rd and by the end of second year 4/5th of adult size. By 5 years of age, brain is almost fully developed and the child is ready for education and training. The measurement of head circumference (which also increases with the

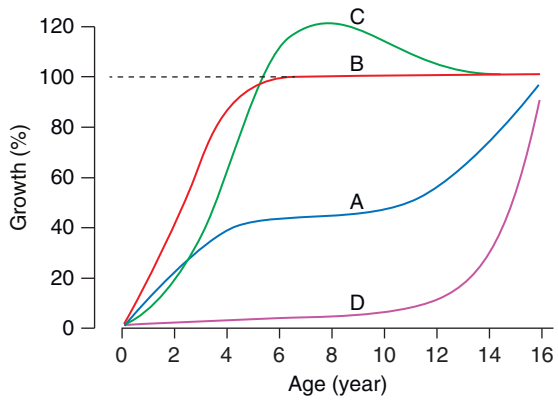


Fig. 12.2-1 Different growth curves: A, general growth curve; B, neural growth curve; C, lymphoid growth curve and D, gonadal growth curve.

Table 12.2-1 Mean weight and height at different ages in Indian children

Age	Weight (kg)		Height (cm)	
	Male	Female	Male	Female
Birth	2.8	2.7	48.5	47.7
3 month	4.5	4.5	60.2	58.5
6 month	6.2	6.0	65	63.6
9 month	7.9	7.5	68.7	67.6
1 year	8.9	8.4	73.9	72.8
3 year	12.6	12	88.8	87.2
6 year	16.3	16	108.5	107.4
9 year	21.5	21.3	123.7	122.9
12 year	28.5	29.8	138.3	139.2
15 year	39.6	36.8	155.5	149.6
18 year	47.4	42.4	163.1	151.7

Source: Indian Council of Medical Research (ICMR).

growth of brain) is thus very important up to 3 to 5 years of age to get information of the developing brain inside.

3. Lymphoid growth curve

The growth of lymphoid organs, such as tonsils, adenoids, thymus, spleen, lymph nodes and lymphoid tissue of the intestine is very rapid in infancy and childhood and is followed by a partial involution at puberty (Fig. 12.2-1C). Because of this reason, the size of tonsils and adenoids at the age of 8–10 years is larger than in the adults. This is important to recognize that more enlargement of tonsils is not an indication for removing them surgically.

4. Gonadal growth curve

The gonads and accessory organs of reproduction remain in dormancy in childhood and grow at a remarkable rate around puberty. Thus, gonadal growth pattern is essentially opposite to the neural growth pattern (Fig. 12.2-1D).

FACTORS AFFECTING GROWTH

Marked differences existing between the growth and development patterns of various races and communities and different individuals in the same race and community are determined by various factors which affect the growth:

1. Genetic factors. The ultimate growth pattern of an individual is largely determined by the genetic inheritance, either maternal or paternal.

2. Hormonal factors. The important hormones which affect the growth and development are:

(i) *Growth hormone* (GH) secreted by anterior pituitary plays an important role in the growth and thus determines the height of an individual during childhood (see page 539). Therefore, deficiency of GH produces growth retardation known as *pituitary dwarfism*.

(ii) *Insulin-like growth factor-I* (IGF-I), as the name indicates, chemically resembles insulin. It is one of the polypeptides, which are collectively known as somatomedins. IGF-I, like GH, promotes protein synthesis, epiphyseal growth. Some of its actions are opposite to GH. For details see page 541.

(iii) *Thyroxine* plays an important role in the growth and development by its direct action, and also indirectly by potentiating the release of GH and somatomedins (see page 555). Congenital deficiency of thyroxine results in a clinical condition called 'cretinism', which is characterized by retardation of physical as well as mental growth.

(iv) *Sex hormones.* The oestrogens and androgens cause maturation and are important in the adolescent age. The anabolic actions of sex hormones, adrenal androgens, growth hormone and IGF-I seem to potentiate each other producing a marked growth spurt during puberty.

(v) *Insulin* is an anabolic hormone and thus its role in growth and development cannot be overemphasized. Congenital deficiency of insulin results in juvenile diabetes mellitus. Retardation of growth is one of the characteristic features of juvenile diabetes mellitus.

Relative importance of different hormones at various stages of growth is (Fig. 12.2-2):

- *Thyroxine* is essential for growth in late fetal life and first few years of post-natal life (Fig. 12.2-2A).
- *Growth hormone* in contrast to thyroxine does not seem to be of critical importance during fetal and early post-natal life. Infants with congenital deficiency of GH have normal height and weight up to about 2 years of age, after which a decrease in the velocity of growth becomes apparent (Fig. 12.2-2B).
- *Insulin*, being an anabolic hormone, remains important during fetal as well as post-natal growth (Fig. 12.2-2C).
- *Sex hormones* are most essential around puberty and play an important role in the development of gonads as well as general growth (Fig. 12.2-2D).

3. Nutritional factors. The major ingredients of diet, viz. proteins, fats, carbohydrates, minerals and vitamins are important for optimal growth, both pre-and post-natally. Protein and its various amino acids are essential for laying down the new tissues, for wear and tear and for specific metabolic functions. Their requirements are increased during the active periods of growth.

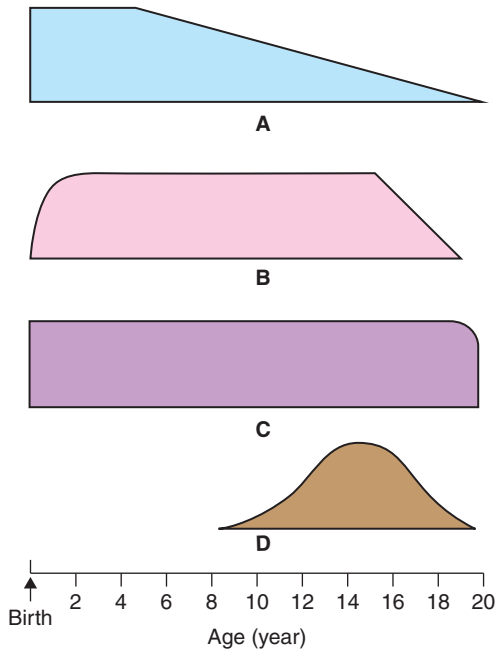


Fig. 12.2-2 Relative importance of different hormones at various stages of growth: A, thyroxine; B, growth hormone; C, insulin and D, sex hormones.

Lack of nutritional factors affects normal growth even when the genetic and hormonal factors are normal. Undernutrition and malnutrition in childhood is responsible for smaller status, poor muscular development, and generalized apathy in underdeveloped countries.

4. Illnesses. Congenital anomalies compatible with life are likely to retard the growth of the child. Acute illnesses temporarily depress growth. Chronic long standing illnesses including long standing infections can markedly interfere with the normal pattern of physical and mental growth and can produce permanent growth retardation.

5. Emotional factors. Lack of love, security and a disturbed child–parent relationship results in various psychological and behavioural problems during childhood and adolescence, which result in distortion of normal development and achievement of maturity.

6. Internal milieu. Normal metabolism and co-ordinate functioning of all the organs result in the optimum growth. Disturbed internal metabolism in various liver and kidney diseases precludes normal growth.

7. Environmental factors. The environment influences the growth pattern right from the intrauterine life. Faulty position of the fetus, faulty implantation of the ovum, rubella syndrome, etc. are amongst the earliest environmental factors influencing growth. Birth injuries, during difficult labour and other factors during natal period interfere with the normal development. Exposure to various seasonal variations has similar effect. The countries in the temperate zone have a small average height because of suboptimal environment. Lack of sunshine and poor personal hygiene also affect the normal growth.

8. Socioeconomic factors also play an important role in growth and development of child.

Table 12.2-2 Actions of commonly known growth factors

Name of growth factors	Action
Growth hormone	General tissue growth
Insulin-like growth factor (IGF-I)	General tissue growth
Erythropoietin	Proliferation of red cell precursors
Thrombopoietin	Proliferation of platelet precursors
Colony-stimulating factors	
Granulocyte-stimulating factor	Proliferation of granulocyte and monocyte
Lymphokines	Proliferation of lymphocytes
Epidermal growth factor (EGF)	Proliferation of epithelial cells, fibroblasts and glial cells
Platelet-derived growth factor	Growth of vascular smooth muscle
Fibroblast growth factor	Proliferation of fibroblast, endothelial cells and vascular smooth muscle
Nerve growth factor	Growth and maintenance of neurons

GROWTH FACTORS

Growth factors refer to a number of polypeptides, which regulate the proliferation of cells in the embryonic and post-natal life and thus ultimately affect the growth of various tissues.

Actions of growth factors. Actions of some of the important growth factors are summarized in Table 12.2-2.

BEHAVIOURAL DEVELOPMENT

Behavioural development of a child is studied through various responses which the child exhibits, following a natural or experimental stimulus. The developmental status depends not only on the age, but also on the environment. The age determines the proper physical and biochemical growth of various constituent organs, and the environment determines the experiences during the process of learning.

Hence both factors are essential to determine a particular behaviour, which in other words could be called a milestone in the development process.

Behaviour pattern

Behaviour pattern can be divided into four groups, for the sake of convenience, to observe the development in different age groups (Gessel):

1. Motor behaviour. *Gross motor behaviour* is the behavioural response in ventral suspension, supine, prone, sitting, standing and walking postures. *Fine motor behaviour* is seen in the form of grasp and manipulation of the objects, e.g. cube, pellet and string.

2. Adaptive behaviour. This includes sensory and motor adjustments to objects and co-ordination of eyes and hands to adjust with the simple problem situations, which are set before the infant.

Table 12.2-3 Developmental milestones at different ages

Milestone (motor)	Age (weeks)
Holding of head with bobbing	12
Head control	16
Sitting with support	20
Sitting without support	26
Standing with support	32
Standing without support	36–40
Crawling on belly	30
Crawling on knees	32
Walking with support	45
Walking without support	52

3. Language behaviour. This includes all visible and audible forms of communication whether by facial expression, gestures, postural movements, vocalisation of words, phrases or sentences and mimicry. Articulate speech depends upon social milieu and readiness of sensory, motor and cortical structures.

4. Personal–social behaviour. Child's individual reaction depends primarily upon the neuromotor maturity and the social culture in which the child lives. These include bladder and bowel control, feeding abilities, sense of priority, self-dependence in play, cooperativeness and emotional responsiveness to various stimuli.

Developmental and intelligent quotient

Developmental quotient (DQ) represents the proportion of normal development that is present at any given age. The measurement of the former depends upon the achievement of the adaptive behaviour, e.g. prehension, locomotion and manipulation.

$$DQ = \frac{\text{Mental age}}{\text{Chronological age}} \times 100$$

Intelligent quotient (IQ). The child can make use of his intellectual capabilities at the age of 5–6 years, because mental development is almost complete by this age. Therefore, intelligent quotient can only be applied at this age, unlike that of developmental quotient which can be tested at any age after birth.

$$IQ = \frac{\text{Maturity age}}{\text{Chronological age}} \times 100$$

Milestones

Some of the important developmental milestones at different ages are depicted in Table 12.2-3.

Chapter 12.3

Physiology of Fetus, Neonate and Childhood

INTRODUCTION

An infant or a child is not a miniature or small adult, rather the difference between a child and an adult is more than that of the size. The physiological responses to environmental stresses, diseases and drugs are a lot different in a child and an adult. It is because of the fact that some of the organs develop during different stages of infancy or childhood, and in other organs, the function is not as well developed as in an adult. The difference is maximum between a newborn and an adult, and this gap gradually narrows down with the growth of a newborn into an adult. There exist both quantitative and qualitative differences between physiological responses of a child and an adult in each organ system. Therefore, this section discusses some basic differences and specific details of systemic physiology of the fetus, neonate, childhood and how it changes during progress towards adulthood.

ROLE OF PLACENTA IN FETAL PHYSIOLOGY

UTERINE AND PLACENTAL CIRCULATION DURING PREGNANCY

Uterine circulation

As described on page 664, the blood supply to the uterus comes through uterine arteries and fluctuates cyclically along with the menstrual cycle to fulfill the metabolic demands of myometrium and endometrium. During pregnancy, the uterine blood flow increases parallel to the increase in fetal weight and uterine size (Fig. 12.3-1A). During early pregnancy, a rise in the levels of oestrogen and progesterone leads to an increase in the uterine blood flow, which meets the increased O_2 demand. Eventually, placenta develops and becomes the circulatory link between the mother and the fetus. Owing to increasing demand of O_2 with the progression

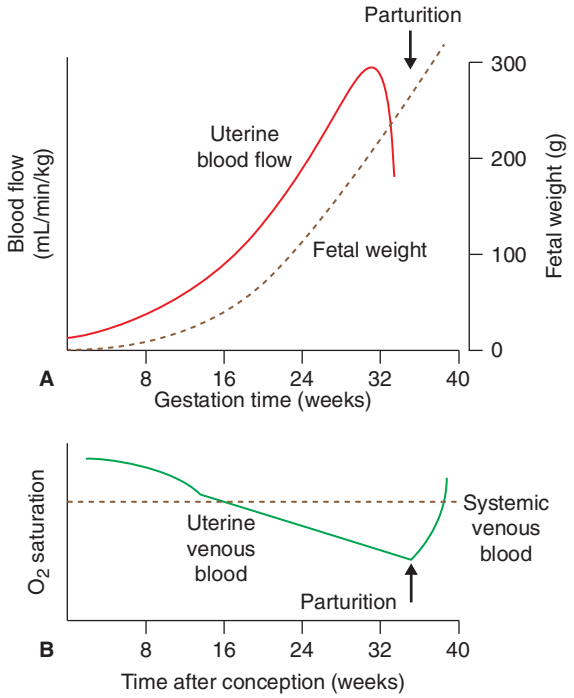


Fig. 12.3-1 Uterine circulation during pregnancy: A, changes in the uterine blood flow and B, changes in the amount of O_2 in the venous blood.

of pregnancy, more and more O_2 is extracted from the uterine blood and consequently in later part of pregnancy the O_2 saturation of uterine blood falls (Fig. 12.3-1B). As shown in Fig. 12.3-1A, the uterine blood flow increases tremendously (200–300 mL/min/kg of uterine mass including the fetus) during late pregnancy. To provide for it, the maternal cardiac output increases by 2–2.5 L/min near full term. Eighty percent of the uterine blood flow enters the placenta. Just before parturition, there occurs a sharp decline in the uterine blood flow, but the significance of this is not yet known.

Placental circulation

Placenta, which forms the circulatory link between the mother and the fetus also works as *fetal lung*, *fetal gut* and *fetal kidney*. It consists of two major portions (Fig. 12.3-2):

Maternal portion of the placenta is in fact a large blood sinus. The maternal blood flows through the uterine arteries into the maternal sinuses and back through the uterine veins of the mother.

Fetal portion of the placenta consists of placental villi. The fetal blood flows through umbilical arteries into the capillaries of placental villi and back through an umbilical vein into the fetus.

EXCHANGE BETWEEN MATERNAL AND FETAL BLOOD ACROSS PLACENTAL MEMBRANE

As shown in Fig. 12.3-2, the placental villi containing fetal blood in capillaries project into and are then bathed by the

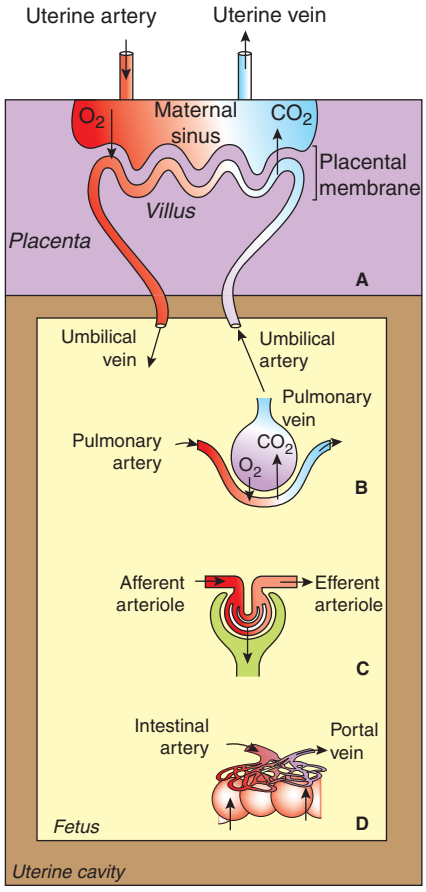


Fig. 12.3-2 Placental circulation: A, diagrammatic depiction of exchange across placental membrane between mother and fetal blood; B, diagrammatic depiction of gas exchange across pulmonary alveolus during extrauterine life. Observe the similarity with the process in placenta. For practical purposes, the umbilical artery can be compared to pulmonary artery and umbilical vein to the pulmonary vein; C, diagrammatic depiction of nutrient absorption from the gut during the extrauterine life. The similar process occurs in the placenta where umbilical artery can be compared to intestinal artery and umbilical vein to portal vein and D, diagrammatic depiction of glomerular filtration during extrauterine life. In the fetus, similar process takes place at placenta where the umbilical artery can be compared to the afferent arteriole and umbilical vein to the efferent arteriole.

blood in the maternal sinuses. Hence, in the placenta, the maternal and fetal blood do not mix with each other but are separated by the so-called placental membrane, which consists (from fetal side) of following layers (Fig. 9.4-2, page 664):

- Endothelium and basement membrane,
- Surrounding mesenchymal tissue (connective tissue),
- Cytotrophoblast and its basement membrane and
- Syncytiotrophoblast.

All exchange of O_2, nutrients and waste products between the maternal and fetal blood takes place through the placental membrane barrier.

Table 12.3-1 Values of pO_2 and pCO_2 in the maternal and umbilical cord blood

Blood vessel	pO_2 (mm Hg)	pCO_2 (mm Hg)
Uterine artery	95	36
Uterine vein	50	40
Umbilical artery	20	50
Umbilical vein	35	43

1. Gaseous exchange at placenta: placenta as lung (Fig. 12.3-2)

As shown in Fig. 12.3-2A, O_2 is taken up by the fetal blood and CO_2 is discharged into the maternal circulation across the placental membrane in a fashion analogous to O_2 and CO_2 exchange in the lungs across the alveolocapillary membrane (Fig. 12.3-2B). However, it is important to note that placental membrane (Fig. 9.4-4) is much thicker and less permeable than the alveolar membrane (Fig. 5.4-6), and therefore, the exchange is much less efficient. Table 12.3-1 shows the values of gaseous interchange in the placenta.

2. Placental transfer of nutrients: placenta as gut

See page 667.

3. Excretion of waste products through placenta: placenta as kidney

See page 667.

SYSTEMIC PHYSIOLOGY OF FETUS, NEWBORN AND CHILDHOOD

CARDIOVASCULAR PHYSIOLOGY

FETAL CIRCULATION

Pattern of fetal circulation

Pattern of fetal circulation shown in Fig. 12.3-3 and represented diagrammatically in Fig. 12.3-4 are:

Umbilical vein brings oxygenated blood from the placenta, which acts as lungs for the fetus. This blood is 80% saturated with O_2 (compared with 98% saturation in the arterial circulation in adults). The umbilical vein, before supplying the blood to the liver, bypasses some of the blood to the inferior vena cava through **ductus venosus** (Fig. 12.3-3).

Inferior vena cava thus receives some blood (80% saturated with O_2) from the umbilical vein through ductus venosus, and other blood from the hepatic veins and systemic veins draining from the trunk and inferior extremities

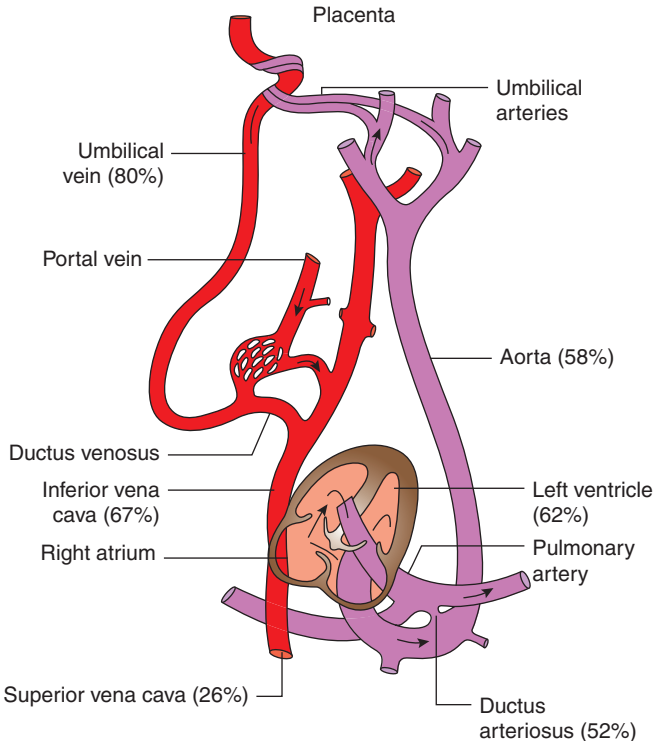


Fig. 12.3-3 Pattern of fetal circulation with pO_2 in different components.

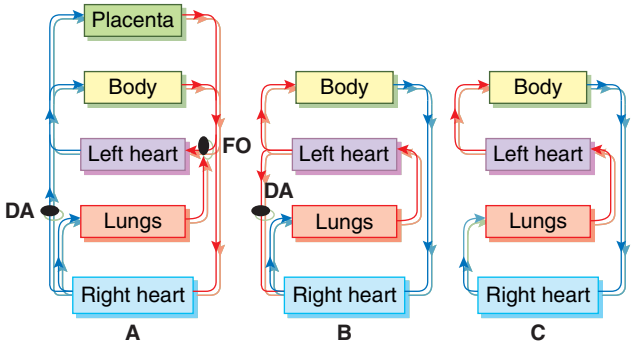


Fig. 12.3-4 Diagrammatic depiction of circulatory pattern: A, fetus; B, newborn and C, adult.

(26% saturated with O_2). The mixed blood from inferior vena cava (with approximate 67% saturation) then enters the right atrium.

Right atrium receives blood from the inferior vena cava (saturation 67%) as well as superior vena cava (saturation 26%). The fate of blood entering the right atrium is very different from that in adult:

- From the right atrium, majority of the blood coming from the inferior vena cava (saturation 67%) passes to the left atrium directly through the *foramen* ovale (an opening in the interatrial septum) and joins the blood coming from the pulmonary vein (saturation 42%). The mixed blood from left atrium (saturation 62%) passes on to the left ventricle.

- From the right atrium, most of the blood coming from superior vena cava (26% saturation) and a small amount of that coming from the inferior vena cava (saturation 67%), passes into the right ventricle. This mixed blood from the right ventricle (saturation 52%) is pumped into the pulmonary artery. But, since the fetal lungs are collapsed, their vascular resistance is very high. Hence only a small fraction of blood passes through the lungs to reach the left atrium via the pulmonary veins. Bulk of the pulmonary artery blood enters the descending aorta directly by a vascular connection called the *ductus arteriosus*.

Left ventricle pumps the blood (with saturation 62%) into the ascending aorta, from where most of the blood goes into the vessels of the head and neck and forelimbs and only a small amount of blood goes to the descending aorta.

Descending aorta, thus receives blood mainly from the pulmonary artery through ductus arteriosus (with saturation 52%) and only a small amount from the left ventricle (with saturation 62%). The descending aorta then supplies the blood (with saturation 58%) to the whole body (minus head and neck and forelimbs) and also to the placenta via *umbilical arteries* for oxygenation.

Special features of fetal circulation

1. The two ventricles work in parallel (of in series in adults), because of the presence of foramen ovale (Fo) and ductus arteriosus (DA), to drive the blood from the great veins into the arteries.

2. The two ventricles have equal thickness. This is due to the fact that right ventricle has to pump blood against considerable resistance. First, because of high resistance of the pulmonary vasculature and secondly, because of the reason that the major part of the right ventricular output is pumped into the aorta via the ductus arteriosus. The latter is possible only if the pulmonary artery pressure is higher than the aortic pressure. In fact, the fetal right ventricular systolic pressure is a few mm higher than the left ventricular pressure.

3. The two ventricles do not have similar cardiac output. The left ventricular output is approximately 20% greater than the right ventricular output. The disparity between the outputs of two ventricles does not produce any haemodynamic complications because, unlike in adults, the ventricles work in parallel, as mentioned above.

4. Fetal heart pumps only 40% of the output to the systemic circulation and 60% to the placenta. This is first because of the fact that fetal lungs are mainly nonfunctional, liver is partially functional and peripheral resistance of fetal vessels is high because of low level of activity. Secondly, on the other hand, the resistance of the placenta is low because of the large cross-sectional area of chorionic villi.

5. Oxygen saturation of the fetal arterial blood supplying the tissues is much lower (approximately 60%) than that of the adults (approximately 98%). However, fetus shows remarkable adaptation to low pO_2, because of following compensatory mechanisms:

(i) Greater affinity of fetal haemoglobin (HbF) for O_2 (see page 328),
(ii) Greater concentration of haemoglobin in fetus (18–20 g/dL),
(iii) Double Bohr's effect allows increased uptake of O_2 by the fetal blood (see page 668) and
(iv) Fetal tissues, as well as blood vessels are highly resistant to the effects of hypoxia.

NEONATAL CIRCULATION

Changes in circulation after birth

1. Arrest of umbilical blood flow and placental transfusion

Factors responsible for arrest of umbilical blood flow and placental transfusion are:

(i) Vasoconstriction of umbilical vessels. Immediately after birth, there occurs a sudden and marked reduction in the blood flow through the umbilical vessels. It results from vasoconstriction of the umbilical vessels in response to:

- Mechanical stimulation,
- Exposure to cold air and
- Secretion of catecholamines from the infant's adrenal medulla due to stress.

(ii) Tying and cutting of the umbilical cord. The process which is initiated by the nature by producing vasoconstriction is completed by the doctor by tying and cutting the umbilical cord.

Precautions to be taken while tying the umbilical cord:

- Milking of the umbilical cord should not be done, because it can send so much blood to the infant that its circulation may get overloaded.
- Infant should be held at or slightly below the level of vagina; as it leads to transfer of an additional 60–80 mL of blood to the infant, which is very useful.
- Optimum time to tie the umbilical cord is 40–60 s after birth. Tying earlier than this will prevent the transfer of additional blood from placenta for the infant, and delay in tying may be associated with risk of blood flow from the infant to the placenta producing haemorrhagic anaemia.

2. Closure of foramen ovale

Factors leading to closure to foramen ovale are:

(i) Reduction in inferior vena cava left atrial pressure gradient. The valve of foramen ovale is held open before the birth by the pressure and momentum of blood flowing up to the inferior vena cava. After birth, closure of the umbilical circulation immediately decreases the volume of blood flowing up the inferior vena cava and causes contraction of ductus venosus within 1–3 h after birth. Consequently, there occurs a reduction in the inferior vena cava-left atrial pressure gradient favouring functional closure of the valve of foramen ovale.

(ii) Decrease in right atrial and right ventricular pressure. Following arrest of umbilical flow, after birth, the infant develops asphyxia (i.e. pO_2 is lowered and pCO_2 is raised). Fetal asphyxia leads to an intense peripheral chemoreceptor discharge resulting in an initiation of breathing. The infant gasps several times and the lungs expand. The inflation of lungs after birth causes 10% decrease in the pulmonary vascular resistance. The decreased resistance causes a large drop in the after load of the right ventricle and lowers the right atrial and right ventricular pressure. The drop in the right-sided pressure causes pressure in the left atrium to exceed the pressure in the right atrium. This pressure reversal causes the flap-like valve to close over the foramen ovale.

The valve then normally fuses to the interatrial septum over the next few days.

Closure of the foramen ovale prevents the right-to-left flow of venous blood and thus improves the oxygenation of systemic arterial blood.

3. Changes in pulmonary circulation

Rapid fall in pulmonary artery pressure, which occurs following inflation of previously collapsed lungs (as described above), is accompanied with a six to ten fold increase in pulmonary blood flow. In the fetus the pulmonary arterial pressure is slightly higher than that in the aorta and most of the output from the right heart passes through the ductus arteriosus to the aorta. After birth the position is reversed, for the aortic pressure rises and pulmonary artery pressure falls so that blood flow through the ductus arteriosus, which remains partially open for many hours after birth, occurs from the aorta to the pulmonary trunk, i.e. in the reverse direction to that which takes place in the fetus (Fig. 12.3-4A, B).

4. Closure of ductus arteriosus

The ductus arteriosus is almost as large as ascending aorta of the mature fetus and has a thick smooth muscle wall. The closure of ductus occurs in three steps:

- Vasoconstriction with partial patency,
- Complete functional closure and
- Permanent sealing with fibrous tissue.

(i) Vasoconstriction of ductus arteriosus occurs rapidly during the first few hours after birth but final functional closure takes place gradually over the next 1–3 days. *Factors responsible for constriction of ductus* are:

- Rise in pO_2 of neonatal blood.
- Vasoconstrictor effect of catecholamines released due to stress of birth.
- Fall in levels of prostaglandins, PGE and PGF_2, which help in keeping the ductus arteriosus patent during fetal life because of their vasodilator effect.

(ii) Functional closure of the ductus with complete muscular contraction occurs gradually over the next 1–8 days.

(iii) Permanent sealing of the ductus lumen occurs 2–3 weeks after birth by replacement of musculature with fibrous tissue.

5. Changes in cardiac muscle

During fetal life, the two ventricles have almost similar thickness, because the pressure in the pulmonary artery is slightly greater than the aortic pressure. However, after birth, the left ventricular wall rapidly grows thicker as systemic arterial pressure rises; and the right ventricular wall becomes thinner than the left, as the pulmonary artery pressure falls after birth. The increase in left ventricular size after birth occurs due to an increase in length and thickness of the individual fibres.

STATUS OF CARDIOVASCULAR SYSTEM AFTER BIRTH

1. Heart rate. Immediately after birth, the heart rate of an infant is approximately 140 beats/min. During infancy and childhood heart rate decreases gradually and the adult values are reached only at puberty.

Sinus arrhythmia, i.e. variation in heart rate during two phases of respiration can be observed in infants and children even during normal quiet breathing. In adults, it is observed only during deep breathing.

2. Blood pressure. At birth, the mean arterial blood pressure is approximately 80 mmHg. During the next few hours, it declines to about 65 mmHg. Thereafter, the blood pressure gradually increases throughout infancy and childhood to reach the adult values by the end of pubertal growth.

3. Blood volume at birth is 300 mL or (90 mL/kg).

4. Cardiac output at birth is about 550 mL/min (which is two times as much in relation to body weight as in the adult).

5. Electrocardiography (ECG). At birth, the ECG record shows all the normal waves; however, the right ventricular preponderance is indicated by the mild right axis deviation.

After a few months of post-natal life, the right axis deviation is no more evident. Left ventricular preponderance, indicated by mild left axis deviation, is established by the age of 6–8 years.

CONGENITAL HEART DISEASES

Congenital heart diseases occur either due to defective development of the embryonic heart or due to defect in the closure of three channels of communication (ductus venosus, ductus arteriosus, and foramen ovale) after birth. A few common congenital heart diseases are:

1. Atrial septal defect results from failure of foramen ovale to close. A very small symptomless defect is not uncommon. A large defect causes problem due to shunting of a large volume of blood from left atrium to right atrium.

2. Ventricular septal defect (VSD) occurs due to defective development of interventricular septum.

3. Patent ductus arteriosus. It is also an acyanotic congenital defect with left to right (L → R) shunt, as the blood flows from a high pressure vessel (aorta) to a low pressure vessel (pulmonary artery).

4. Tetralogy of Fallot. It occurs due to defective development of embryonic heart and, as the name indicates, consists of four components:

- Ventricular septal defect,
- Overriding of the aorta at the level of VSD,
- Pulmonary stenosis (subvalvular) and
- Right ventricular hypertrophy.

RESPIRATORY PHYSIOLOGY

FETAL RESPIRATION

Placenta as lung

As described earlier (see page 668), the placenta acts as the site of gas exchange for the fetus before birth.

Oxygen and carbon dioxide transport in the fetus

See page 668.

Fetal breathing movements

Though gaseous exchange does not occur in the fetal lung, the breathing movements do occur during the fetal life. Since the lungs are filled with fluid with a viscosity and density many times that of air, the breathing movements lead to only small alterations in the pulmonary volume. Thus, there is little mixing of amniotic fluid and lung fluids. It has been suggested that the purpose of these breathing movements in utero is to exercise and train the respiratory muscles for their function after birth. Ultrasound scanning technique reveals fetal breathing movements as early as 11 weeks gestation. Initially, irregular, they gradually become more regular.

Fetal pulmonary blood flow and peripheral chemoreceptors

Pulmonary blood flow, during fetal life, is just a fraction of the right ventricular outflow because, due to high pulmonary vascular resistance, most of the right ventricular output is diverted to the aorta through the ductus arteriosus.

Peripheral chemoreceptors are fully developed in a fetus before birth. However, due to some unexplained mechanisms, there is no chemoreceptor discharge in spite of very low pO_2 in the fetal blood.

Pulmonary surfactant (See page 306)

RESPIRATORY ADJUSTMENTS AT BIRTH

Birth is the most traumatic event that the respiratory system must withstand during the entire life span of an individual. It involves sudden transfer from a situation in which no breathing effort is necessary to one in which continual breathing effort is indispensable.

Initiation of breathing

The essential changes which occur while shifting from perinatal respiration to post-natal respiration are summarized:

Pulmonary fluid, which fills the airway in the fetus, keeps the respiratory system at approximately functional residual capacity (FRC). Hence, a major requirement after birth is speedy replacement of fluid by the air so that the respiratory movements can be easier and more useful in terms of gas exchange. From the lungs, the fluid is removed by following forces:

- Part of this vital task is accomplished when thorax is squeezed while passing through birth passages during delivery and
- Part of the fluid is absorbed into the pulmonary capillaries and lymphatics after birth, and some fluid is removed by evaporation.

Alveolar epithelium, which has got a fluid secretory function in the pre-natal period, changes its function to fluid reabsorption after birth.

Factors which stimulate first breath after the birth have been an interesting area of research and speculation, probably, these may be:

- Squeezing of thorax during birth,
- Lower temperature outside as compared to inside the uterus,

- Sound, light, gravity and tactile and painful stimuli and
- Fall in arterial pO_2 and rise in pCO_2 (asphyxia) due to suspension of fetal respiration for a small period during birth and mainly following arrest of umbilical flow (see page 973) is the main stimulant for initiation of breathing.

All the above factors are mediated through nerves, so these may be depressed if the mother receives general anaesthesia during labour, as the fetus is also anaesthetized.

Intrapleural pressure generated during first few breaths is very high, about 60 mmHg, since the lungs are partly fluid filled. The intense efforts required to expand the lungs slowly decrease and normalize in 2 weeks. During this period, *surfactant* is also produced lowering the surface tension which stabilizes the alveoli and allows the lung volume to increase steadily during first few weeks of life.

Effects of initial breathing. After the first gasping breath, normal breathing pattern is gradually established within hours of the breathing. The neonate recovers from hypoxia, hypercapnia and acidosis present throughout fetal life and aggravated markedly by the process of birth and subsequent ligation of the umbilical cord.

Neonatal resistance to hypoxia. It may be mentioned that even when the breathing is not initiated up to 10 min after birth, the neonate shows a normal post-natal growth and development. Resistance to such a prolonged and severe hypoxia is a peculiar feature of neonatal physiology and does not occur in the later life.

Neonatal respiration and haematological changes. Haematological changes also take place after birth, because the adaptations in the form of high haemoglobin level and fetal type of haemoglobin, which is necessary to cope with the hypoxic environment in uterus (pO_2 of oxygenated fetal blood is 30 mm Hg and that of mother's blood in placenta is 50 mm Hg) are no more required after birth. The haemoglobin level falls by haemolysis outpacing erythropoiesis and the haemoglobin type changes to the adult variety soon after birth (see page 109).

STATUS OF RESPIRATORY SYSTEM AFTER BIRTH

1. Number of alveoli in the lungs. About 20 million alveoli are present in the lungs at birth, and their number increases to the adult value of 300 million alveoli by the age of 8–10 years. After this age till the end of pubertal growth, the lungs increase in size because of the increase in the size of alveoli rather than increase in their number.

2. Respiratory rate of a newborn is very high (30–60/min) and gradually decreases during infancy and childhood. The adult values of respiratory rate are reached at about 10 years of age.

3. Tidal air volume at birth in a premature infant is 10–12 mL, while in a full-term newborn is 16–18 mL and keeps on increasing throughout childhood, till the pubertal growth is completed.

4. Vital capacity also continues to increase throughout childhood. It reaches to about 3.3 L at the age of 15 years.

5. Minute ventilation at birth is about 140 mL/min. This is two times as great as in relation to body weight as that of an adult.

6. Functional residual capacity in a newborn is only half that of an adult in relation to body weight.

APPLIED ASPECTS

Respiratory distress syndrome (See page 307)

BLOOD AND IMMUNE MECHANISMS

ERYTHROPOIESIS, LEUCOPOIESIS AND THROMBOPOIESIS

For details see page 104.

FETAL HAEMOGLOBIN (HBF) AND ADULT HAEMOGLOBIN (HBA)

Characteristic features of HbF and HbA are described on page 111. The most salient point to be noted is that HbF has higher affinity for oxygen than the HbA. Up to 32 weeks of gestation, fetus RBCs contains only HbF; after which HbA begins to be synthesized in low concentration. For details see page 109.

CHARACTERISTICS OF BLOOD IN NEWBORN

1. Haemoglobin concentration is high (16–18 g/dL). It is related to low pO_2 of the fetal blood. Changes in Hb seen with age are depicted in Table 12.3-2.

Table 12.3-2 Hb concentration, RBC count and WBC count in relation to age

Age	Hb concentration (g/dL)	RBC count (million/mm^3)	TLC (per mm^3)	DLC	
				Poly (%)	Lymphocyte (%)
1 day	18	6.0	20,000	70	20
1 month	16	4.7	12,000	30	60
3 month	10	4.0	11,000	35	55
1 year	12.5	4.5	10,000	45	50
5 year	13.0	4.7	8000	55	40
10 years	13.0	4.7	8000	60	35

2. RBC count at birth is about 5.5–6.5 million/mm^3, which gradually decreases to 3–4 million/mm^3 by approximately 10–12 weeks of age due to absence of hypoxic stimulus of fetal life and returns to normal within another 2–3 months (Table 12.3-2).

3. WBC count at birth is nearly 20,000/mm^3 with preponderance of neutrophils (70%). It decreases to about 12,000/mm^3 with a marked decrease in neutrophils (30%) by the end of 1 month. From the first month to one year of life lymphocytes predominate. Normal adult values of WBC count are reached at 5–10 years of age (Table 12.3-2).

4. Coagulation factors. *At birth*, though all the clotting factors and fibrinolysins are present in the plasma, but the levels of prothrombin factor V, VII, and X are subnormal. *In the first week* of post-natal life, the levels of these clotting factors may decrease to a level to produce the so-called *haemorrhagic diseases of newborn*. This occurs primarily because of the deficiency of vitamin K and consequently decreased synthesis of vitamin K-dependent clotting factors by immature hepatocytes. That is why to prevent this disease, injection of vitamin K is given immediately after birth as prophylaxis.

At about 2–12 months of age, the adult values of all the clotting factors are reached.

5. Immunologic functions. The newborn is competent to produce both T lymphocyte as well as B lymphocyte-mediated response.

- *IgG levels*. During intrauterine life, IgG levels are actively transferred from the maternal blood to the fetus. Consequently, at birth the newborn's plasma level of IgG is even higher than the maternal plasma level reach a level of 400–500 mg/dL within 1 week as compared to over 1200 mg/dL at birth. At about 4 months of age, the infant's immunological apparatus starts manufacturing IgG and adult levels are reached by the age of 8 months.
- *IgA and IgM* are not transferred to the fetus; consequently, their levels are much lower at birth as compared to mother's level. After birth, IgM begins to rise rapidly, but adult levels are reached by the age of 12–16 years.

PHYSIOLOGICAL ANAEMIA

At birth, the cord blood shows high concentration of erythropoietin, high Hb concentration (18 g/dL), and high RBC count (5.5–6.5 million/mm^3). It is related to low pO_2 of fetal blood.

After birth, the hypoxic stimulus present throughout intrauterine life, disappears and the pO_2 is raised. Consequently, after 1 week of life, there occurs complete cessation of RBC production in the bone marrow. As a result, by 10–12 weeks of life, the Hb concentration may fall to as low as 9–10 g/dL and RBC count may be as low as 3–4 million/mm^3. This produces the so-called *physiological anaemia* of the newborn.

NERVOUS SYSTEM

Neural growth is completed by 4–5 years of age (see page 963):

- At the end of 1 year, 50%,
- At the end of 3 years, 75% and
- At the end of 5 years, 90% of post-natal growth is completed.

Myelination begins during fetal life, but continues for several years postnatally.

Blood–brain barrier is not well developed at birth. This accounts for higher levels of protein, sugar and high white cell count in the cerebrospinal fluid of newborn as compared to that in the later childhood. This also explains the occurrence of kernicterus in an infant at that level of serum bilirubin, which is harmless in an adult.

Visual apparatus becomes fully developed by 5 years of age. Salient points to be noted are:

- *Macula and fovea* of the retina are structurally and functionally differentiated by 4–6 months of age.
- *Colour perception* is fully developed by the age of 4–6 months.
- *Size of eyeball* at birth is small, so the infant is hypermetropic. Adult size is obtained by 5 years, when the child becomes emmetropic.
- *Visual acuity* becomes 6/6 by the age of 5–6 years.

Auditory apparatus is almost fully developed at term. Therefore, response to loud noise can be elicited just after birth.

Taste sensation is present at birth, but becomes sharp by the age of 2–3 months.

Visceral reflexes involved in swallowing, micturition, defaecation, sneezing and coughing are fully developed at birth.

Deep reflexes. Most of the tendon jerks can be elicited at birth except the ankle jerk. The Babinski's sign appears only a few weeks after birth and can be elicited during the next 1–2 years.

GASTROINTESTINAL PHYSIOLOGY

GIT: DURING FETAL LIFE

Placental transfer of nutrients: placenta as gut

As mentioned earlier, the transfer of nutrients to the fetus occur from the maternal blood through placenta (see Fig. 12.3-2D and page 667).

During fetal life, constant and plentiful supply of glucose reaches from the mother.

Fetal gastrointestinal tract

After 20 weeks, fetus ingests and absorbs large quantities of amniotic fluid.

At 24 weeks, fetal GIT functions approach that of a normal newborn infant. Small quantities of *meconium* are continuously formed in the GIT and excreted from the bowels into the amniotic fluid. The meconium consists of unabsorbed residue of amniotic fluid and secretions from the GIT mucosa and glands.

GIT: AFTER BIRTH

The fetus gets the nutrients supply from the maternal blood through the placenta, which is suddenly cut off after ligation of the cord. Since the newborn has to produce its own nutrient's supply, so the gastrointestinal function is fairly well developed at birth.

Liver functions

The newborn has to fulfill the caloric requirements from carbohydrate-poor, fat-rich milk diet, a number of adaptive changes in the hepatic enzymes help in this transition.

Liver functions are quite deficient during first few days, as the liver of newborn:

- Poorly conjugates bilirubin causing its less excretion,
- Poorly performs gluconeogenic function decreasing blood glucose level to 30–40 mg/dL,
- Inadequately synthesizes plasma proteins producing hypoproteinaemia and
- Inadequately synthesizes clotting factors producing even haemorrhagic disease of newborn, sometimes, especially in the premature babies.

Fetal liver glycogen and blood glucose. During late fetal life, the activity of glycogen synthetase in the fetal liver increases. As a result, fetal liver is very rich in glycogen at birth. After ligation of the cord, the blood sugar level of the infant begins to fall and may reach 60 mg/dL within 1 h. The problem of neonatal hypoglycaemia is more pronounced in premature babies.

Hepatic bilirubin excretion. *During the fetal life*, the RBCs have a life span of approximately 70 days, and thus about 1.4% of circulating red cell mass is destroyed every day producing a large amount of bilirubin. *After birth*, in the early post-natal life the hepatic conjugation enzyme UDP-glucuronyl transferase is not fully active, so conjugation with glucuronic acid is poor and so is the bilirubin excretion.

This explains the occurrence of physiological jaundice in the first week of post-natal life. Further the risk of neonatal jaundice is more in premature neonates and those suffering from Rh-incompatibility. Such infants are under the risk of developing kernicterus, since the blood–brain barrier is not well developed at birth (see above in nervous system).

RENAL PHYSIOLOGY, AND FLUID AND ACID–BASE BALANCE

Anatomically, the development of nephrons is fairly complete at birth and no new nephrons are formed in the post-natal life. However, functionally the kidneys in infancy are immature and can manage to maintain normal blood chemistry as long as there is no homeostatic disturbance. In other words, there is very little margin of safety, hence important problems of infancy are dehydration, overhydration and acidosis.

TEMPERATURE REGULATION IN NEWBORN AND INFANTS

The temperature regulation mechanisms, as described in Chapter 12.1 on 'Physiology of Body Temperature Regulation', are operative in newborns and even in premature babies. The salient points which need special mention for temperature regulation in newborns and infants are:

1. Maintenance of body temperature presents greater problems in newborns and infants because of following reasons:

- *Surface area* in relation to the body weight is greater than the adults.
- *Basal metabolic rate*, throughout the childhood, is relatively higher than the adult values and
- *Sweating mechanism* is not fully developed during infancy.

2. Thermoneutral zone (see page 957), for naked newborn infants (32–34°C in term infants and 35°C in premature infants) is higher than the naked adult (26–28°C). The fact underlines two important implications:

- Infants have poor tolerance to cold and
- While adjusting the temperature of the incubator for premature babies and of the labour room in general, the thermoneutral zone of the infant should be kept in mind.

3. Non-shivering or chemical thermogenesis is the most effective mechanism against cold in infants. Presence of brown adipose tissue in infants accounts for this fact (for details see page 955).

SEXUAL GROWTH AND DEVELOPMENT

See page 623.

Chapter 12.4

Geriatric Physiology

INTRODUCTION

Ageing

Ageing is a natural process. No one knows when old age begins. The biological age of the person is not identical with his chronological age. While ageing merely stands for growing old, senescence is an expression used for the deterioration in the vitality or the lowering of the biological efficiency that accompanies ageing. From a physiological standpoint, human ageing is characterized by a progressive constriction of the homeostatic reserve of every organ system. This decline, often referred to as homeostenosis, is gradual and progressive, although the rate and extent of decline vary.

The decline of each organ system appears to occur independently of changes in other organ systems and is influenced by diet, environment and personal habits as well as by genetic factors.

In other words, ageing can be defined as the time-related deterioration of the physiological functions necessary for survival and fertility. The science of ageing is often referred to as *gerontology*. The scientists studying the science of ageing are known as *gerontologists*, and the branch of medicine dealing with the problems of ageing is called *geriatric medicine*.

AGE-RELATED CHANGES IN DIFFERENT ORGAN SYSTEMS

I. CARDIOVASCULAR CHANGES

Changes in heart

1. Myocardium may show following changes:
- Deposition of yellow brown lipofuscin pigment.
- Degenerative changes in the myofibrils and mitochondria.
- Fibrotic lesions and sometimes amyloid deposits.
- Capillary density may be decreased.

2. Valves show thickening and structural changes making the:
- Aortic valve somewhat stenotic and
- Mitral valve slightly incompetent.

3. Functional changes in heart of elderly include:
- *Heart rate* in resting conditions is unchanged. But the maximum heart rate during exercise declines.
- *Maximum cardiac output* in response to exercise is decreased at the rate of 1% per year after the age of 40 years.

4. Sinoatrial (SA) node automaticity and baroreceptor sensitivity is decreased with age. This leads to impaired blood pressure response to standing and volume depletion.

5. Electrocardiogram (ECG) does not show any significant change with age. Therefore, any significant change in ECG of an elderly individual should be considered pathological.

Changes in blood vessels and blood pressure

Blood vessels show a gradual decrease in the number of elastic fibres and a progressive change in the characteristics of elastic tissue. There occurs deposition of calcium salts in the elastic and muscular type of arteries as well as deposition of more collagen fibres resulting in a decrease in the distensibility of the blood vessels.

Blood pressure. Systolic blood pressure is raised because of loss of elasticity in the aorta and its major branches, but there is little change in the diastolic blood pressure, resulting in widening of pulse pressure.

Blood flow to the various organs, such as heart, brain and especially kidney is decreased.

II. CHANGES IN RESPIRATORY SYSTEM

Structural changes in lungs. Alveoli become flatter and shallow, while alveolar ducts enlarge. Number of alveoli declines gradually due to the progressive loss of interalveolar septa.

Pulmonary compliance is increased due to decrease in elasticity of the lungs.

Compliance of thoracic cage and mobility of the ribs are decreased due to calcification of costal cartilages.

Pulmonary blood vessels show age-related increase in wall thickness.

Functional changes occurring as a result of above-said changes in the lungs and thoracic cage are:

- *Functional residual capacity* of the lungs is increased by 50%.
- *Residual volume* is increased by 100%.
- *Vital capacity*, forced expiratory volume in first second (FEV_1), maximum breathing capacity (MBC) and diffusion capacity for oxygen are significantly decreased.
- *Respiratory response to hypoxia and hypercapnia* is sluggish in the elderly.
- *Airways become more susceptible to collapse*, especially during expiration, because of reduced elastic recoil of thoracic cage. The collapse is more likely during exercise because of the high expiratory flow rate.
- *Arterial pCO_2 and pO_2.* The arterial pCO_2 is not changed, but arterial pO_2 is decreased by 10–15% due to an increase in the physiological dead space. But it has no serious detrimental effect on the body.
- *Impairment of bronchiolar escalator function*, which occurs in old age, causes more serious problem, especially in smokers.

To summarize, the respiratory functions of elderly show an overall impairment of ventilation, diffusion, ventilation/perfusion mismatch as well as regulation.

III. GASTROINTESTINAL TRACT CHANGES

Age-related changes noted in relation to gastrointestinal tract include:

1. Diminution of masticatory efficiency occurs due to teeth problems. With advancing age, teeth show *attrition* due to loss of first enamel and then even dentine and cement. In addition, several teeth are lost as a result of caries or periodontal diseases.

2. Difficulty in swallowing (dysphagia) may occur in extreme old age, because of frequent weakness of pharyngeal musculature and abnormal relaxation of cricopharyngeal muscle. Disordered oesophageal motility, especially at the lower end of oesophagus, may further compound this problem.

3. Reduction in gastric secretion leading to *achlorhydria* is seen in 25% of the individuals above 60 years of age. It results from age-related mucosal atrophy. Decreased gastric acidity causes decreased Ca^{2+} absorption from empty stomach. Achlorhydria also results in deficiency of iron and vitamin B_{12}.

The secretion of pancreatic amylase is not affected in old age. Thus, as a whole digestion and absorption of foodstuffs does not seem to be affected in old age except for the deficiency of iron and vitamin B_{12}.

4. Age-related changes in small intestine include reduction in villus height and reduction in lactase activity in the brush border. These changes decrease absorptive capacity in the elderly. However, there is no marked decrease in the digestive or absorptive processes. Therefore, nutritional deficiency in an elderly individual cannot be attributed to malabsorption and actually reflects deficient intake of the nutrients.

5. Changes in liver include decrease in number but increase in size of hepatocytes and an increase in the fibrous tissue. But because of large reserves, the hepatic function is maintained at normal level even after 70 years of age. However, synthetic functions of liver, such as protein synthesis and microsomal mixed function, oxidase activity required for hepatic metabolism of drugs and steroids are reduced in old age. Consequently, all liver function tests show normal results except for a decrease in albumin/globulin ratio. Pigment excretion also remains normal.

6. Colon motility may be decreased in extreme old age resulting in constipation.

IV. RENAL AND GENITOURINARY CHANGES

1. Kidneys show progressive reduction in weight. Functional renal changes include:

- *Decreased glomerular filtration rate* occurs in elderly because of 30–40% decrease in the number of renal glomeruli by the age of 80 years. It leads to impaired excretion of certain drugs, which may produce toxicity at doses well tolerated in younger individuals.
- *Decrease in tubular function*, both secretory and absorptive activity, leads to decreased urinary concentration and dilution abilities. This leads to delayed response to salt or fluid restriction/overload. There occurs nocturia. Maximum urinary osmolality is about 750 mOsm/kg (specific gravity 1.0) at the age of 80 years.
- *Renal function* becomes borderline. Because of large renal reserves, plasma concentration of creatinine and other nitrogenous waste products are not elevated. However, any type of circulatory stress may precipitate renal failure.

2. Prostate enlargement in elderly males is a quite frequent cause of increased residual urine volume. It may take the form of a disease (urinary incontinence or urinary retention).

3. Vaginal/urethral mucosal atrophy occurring in elderly females leads to dyspareunia and bacteriuria.

V. CHANGES IN ENDOCRINAL SYSTEM

Endocrinal changes have been even implicated as the underlying mechanism of ageing. Age-related decrease in the endocrinal function may occur due to any of the following changes:

- Decrease in the plasma concentration of hormone due to decreased production or due to decrease in the concentration of binding proteins involved in the transport of hormone,
- Decreased responsiveness of the target cells,
- Alteration in number or sensitivity of the hormone receptors or
- Diminished response to physiological stimuli for secretion of the hormone.

Age-related changes occurring in endocrinal system

1. Thyroid hormone secretion is definitely decreased. Up to a 50% decrease in the production is noted by the age of 80 years.

2. Impaired glucose homeostasis is frequently seen in old age. This seems to be due to diminished sensitivity of tissues to insulin, as the plasma levels of insulin are unaffected. It has been observed that 2 h after administration of glucose load, plasma glucose is about 30 mg/dL higher at the age of 70 years than in the young adults.

3. Reproductive hormones show most consistent age-related changes. In females, plasma levels of oestrogen and progesterone are decreased after menopause. In males, testosterone levels are decreased around the age of 70 years.

4. Anterior pituitary hormones secretion is not decreased.

- Follicle-stimulating hormone and luteinising hormone levels in females are rather increased due to negative feedback effect exerted by the decreased plasma levels of oestrogen and progesterone.
- Gonadotropin levels in males are raised because of negative feedback effect of lowered testosterone levels. The raised gonadotropin level induces an increase in testicular Leydig cell volume. But because of the age-related changes, even the higher Leydig cell volume is not able to achieve the normal testosterone output.

5. Antidiuretic hormone, renin and aldosterone levels are decreased in old age. Changes in these hormones affect the renal functions, especially urinary concentration and dilution mechanism.

6. Vitamin D absorption and activation is decreased with age and contribute to *osteoporosis* occurring in old age.

VI. CHANGES IN BLOOD AND IMMUNE MECHANISMS

1. Blood volume and blood cells. The blood cells (red blood cells, white blood cells and platelets) of elderly individual are not significantly different from those of the young individuals.

2. Haemopoietic marrow reserve is gradually decreased because of replacement with fatty marrow as age advances. Long bones are first affected followed by flat bones and the last to be involved are vertebrae.

3. Anaemia in elderly usually occurs due to deficiency of iron and vitamin B_{12}.

4. Senile purpura occurs due to defect in capillary endothelium. Platelet count and function of blood coagulation usually remain normal in old age.

5. Raised ESR (up to 40 mm in first hour) seen in elderly is related to increased plasma fibrinogen levels.

6. Immunological function is markedly depressed in old age. Both cell mediated immunity and humoral immunity are declined.

- *Reduced immune surveillance* by T lymphocytes, at least in part explains the greater incidence of malignancy in old age.
- *Involution of thymus* is a well known age-related change.
- *T-cell autoreactivity and autoantibody titre* is, however, increased, probably due to diminished tolerance to antigens normally recognized as self.

VII. CHANGES IN MUSCULOSKELETAL SYSTEM

1. Muscular power, characteristically reduces progressively with ageing due to loss of muscle fibres (decreased lean body mass). The loss of muscle fibre is much greater than the number explained by the loss of motor neurons in the central nervous system (CNS). Muscle fibres show deposition of lipofuscin and many other degenerative changes with age.

2. Muscle twitch reveals a prolongation of latency, contraction period and relaxation period. The maximum tension developed in a muscle in elderly is much less than in young adult.

3. Osteoarthritis, i.e. age-related degenerative changes in the joints start at the age of 40 years and become well marked by the age of 60–70 years.

4. Osteoporosis, i.e. age-related decrease in bone density is a characteristic feature of ageing. It is more marked in postmenopausal women than men of the same age group. Osteoporosis predisposes the elderly to fractures.

5. Changes in stature and posture occur mainly due to changes in the vertebral column, as long bones of the limbs do not show any significant change. Initially, there occurs thinning of the intervertebral disc only, osteoporotic changes cause a decrease in height of individual vertebra after the age of 50–60 years. As a result of these changes, the height decreases progressively from 50 years onward with a prominent change at 70–80 years. Further, kyphosis and slight flexion at the hip and knee makes an aged person look still smaller. A decrease of about 5 cm in height is reported to occur from 20 to 80 years of age, which increases to 10 cm by the age of 90 years.

VIII. CHANGES IN SKIN AND HAIR

1. Wrinkling of skin due to decreased elasticity, increased thinning of epidermis and dermis and decreased subcutaneous fat is the hallmark of ageing.

2. Greying of hair, due to loss of melanin pigment, is universal in ageing.

3. Baldness in males is quite common, though growth of beard is not effected.

4. Loss of axillary and pubic hair in females occurs due to decreased levels of adrenal androgens.

5. Increase in facial hair growth may occur in females due to unopposed action of the residual adrenal androgens in the absence of oestrogens.

6. Sweat glands decrease in size and number, therefore secretion of sweat as well as sebaceous glands is decreased.

IX. CHANGES IN CENTRAL NERVOUS SYSTEM

Both structural and functional changes in CNS are quite common with ageing.

1. Brain atrophy and neuronal loss is the most obvious change. By the age of 70 years, there may be 45% cell loss in cerebral cortex and 25% loss in cerebellum. Atrophy of frontal lobes leading to shrinkage of gyri and enlargement of sulci is quite common.

2. Degenerative changes may occur in substantia nigra and lentiform nucleus. Degeneration in spinal cord occurs to a lesser extent than in the brain.

3. Other histological changes include accumulation of lipofuscin granules in almost all the neurons and glial cells, loss of synapses and gradual loss of dendrites.

4. Cerebral blood flow is decreased by 40% at the age of 70 years and oxygen utilization is reduced by 25%.

5. Functions of neurotransmitters are impaired. Specifically, cholinergic deficit has been demonstrated in the *Alzheimer's disease* (see also page 881) and dopaminergic deficit in Parkinson's disease (see also page 730). Milder forms of cholinergic deficit may be responsible for *senile dementia*, and some degree of dopaminergic deficit may be responsible for the *hypokinesia* of old age. Decrease in catecholamine synthesis may be responsible for the depression in old age.

6. Reflexes tend to sluggish or even absent. The ankle jerk is lost in most of the elderly individuals. Decreased righting reflexes result in an increased body sway in old age.

7. Sleep changes occur in the form of decrease in stage 3 and 4 of non-REM sleep, while the total duration of sleep may not be decreased much. Very old people may not go into stage 4 of sleep at all and have early awakening. Numerous brief arousals occur and account for feeling of no sleep or insomnia.

X. CHANGES IN AUTONOMIC NERVOUS SYSTEM

There is also an age-related impairment of autonomic nervous system (ANS) function associated with an increased sensitivity to humoral factors. Common manifestations of decreased ANS function are:

1. Impaired temperature regulation in elderly individuals occurs because of the fact that exposure to moderate cold or hot environment does not produce expected vasoconstriction and shivering, or vasodilatation and sweating, respectively. Consequently, elderly are more prone to get hypothermia on exposure to cold and hyperthermia on exposure to heat.

2. Postural hypotension is of frequent occurrence in elderly. It is known to occur because of partial failure of baroreceptor mechanism.

XI. CHANGES IN SPECIAL SENSES

(i) Age-related ocular changes

1. Presbyopia refers to an age-related physiological insufficiency of accommodation leading to failing vision for near or progressive increase in near point. For details see page 897.

2. Age-related cataract, or the senile cataract, refers to opacification of the crystalline lens leading to progressively decreasing vision. Its age of onset and maturation is influenced by the hereditary and environmental factors.

3. Age-related corneal degeneration manifests as a ring-shaped whitish opacity near the limbus, which is called *arcus senilis*.

4. Age-related macular degeneration (ARMD) is a cause of irreversible blindness in many elderly individuals after the age of 70 years.

5. Dry eye may occur in elderly individuals, more so in post-menopausal females, because of age-related decrease in tear secretion.

(ii) Age-related changes in ears

1. Presbyacusia, i.e. age-related impairment of hearing, especially for higher frequencies occurring due to degenerative changes in the organ of Corti (hair cells), ganglion cells as well as of temporal cortex, is not uncommon. Other factors contributing to presbyacusia are loss of elasticity of tympanic membrane and basilar membrane, loss of neurons in the cochlea and atrophy of stria vascularis. The impaired sensitivity often leads to difficulty in understanding speech and disturbances of localization of sounds.

2. Otosclerosis, characterized by age-related decrease in motility of middle ear ossicles, is another cause of deafness in old age.

3. Impairment of postural reflexes may occur due to age-related degenerative changes in hair cells of the crista ampullaris and decreased endolymph production because of atrophy of the stria vascularis.

(iii) Age-related changes in taste and smell

1. Impairment in taste sensation in elderly is attributed to decrease in number of taste buds from an average of 250 buds per papilla in childhood to about 90 buds per papilla by the age of 80 years.

2. Impairment in sensation of smell in elderly is attributed to decrease in smell receptors and partly to loss of neurons in cerebral cortical centres.

THEORIES OF AGEING

Many theories have been put forward to explain the process of ageing, but none is able to explain all the queries. Most of the theories fall in one of the two main groups:

- Genetic theories of ageing and
- Random damage theories.

I. Genetic theories of ageing

These theories consider ageing to be the inevitable result of the genetic programme.

Of the many genetic theories, some important are:

1. Programmed senescence theory. Programmed senescence theory of ageing hold that ageing follows a biological time table, perhaps a continuation of one that regulates childhood growth and development. According to *programmed senescence theory*, ageing is the result of the segmental switching on and off of certain genes; with senescence being defined as the time when age-associated deficits are manifested.

2. Mutation theory. This theory suggests that since animals usually succumb to natural forces long before reaching their maximal life span, ageing might reflect mutations that impair long-term survival. These mutations would accumulate in the genome because there is no selection pressure to delete them.

II. The 'random damage' theories of ageing

All the 'random-damage' theories are based on the possibility that the balance between ongoing damage and repair is disrupted. These theories share the observation that cell and organ repair capacity declines with age.

The 'random-damage' theories include:

1. Free radical theory. Oxidation reactions in the cells are associated with formation of free radicals, such as superoxide and hydroxyl radicals. For free radical scavenging the *antioxidant mechanisms* exist in the body in the form of glutathione, vitamin E, vitamin A and vitamin C. When these antioxidant mechanisms are overwhelmed, there occurs damage by the free radicals. The free radicals, can possibly damage vital macromolecules, such as DNA and proteins, and cause peroxidation of lipids in the membranes around cells and organelles. Although free radical theory is very popular and antioxidants that are being prescribed

over enthusiastically by the physicians, neither has lipid peroxidation been demonstrated at cellular level, nor has the beneficial effect of antioxidants been proven.

2. Cell replication theory. Depending upon the replicating capabilities, the cells in the body can be divided into three categories:

- *Cells which continuously replicate* include blood cells, epidermis cells and gastrointestinal cells.
- *Cells which replicate only under stress* (e.g. injury) include endothelial cells, hepatocytes and fibroblasts and
- *Cells which do not replicate at all* include neurons, myocardial and skeletal muscle cells.

It has been stated that replicating cells have a definite replicating limit. The cell replication theory suggests that ageing may represent a stage of life when replication of cell ceases; i.e. when repair is not capable to cope up with the damage.

Subsequent researches revealed that this replicative senescence was due to arrest of the cell cycles at the G_1/S phase, the point at which DNA synthesis begins. Recently, cell replication has also been linked to the length of *telomeric DNA*. With each cell division, roughly 50 of the total 2000 base pairs of the telomere are lost. Telomeric shortening might thus result in loss of gene accessibility, which is caused by metabolism. Together with cytoplasmic factors mediating arrest of DNA synthesis, telomeric shortening could also limit the cell's ability to divide and thereby replace cells loss to apoptosis.

3. Cross-linking theory. This theory highlights that an accumulation of cross-linked proteins damages cells and tissues, slowing down bodily processes and results in ageing. In a process called *non-enzymatic glycosylation or glycation*, glucose molecules attach themselves to proteins, setting in motion a chain of chemical reactions that ends in the proteins binding together or cross-linking, thus altering their biological and structural roles. The process is slow but increases with time.

Crosslinks, which have been termed *advanced glycosylation end products* (AGEs), seem to toughen tissues and may cause some of the deterioration associated with ageing. Advanced glycosylation end products have been linked to stiffening connective tissue (collagen), hardened arteries, clouded eyes, loss of nerve function and less efficient kidneys.

MODULATING THE PROCESS OF AGEING

The quest for remaining youthful and preventing ageing has lead to many trials on modulating the process of ageing. However, ageing has proved to be an almost inevitable process. The only measures which have shown some progress in this regard are: caloric restriction and exercise.

Caloric restriction. To date the only intervention known to delay ageing and prolong the life span is caloric restriction, which has been proved in experimental animals. Although the underlying mechanism is still not determined, it is specific to caloric restriction rather than to reduction of any dietary compound (e.g. fat intake) or supplementation with vitamins or antioxidants. Further, the effects of caloric restriction in humans are still unknown.

Exercise. There is still no conclusive evidence to document that exercise prevents ageing or not. However, definitely, exercise improves work capacity as assessed from maximum oxygen uptake. Physical exercise also improves cardiac performance and reduces musculoskeletal disability. Physical exercise is also reported to prevent age-related decline in resting metabolic rate.

“This page intentionally left blank"

Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z