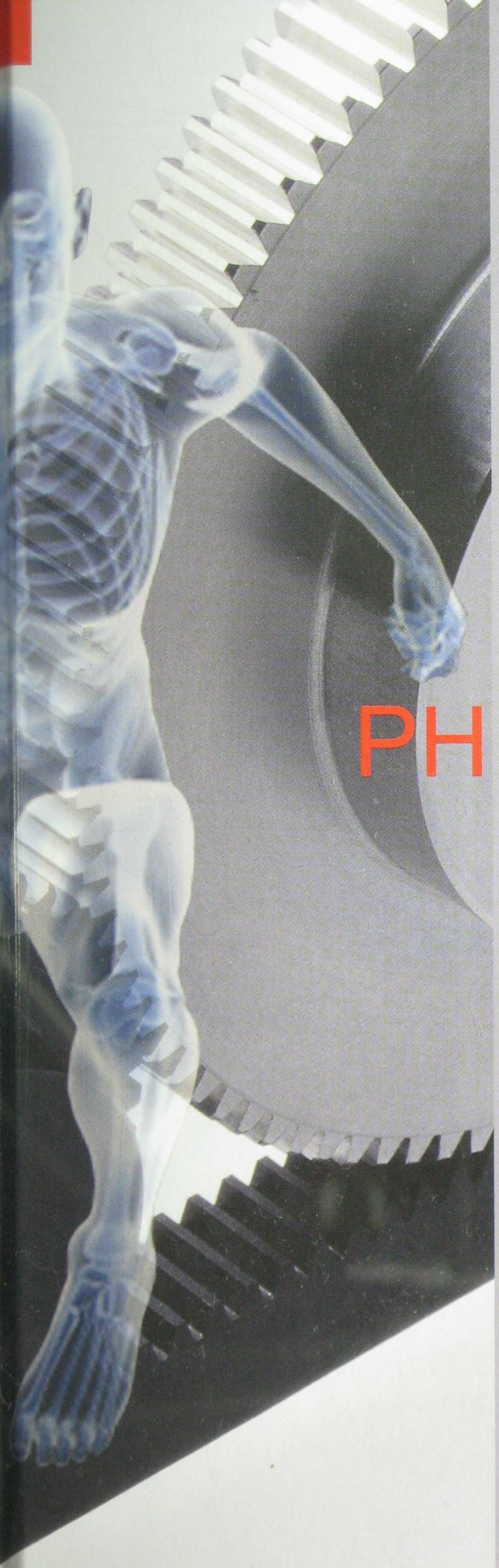


Edited by
Vasyl M. Moroz
Oleksiy A. Shandra

PHYSIOLOGY

NR
PUBLISHERS

PHYSIOLOGY

Edited by
V. M. Moroz
O. A. Shandra

2nd edition

Vinnytsia
Nova Knyha
2016

UDC 28.903я73
BBK 612(075.8)
P57

Recommended by the Ministry of Health of Ukraine as a textbook for students of higher medical educational institutions with the IVth level of accreditation, with English as the language of instruction (report # 2, 26.10.2011)

Authors:

V. M. Moroz, O. A. Shandra, R. S. Vastyanov, M. V. Yoltukhivsky, O. D. Omelchenko

Reviewers:

Karpov L. M. – Head of the Department of Human and Animal Physiology of Odessa National I. I. Mechnikov University, D. Sc. (Biology), Prof.;
Tkachuk S. S. – Head of the Department of Physiology of Bukovinian State Medical University, D. Sc. (Medicine), Prof.;
Rodinsky O. G. – Doctor of Medical Sciences, Professor, Head of Physiology Department, Dnipropetrovsk State Medical Academy.

Phisiology : textbook / edited by V. M. Moroz, O. A. Shandra. — 2nd edition. — Vinnytsia : Nova Knyha, 2016. – 728 pp.
P57
ISBN 978-966-382-572-4

"Physiology" is a textbook on Human Physiology, written in English for best underdstanding of physiology by English-speaking students, contains basic educational materials according to the Physiology Course Program. The edition contains information about all parts of discipline paying attention to basic requirements of the credit module system of teaching in relation to improvement of student self-study. Together with basic mechanisms of organs and organ systems functioning, principles of organism neuro-humoral regulation in normal conditions, adaptative-compensatory mechanisms in stressed conditions or other factors, initial data about functioning failure of each organ are given in the textbook.

Clear, detailed content of each chapter with selection of major details on every theme, understandable definition of main data, definitions and classifications with additional graphical material allow students to easily find out the main aspects of each theme.

The textbook can also be useful to teachers working in educational medical institutions with English-speaking students.

UDC 28.903я73
BBK 612(075.8)

ISBN 978-966-382-572-4

© Authors, 2016
© Nova Knyha, 2016

Contents

PREFACE 6

UNIT I. PHYSIOLOGY OF EXCITATIVE TISSUES (Prepared by O. A. Shandra) 7
Chapter 1. Introduction into physiology. The excitative tissues. Membrane potential and potential of action 7
Chapter 2. Physiology of muscles. Skeletal and smooth muscles 28
Chapter 3. Physiology of synaptic transmission 47

UNIT II. GENERAL PHYSIOLOGY OF CENTRAL NERVOUS SYSTEM (Prepared by R. S. Vastyanov) 54
Chapter 4. General principles of biologic regulation. Neural regulation. The principles of reflectory regulation 54
Chapter 5. Excitation and inhibition in central nervous system. The principles of reflectory activity coordination 66

UNIT III. ROLE OF CNS IN REGULATION OF SOMATIC FUNCTIONS (Prepared by O. A. Shandra & R. S. Vastyanov) 80
Chapter 6. Organization and control of motor movements. Tone of skeletal muscles. The pyramidal and the extrapyramidal systems. Spinal cord, basal ganglia 80
Chapter 7. Forebrain and cerebellar regulation of motor function and organism systemic activity. Physiology of thalamus and brain cortex. Electric activity of CNS and its clinical importance 97

UNIT IV. PHYSIOLOGY OF AUTONOMIC NERVOUS SYSTEM (Prepared by R. S. Vastyanov) 119
Chapter 8. Neural regulation of vegetative functions. Structural-functional peculiarities of autonomic (vegetative) nervous system 119

UNIT V. ENDOCRINE PHYSIOLOGY (Prepared by O. A. Shandra) 134
Chapter 9. General endocrinology. The hypothalamic-pituitary system 134

Chapter 10. Endocrine regulation of physical and mental development 155
Chapter 11. Endocrine regulation of homeostasis and organism nonspecific adaptation 165
Chapter 12. Physiology of male and female sexual glands 215

UNIT VI. BLOOD PHYSIOLOGY (Prepared by R. S. Vastyanov) 250
Chapter 13. General characteristics of blood system. Physiological constants of blood. Physiology of erythrocytes. Hemoglobin. Blood groups 250
Chapter 14. Protective functions of blood. Physiology of leykocytes. Immunity and its types 268
Chapter 15. Physiology of blood coagulation 381

UNIT VII. CARDIOVASCULAR PHYSIOLOGY (Prepared by O. A. Shandra & R. S. Vastyanov) 293
Chapter 16. Physiological properties of cardiac muscle. Mechanical properties of the heart 293
Chapter 17. Electrical displays of cardiac activity 326
Chapter 18. Neurohumoral regulation of cardiac activity 356
Chapter 19. Circulatory system. Hemodynamics. Classification of vessels 377
Chapter 20. Microcirculation and lymph flow. Regional bloodflow 391
Chapter 21. Arterial pressure. Neurohumoral regulation of circulatory system functional activity 404

UNIT VIII. RESPIRATORY PHYSIOLOGY (Prepared by R. S. Vastyanov) 437
Chapter 22. Structural organization of the respiratory system. The mechanism of inspiration and expiration 437
Chapter 23. Gas exchange and gases transportation by blood 462
Chapter 24. Regulation of respiration 480

UNIT IX. GASTROINTESTINAL PHYSIOLOGY (Prepared by R. S. Vastyanov) 491
Chapter 25. Digestive tract functions. Digestion in the oral cavity. Swallowing 491
Chapter 26. Digestion in the stomach 503
Chapter 27. Digestion and absorption in the small intestine 621

Chapter 28. Physiology of pancreatic gland and liver.
Digestive functions of pancreatic juice and bile 541

UNIT X. ENERGY EXCHANGE AND TEMPERATURE PHYSIOLOGY
(Prepared by O. A. Shandra) 551
Chapter 29. Energy exchange. Basal exchange. Nutrition 551
Chapter 30. Temperature and its regulation 574

UNIT XI. PHYSIOLOGY OF KIDNEYS AND ACID-BASE BALANCE
(Prepared by R. S. Vastyanov) 587
Chapter 31. Structural organization of excreting system.
Kidneys and their role in urine formation 587
Chapter 32. Regulation of water-electrolyte balance 605
Chapter 33. Acid-base balance 622

UNIT XII. SENSORY PHYSIOLOGY
(Prepared by V. M. Moroz, V. M. Yoltukhivsky & O. D. Omelchenko) 635
Chapter 34. General properties of sensory systems.
Proprioceptive, taste and olfactory sensory systems 635
Chapter 35. Physiology of vision 658
Chapter 36. Physiology of auditory and vestibular sensory systems 672

UNIT XIII. PHYSIOLOGY OF HUMAN BEHAVIOUR
(Prepared by R. S. Vastyanov) 684
Chapter 37. Unconditioned and conditioned reflectory activity 684
Chapter 38. Physiology of memory 694
Chapter 39. Physiology of sleep 707

Preface

The first edition of the 'PHYSIOLOGY' handbook was written 4 years ago. Over the years the handbook has become popular among English-speaking students studying in our country. We use the book as the source of knowledge in the field of Physiology during the lectures and practicals where students have to evaluate their theoretical and practical skills.

The main purpose of the first edition 'PHYSIOLOGY' was to introduce students into the wide spectrum of physiological processes manifesting in human organism. That's why we allowed ourselves to give readers certain data concerning biological, anatomical, histological and biochemical boxes closely related to the chapters reviewed. We prepared each chapter in such a way to be finalized with the 'bridge' to pathological physiology thus giving an idea what and in which manner has been happening in actual pathological event.

With time we decided to change the approach to teaching physiology to students. Preserving the same purpose in the second edition handbook we come to a decision to outline only physiology study material and concentrate only on the main mechanisms of human organism organs and systems adaptation in case of the environmental and other factors influencing on it. Hence, we eliminate from the text the biological, morphological and biochemical aspects of the chapters.

We are grateful to colleagues and members of the Physiology Departments of Odessa and Vinnitsa National Medical Universities for their assistance in preparing the second edition of 'PHYSIOLOGY' handbook.

Tables, schemes and figures have been borrowed from various educational and electronic editions and sources on physiology and used in the book exclusively with the educational purposes.

Unit I PHYSIOLOGY OF EXCITATIVE TISSUES

Chapter 1. Introduction into physiology. The excitative tissues. Membrane potential and potential of action

Physiology (from the Greek *physis* = nature; *logos* = study) is the study of biological function — of how the body works, from cell to tissue, tissue to organ, organ to system, and of how the organism as a whole accomplishes particular tasks essential for life including the mechanisms of its adaptation to the changing conditions of the environment.

It is a broad field that encompasses many disciplines and has strong roots in physics, chemistry and mathematics. Additionally, this discipline neighbours with the biochemistry, anatomy, histology, pharmacology and other fundamental and clinical disciplines. Therefore, human physiology provides the scientific foundation for the field of medicine and all other professions related to human health and physical performance. The scope of topics included in a human physiology course is therefore wide-ranging, yet each topic must be covered in sufficient detail to provide a firm basis for future expansion and application. A basic understanding of physiological mechanisms can instill a deeper appreciation for the complexity and beauty of the human body and motivate the student to learn still more.

Physiology is considered to be an ancient science. The Greek philosopher *Aristotle* (384–322 B.C.) speculated on the function of the human body. *Galen* (A.D. 130–201) wrote widely on the subject and was considered the supreme authority until the advent of the Renaissance. Physiology became a fully experimental science with the revolutionary work of the English physician *William Harvey* (1578–1657), who demonstrated that the heart pumps blood through a closed system of vessels. However, the father of modern physiology

is the French physiologist *Claude Bernard* (1813–1878), who observed that the *milieu interieur* ("internal environment") remains remarkably constant despite changing conditions in the external environment. In 1932, the American physiologist *Walter Cannon* (1871–1945) coined the term homeostasis to describe this internal constancy. Cannon further suggested that the many mechanisms of physiological regulation have but one purpose — the maintenance of internal constancy.

Physiologists assume that the same chemical and physical laws that apply to the inanimate world govern processes in the body. Because the functions of living systems are carried out by their constituent structures, knowledge of structure from normal anatomy to the molecular level is the tool to a physiology understanding.

The scope of physiology ranges from the activities or functions of individual molecules and cells to the interaction of our bodies with the external world. In recent years, we have seen many advances in our understanding of physiological processes at the molecular and cellular levels. In higher organisms, changes in cell function always occur in the context of a whole organism, and different tissues and organs obviously affect one another. The independent activity of an organism requires the coordination of function at all levels, from molecular and cellular to the organism as a whole. An important part of physiology is understanding how different parts of the body are controlled, how they interact, and how they adapt to changing conditions.

For a person to remain healthy, physiological conditions in the body must be kept at optimal levels and closely regulated. Here we have to introduce one of the important definitions in the branch of physiology — *homeostasis* that means the constant internal environment of the organism (firstly defined by W. Cannon). The term *homeostasis* has traditionally been applied to the internal environment — the extracellular fluid that bathes our tissues — but it can also be applied to conditions within cells.

Most of our present knowledge of human physiology has been gained in the twentieth century. Further, new knowledge is being added at an ever more rapid pace, fueled in more recent decades by the revolutionary growth of molecular genetics and its associated biotechnology, and by the availability of ever more powerful computers and other equipment. With the time, human physiology becomes clinical discipline that makes possible to directly investigate human physiological processes using new technique and highly resolute digital devices.

I. PHYSIOLOGY OF EXCITATION

From the general biology course you know that all living organisms are capable of reacting to changes of the external or internal environment by changing the structure, occurrence, strengthening or weakening vigorous activity. Agencies, which cause these changes, are referred to as stimulants; and the changes in a living organism which are caused by them are referred to as reactions to stimulants. It contains the indissoluble connection of living organisms with the environment.

In high-specialized tissue, such as nervous, muscular and glandular, specific change, typical only to these tissues, occurs under the influence of stimulants and is defined by the term "excitation". Therefore, excitability is the property of living structures to accept changes from the external and internal environment and to change its specific activity. Excitation is a characteristic response of an excitable system to stimulation. Excitability — the ability of living tissue to answer stimulation with active reactions, where the basic component is changing physical, chemical and biochemical properties of a superficial cell membrane, which supply specific sensitivity and reaction to stimulation.

Excitability of various tissues is very different. A measurement of excitability is threshold force of stimulation. Threshold force of stimulation is the least amount of force of stimulation which is capable of causing propagated excitation. The threshold of stimulation and excitability is in inverse proportion, that is, the less the threshold is, the higher the excitability and vice versa.

Except for the threshold force of stimulation, the excitability of tissues is characterized by *lability* and *chronaxy*.

The level of excitability of tissues is changeable.

Excitability changes as a result of previous stimuli. In this case, absolute and relative refractory, a phase exaltation and an additional long-term period of decrease in excitability can be distinguished.

Absolute refractory for skeletal muscular fiber is 2.5–4 milliseconds, for thick nervous fibers — 0.4 milliseconds, for the body of a nervous cell — 2.5–4 milliseconds, for cardiac muscles — a few hundreds of milliseconds. These cyclic changes of excitability result in that the frequency of excitation cannot increase indefinitely. At high frequencies of stimulation there comes a change (transformation) in rhythm of increasing excitation. With long-term stimulation, excitability decreases. Such decrease in excitability is referred to as adaptation. Time, during which there is adaptation, is different for different

excitable tissues. So, receptors of a muscular shaft adapt in several minutes, tactile skin receptors — in several seconds.

We can distinguish two kinds of excitation: local, the local answer, and propagated action potential. Local excitation originates with an action that has a small force behind the subthreshold stimulation (approximately 50–90 % threshold forces). At a certain force of stimulation, local excitation reaches a critical level of depolarization, after which there is a propagated excitation which predetermines the external reaction of an organism. Local excitation is observed in receptors of post-synapse membrane. The characteristic of local excitation: in its development is in the absent latent period. It has a graduated character — the more stimuli force, the more local excitation is expressed, i.e. it submits to the law of "gradation". The absence of refractory is characterized by increase in excitabilities, possesses the ability to summation. Excitation fades several millimeters from the place of excitation.

The propagated excitation has the opposite characteristic. It has the latent period which is necessary so that local graduated excitation can reach the necessary threshold level, provided by the occurrence of propagated excitation. It has a stimulation threshold, refractory, submits to the law "all or none ", does not possess the ability to summation. Excitation, in this case, covers the entire stimulated unit and is distributed without an indulgence because of stimulating actions of local electrical currents.

II. THE BIOELECTRIC PHENOMENA DURING EXCITATION

Among the various functions of cellular membranes one of the basic is the generation and conduction of bioelectric activity which is defined by the difference of electro-potentials between two points of a living tissue. The question of the origin and conduction of the biopotentials has interested many scientists. The big interest to these phenomena of nature was founded partially by the well-known physicist Faraday, he wrote: "No matter how sensitive the electric phenomena would be, present in inorganic materials, they do not compare with what is connected with the activity of the nervous system and electric processes."

The basic kinds of bioelectric potentials are:

1. Membrane potential / resting potential (RMP)
2. Action potential (AP)

All living cells have a different electric charge on the external and internal surfaces of the membrane. Thus, the membrane is polarized.

In a condition of rest, the external surface has positive charge and the internal a negative charge. This potential difference between the internal and external surfaces of a membrane in a condition of rest is called the resting membrane potential (RMP).

It is possible to prove the presence of RMP by injecting in a nervous fiber or a cell a thin glass microelectrode filled with KCl solution. The diameter of the tip of such a microelectrode is very thin and constitutes 0.5–1 micrometers. Another inert electrode is positioned in the extracellular environment. If the membrane is not yet punctured, the oscilloscope does not fix a current, but as soon as the tip of a microelectrode passes through the membrane, the device will certify the occurrence of an electro current.

A big contribution to the studying of RMP was introduced by Ukrainian physiologists led by academician P. G. Kostyuk which for the first time in the USSR applied microelectrode techniques.

Characteristic of RMP are such parameters as size (it is expressed in mV), polarity and duration. RMP of a nervous fiber, for example, equals — 90 mV, in neurons — 70 mV. RMP exists while a cell lives until the moment of stimulation.

According to the modern theory of Hodjkin, Huxley and Katsa, the reason of occurrence of RMP are irregular distributions of ions from the outside and within the cell membranes. The concentration of K^+ ions inside nervous and muscular cells is 20–100 times more than extracellular concentration.

The concentration of Na^+ ions, on the contrary, from the outside is 5–15 times more than inside the cell. Endocellular concentration of Cl^- ions is 20–100 times lower than extracellular.

Ionic asymmetry results in the occurrence between the cell cytoplasm and the environment of concentrated ionic gradients which have a direction: the potassium gradient is directed from within to the outside, and sodium and chlorine — from the outside deep inside. In rest, membrane pores are so narrow that only fine K^+ ions can pass through them, therefore membranes more permeable for K^+, than for Na^+. Any time if a K^+ ion encounters a pore aperture, they diffuse through the membranes. Because from the internal side of a membrane K^+ ions are much greater, such interferences will pass more frequently than from the outside so that more ions will pass from within to the outside.

So, the K^+ ion provides a charge for the external membrane surface, according to the Herman's first law of electrophysiology, in rest is positive. That

is how the potential difference between external and internal surfaces of a membrane, RMP, is created. Na^+ makes only a small contribution to the RMP because membrane permeability to Na^+ is very low compared to that of K^+. The contribution of Cl^- ions need not be considered because the resting membrane potential in the muscle cell is the same as the equilibrium potential for Cl^-. Therefore, there is no net movement of chloride ions.

With excitation RMP decreases. This reduction of RMP with the excitation of tissue, from which it is made, is referred to as negative fluctuation of resting potential, or action potential which affects AP. It is possible to explain such reduction of RMP from the formula: force of a current is directly proportional to the potential difference and inversely proportional to resistance; J = E / R. From this, force of a current can decrease due to the reduction of potential difference, or due to the increase of resistance. Resistance of a tissue during excitation not only does not increase, but sometimes also decreases. Significant reduction of current force moves due to the reduction of potential difference. The excited section of tissue is electronegative in relation to the electropositive section not excited at the present moment. It makes the 3rd law of electrophysiology. (the 2nd law: the surface of the damaged tissue has a negative charge).

For the registration of AP 2 outputs are used: single-phase and double-phase. Single-phase output — one electrode is placed on the cell's surface, and the second — inside a cell or on its damaged place. Double-phase output — both electrodes are placed on the cell's surface. According to the arrangement of electrodes AP is single-phase or double-phase. After applying stimulation and the occurrence of excitation, negativity arises, first under one electrode and then as excitation is propagated, under the second electrode. This is registered as double-phase AP.

The AP size of nervous cells of mammals is +30 — +40 mV or 100 –110 mV, skeletal and cardiac muscular fibers — 110–120 mV.

The duration of AP of nervous cells is 0.5–2 milliseconds, of skeletal muscles 5–10 milliseconds, in the myocardium — more than 200 milliseconds.

The nature of AP. In what way does resting potential, usually held at a constant level, rises to such a measure so that AP occurs?

AP always occurs with the depolarization of membranes, that is when the cellular membrane looses its normal charge, approximately -50 mV.

The level of potential, at which depolarization gives rise to action potential, is referred to as "a critical level of depolarization."

AP represents a consecutive depolarization and repolarization of membranes — the auto-regenerative process, constant for each cell, is started as

soon as the depolarization level of a membrane passes to threshold potential. Each type of cell has a constant characteristic for the given kind of time course for AP. It practically does not depend on the frequency of cell excitation. Since the form for AP is constant, it is said that excitation occurs according to the law of "all or none".

AP mechanism.

1. The occurrence of AP is connected to short-term increasing in membrane permeability for Na^+ ions with the subsequent increase in diffusion of these ions against the concentration gradient inside the cell. So, if MRP is caused by K^+ ions, then AP is caused by Na^+ ions. Therefore, MRP refers to potassium and AP refers to sodium ions. Diffusion of sodium ions deep in the cells results in reduction of MRP. Reduction of MRP to some critical level (usually 10–30 mV) will cause positive feedback, that will result in the further increase in penetration of membranes for Na^+ ions, which is accompanied by avalanche-like increased diffusion of sodium deep into the cells. The amplified stream of positively charged Na^+ ions deep into cells causes first the disappearance of excessive negative charge, and then results in the recharging of the membranes: occurs reverse MRP — the external membrane surface becomes negative in relation to the internal. The *depolarization phase* occurs, which results in the occurrence of a short-term ascending current of Na^+ directed deep into a cell, which graphically is responsible for the time of ascending or rising phase of AP (Fig. 1–1). The important role that Na^+ plays in the origin of AP proves experimental data to be true that AP can be generated only with high extracellular Na^+ concentrations. With the lack of extracellular Na^+, the Na^+ input current cannot increase independently by what amount K^+ grows (Na^+ gradient, and so depolarization phase of AP cannot occur).
2. The second stage of AP occurrence is connected to a longer period of increased penetration of membranes for K^+ ions and intensified diffusion of these ions from the cell to the outside. The increase of a K^+ ionic flow results in the reduction of penetration of membranes for Na^+ ions and a reduction of their flow. In result, this flow of Na^+ deep into cells sharply decreases, and for K^+ increases. This results in the occurrence of longer K^+ starting current, which answers with a repolarization phase, which is the restoration of polarity of a membrane in time and answers with ascending phase of AP (Fig. 1). So, membrane repolarization does not occur as the result of inverse moving of Na^+ ions, but as the result of equivalent

amount of K^+ ions exiting from a cell. If an increase of conductivity for K^+ prevents the entrance of some substances, for example, tetraethyl, the membrane repolarizes after AP much slower. This shows that the increase of conductivity for K^+ is an important factor in re-polarizing membranes.

So, AP is caused by a cyclic process of Na^+ input into a cell and the subsequent output of K^+.

Today, on the molecular level, those channels through which Na^+ and K^+ ions move after the occurrence AP, are intensively studied. These ionic channels are specific and differ from the usual mechanisms of transmembrane carrying of ions, with which you are familiar from biology and medical physics. The ionic channel consists of their own transport system called gating mechanism ("gate").

During the last years, the attention of researchers has been chained to a problem of molecular mechanisms, fast Na^+ systems. The task has facilitated with the opening of such a substance as tetrodotoxine (synthesized in tissues of the some fish and salamanders), which in concentrations of 10–7 mmole/liter depresses or blocks fast Na^+ system. The molecule tetrodotoxine blocks only the entrance of Na^+ channel, and the unblocked channels remain intact. Proceeding from this, it is possible to pick up tetrodotoxine concentration, which blocks half of the channels and determines the number of channels on a unit of area of a membrane. Accordingly, such calculations exist, approximately 50 Na^+ channels on 1 micrometer of membrane, i.e. channels are on the average located at a distance of 140 nanometers from one another. The channel pores

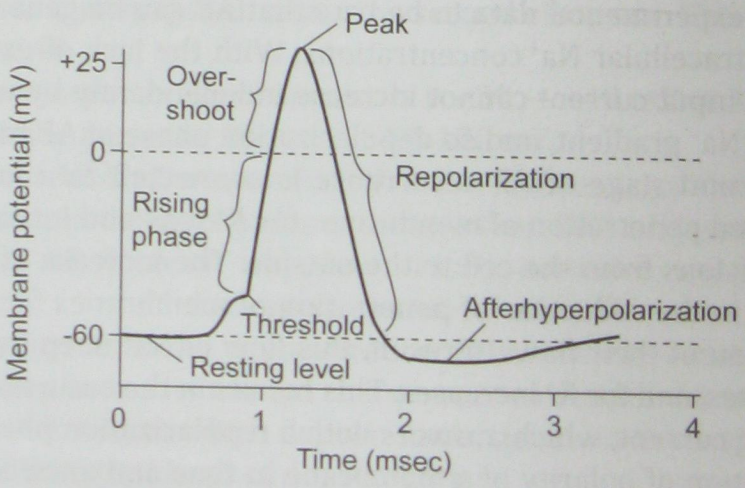


Fig. 1–1. The phases of an action potential.

open only for short time because in 1 millisecond inactivation occurs.

Gates can be in three positions: they are completely closed or completely open and inactivated. The states of voltage-gated sodium and potassium channels are correlated with the course of the action potential (Fig. 1–2). At the RMP, both channels are in a closed, resting state (Fig. 1–2, A). During the depolarizing phase of the AP the voltage-gated sodium channels are activated (open), but the potassium channels open more slowly and, therefore, have not yet responded to the depolarization (Fig. 1–2, B). During the repolarizing phase, sodium channels become inactivated, while the potassium channels become activated (open; (Fig. 1–2, C). During the afterhyperpolarization, the sodium channels are both closed and inactivated, and the potassium channels remain in their active state (Fig. 1–2, D). Eventually, the potassium channels close and the sodium channel inactivation is removed, so that both channels are in their resting state and the membrane potential returns to resting membrane potential. Note that the voltage-gated potassium channel does not have an inactivated state.

AP properties:

1. *The first property for AP is depolarization.* If depolarization reaches a critical level, the threshold value, then AP rises to a constant size irrespective of whether stimulation will amplify or be extended. The amplitude of AP is therefore "all or none". When depolarization is below a threshold value, the voltage-regulated gates are closed; when depolarization reaches threshold ,a minimum potential charge(the AP) is produced. Since the change from -65mV to +40mV and back to -65mV lasts only about 3 mse , the image of an AP on an oscilloscope screen looks like a spike. AP is therefore sometimes called spike potential. This property is known in physiology as the rule "all or nothing ". The reason for such property of AP is that it is caused by a cellular reaction, which is supported and carries, as is said, an auto-regenerative character. Stimulation at the starting moment is a push for the endocellular mechanism which generates AP (positive feedback effect)
2. *The second property of AP is refractoriness.* To induce the second AP just after the occurrence of the first is impossible. If a membrane is discharged during the depolarization phase, excitation does not occur, not with the value of the threshold potential for previous AP, not with any stronger depolarization. Such a condition of full unexcitability, which in nervous cells lasts 1 millisecond, is referred to as the absolute refractory period. It is

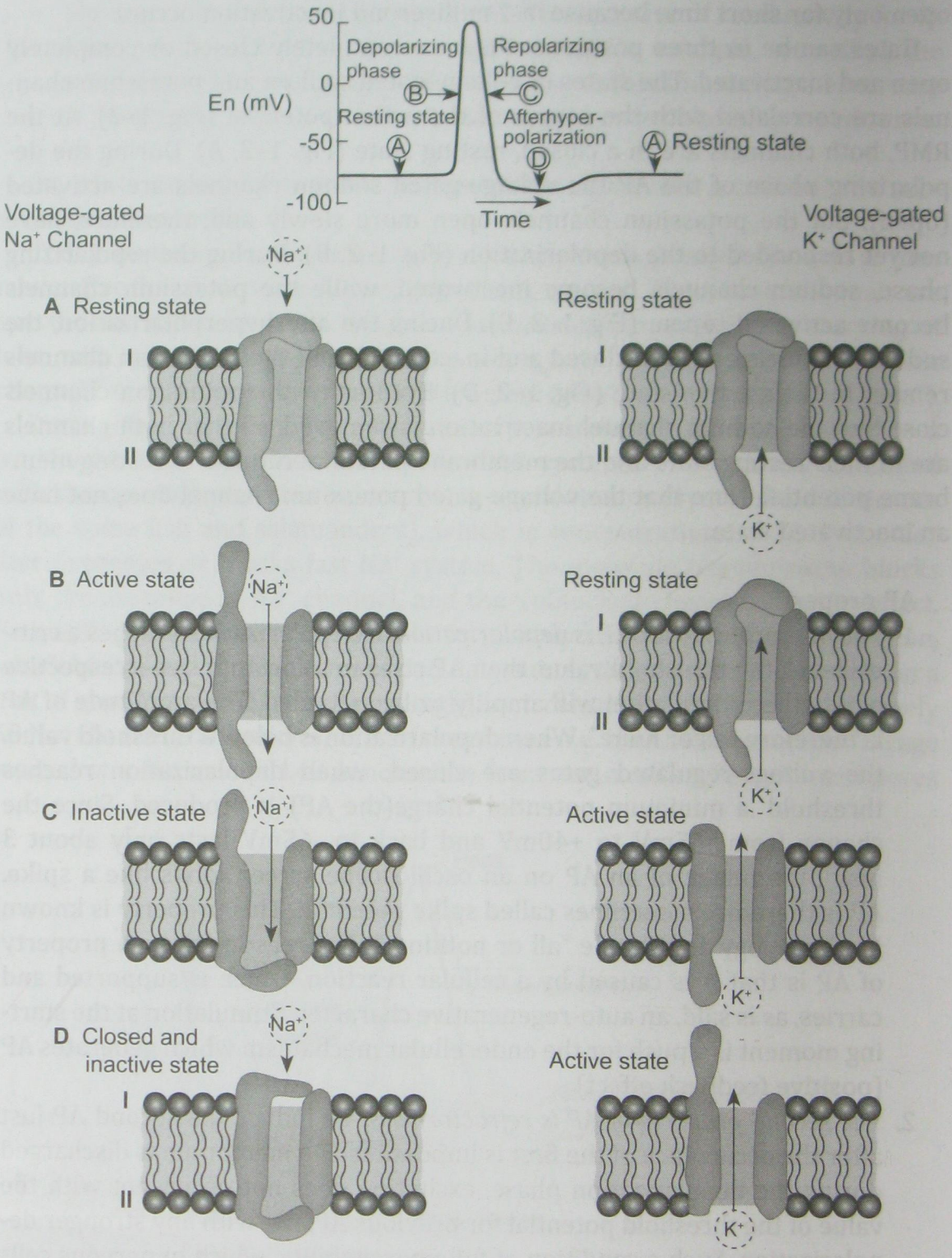
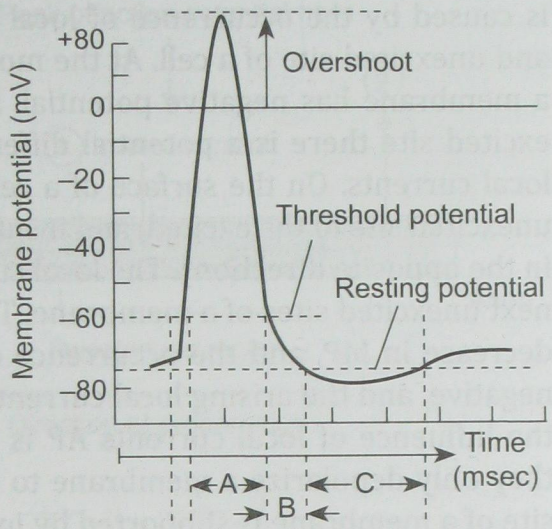


Fig. 1–2. Corres-pondence of AP stages and voltage-dependent channels state.

followed by a relative refractory period (Fig. 1–3) (refractory — unexcitability) which corresponds to a phase of repolarization of AP. Thus, excitability is gradually restored to an initial level (characteristic of RMP); the tissue is capable of being excited only under the influence of the super threshold stimulation, and then amplitude will be smaller.



A=Absolute refractory period (active potential and partial recovery)
B=Relative refractory period
C=Positive relative refractory period

Fig. 1–3. Time-course of AP and the refractoriness stages.

The following depolarization phase of AP corresponds with the exaltation phase or the absolute refractory period. Thus, tissue is capable of being excited even with the action prior to threshold stimulation.

During the following repolarization phase of AP, excitability is reduced a little — a relative refractory period. Refractory is caused by the inactivation of Na^+-system during the previous AP. Even though during membrane repolarization, the inactivation condition transmits, such restoration represents a gradual process, which continues for several seconds during which the Na^+ — system is not yet capable of becoming active. Hyperpolarization is equal to the positive relative refractory period.

Value of absolute refractory period. The absolute refractory period limits the maximal frequency of AP generation. For example, if the absolute refractory period comes to an end 2 milliseconds after the beginning of AP, the cell can be excited with a frequency of maximum 500/sec, and if it lasts 500 milliseconds — 2/sec. Its third property follows from the second AP property — rhythmic character.

The fourth property — the ability of AP to leave after itself a long-term trace of excitability change. The basis of excitability change during the following processes are changes of electric polarization of a membrane.

The fifth property of AP — the property to self-distribution (propagation). AP is distributed along a membrane. It can arise and be distributed and in the event that protoplasm is replaced with an artificial salt solution. Distribution of AP

is caused by the occurrence of local currents circulating between the excited and unexcited site of a cell. At the moment of excitation the external surface of a membrane has negative potential. As a result, between the excited and unexcited site there is a potential difference which results in the occurrence of local currents. On the surface of a cell membrane this current flows from the unexcited site to the excited; and inside — from the excited to the unexcited (i.e. in the opposite direction). The local current conducts a simulative action on the next unexcited sites of a membrane. This results in an increase in permeability, decrease in MP, and the occurrence of AP. This site in turn becomes electronegative, and the arising local current in it stimulates the following site. Under the influence of local currents AP is distributed without attenuation because they only depolarize a membrane to its critical level, and AP in each separate site of a membrane is supported by independent ionic streams.

In nervous fibers, the character of AP propagation depends on the presence or absence of myelin sheaths in them. In unmyelinated fibers AP propagates continuously along the whole membrane (Fig. 1–4). All of the membrane sites then become excited.

In myelinated fibers, in which myelinated segments are interrupted by a node of Ranvier, AP is distributed from one node to another, saltatory conduction, because only in these places the membrane is electrically active (Fig. 1–5). In the node of Ranvier, fibers are concentrated potential-depended Na^+-channels. So, the quantity of Na^+-channels in 1 myelin surface square micrometer is 20–25, in the node of Ranvier — 2000–12000 and 25 — in the terminal axon. Local current in myelin fibers circulate only between the nodes of Ranvier. It seems as if AP is jumping. Such a way of distribution is referred to as saltatory. As a result of saltatory conduction, AP in myelin fibers is higher than in unmyelinated fibers.

The speed of AP propagation still depends on the time of charge of spaciousness of a membrane which is defined by the capacity of a membrane and size of longitudinal resistance of cytoplasm. The speed of AP distribution in nervous fibers grows with the increase in diameter of a fiber because with the increase in diameter resistance decreases per unit of length of a fiber.

In AP, we can distinguish 2 parts — high-voltage or spike AP, and a low-voltage part. Spike potential has ascending and descending phases. AP has some phases. AP begins with a very fast shift in a positive direction — a phase of increase, or an ascending phase which lasts only 0.2–0.5 milliseconds.

During the ascending phase of a membrane the normal charge is ruined, or "polarization", therefore an increasing phase is also called a depolarization

phase. As a rule, depolarization passes the zero line and MP becomes positive. This positive AP phase is referred to as "overshoot". The descending phase (in duration 2 times surpasses the ascending) responds with a repolarization of membranes. The low-voltage part of AP, for the first time, was registered by D.S. Vorontsov known as the physiologist which many years worked at A. A. Bohomolets Institute of Physiology of the National Academy of Sciences of Ukraine. This part consists of an after-depolarization (positive after-potential) and after-hyperpolarization (negitive after-potential). The amplitude of these after-potentials does not surpass several mV, and the duration can reach up to hundreds of milliseconds. The after-potentials are registered only at single-phase output.

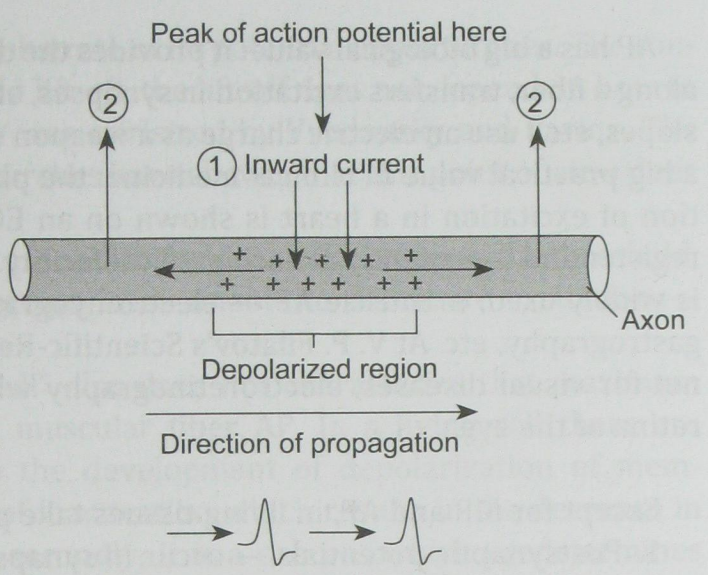


Fig. 1–4. Propagation of an AP in an unmyelinated axon.

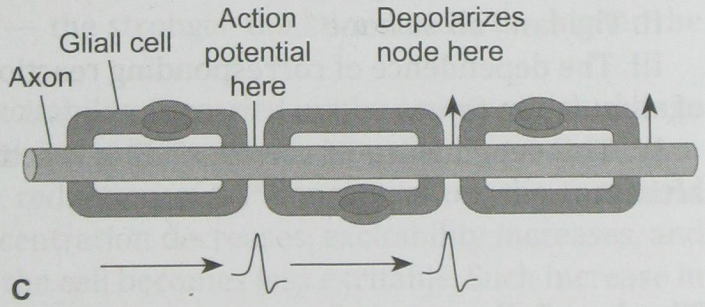


Fig. 1–5. Propagation of an AP in a myelinated axon.

Because bioelectric activity displays the excitation process, its registration has a big theoretical and practical value. Therefore, by registering AP it is possible to judge the speed of distribution of a pulse. Certainly, for this double-phase AP is used. The time between phases and distance between electrodes are determined. With these two values the time of nerve impulse distributions is determined.

It is possible to use a single-phase output. If on the AP record the moment of stimulation is marked, it is possible to determine the duration of the latent period.

AP has a big biological value: it provides the distribution of nervous impulses along a fiber, transfers excitation in synapses, etc. Some of animals (electric eel, slopes, etc.) use an electric charge as a weapon of attack and protection. AP has a big practical value in clinical medicine: the place of occurrence and distribution of excitation in a heart is shown on an ECG, which is an opportunity to register and is presently at each medical facility. In clinical practice registration is widely used: in muscle AP — electromyography; in the stomach — electrogastrography, etc. At V. P. Filatov's Scientific-Research Institute there is a cabinet for visual diseases, electroretinography where AP is registered from the retina of the eye.

Except for MP and AP, in living tissues take place:

1. Postsynaptic potentials — occur in synapses.
2. Receptor potentials — registered in receptors.
3. Secretory potentials — occur on membranes of secretory cells.

Laws for stimulation of excitable tissues.

I. The law of force gradient.

II. The law "all or none ".

III. The dependence of corresponding reactions on the steepness of increase of stimulation force.

IV. The dependence of corresponding reactions on the time of stimulation action.

The law "all or none"

When studying the dependence of stimulation effects and the force of the enclosed stimulus, the law "all or none" was established. According to this law, subthreshold stimulations do not cause AP ("none"), with threshold stimulus excitation immediately gains maximal size ("all") and does not increase any more with further increase in stimulation.

This law for the first time was disclosed by Boudich (1871) when researching the heart, and further confirmed with other excitable tissues. English physiologists suggested in the law of gradation the fact that gradation is imaginary.

For a long time, the law "all or none" was incorrectly taught as a general principle of reaction of excitable tissues. It was taught that "none" meant the full absence of an answer to subthreshold stimulus, and "all" was viewed as the

full exhaustion of an excitable substrate of its potential opportunities. The contraction brought to the threshold lift-off, the lift-off does not depend on how to make arson. In Russia, this law was cririsized by Vvedensky and Pavlov. The succeeding microelectrode researches have shown, that this point of view does not correspond with reality.

It appears that with the force of subthreshold stimulus, close to the threshold, in a direct subthreshold site occurs local, unable to distribute excitation — a local answer.

Moreover, it appears, that "all" also does not characterize that maximum which can reach nervous and muscular fiber AP. In a living cell there are processes which actively stop the development of depolarization of membranes. At first, it is a process of inactivation which results in the decrease in permeability for Na^+ ions and, secondly, it is a process which predetermines the increase in permeability for K^+ ions.

The last may be proved by the influence of any actions on a nervous fiber (for example hypoxia and local anaesthetics — Novocain, relaxants, some poisons, drugs) that decrease the entrance of sodium currents which provide AP generation; it ceases to submit to the rule "all or nothing "; AP amplitude depends on the stimulus force — the stronger the stimulation, the higher the answer amplitude.

An important influence on excitability is carried out by agents which change the concentration of Ca^{2+}. An increase of Ca^{2+}concentration facilitates the activation of Na^+- systems. With the reduction of Ca^{2+}concentration, the threshold increases. From this, if Ca^{2+} concentration decreases, excitability increases, and if Ca^{2+} concentration increases, the cell becomes less excitable. Such increase in excitability underlies tetany syndromes and other conditions connected with deficiency of Ca^{2+} in blood, muscular contractions and spasms occur.

Similar with Ca^{2+} effects, heart glycoside and some anti-arrhythmic drugs developed. These drugs decrease tissue excitability, as a result of the influence on the inactivation Na +-systems and AP generation.

Therefore "all or none" is considered not as the general law of reaction of excitable tissues, but only as a rule which characterizes the AP distribution features.

III. THE ROLE OF TIME DURING STIMULATION AND EXCITATION

Emil du Bois-Reymond did an experiment — he took a nervous-muscular drug and stimulated it through a wire of direct current from an accumulator. He determined the threshold of stimulation. A slide of 5 cm (20 mV) — a closed current. He placed a slide on "0" and gradually with the slide, increase the force of a current, the nerve can burn, and excitation (i.e. contraction of a muscle as a parameter) was not received. So, it has been established, that not only the force of a current matters but also the degree of its increase. This is the Emil du Bois-Reymond law.

Emil du Bois-Reymond (1848) formulated the law of stimulation — stimulation of a current does not depend on the absolute value of force or density of a current, but it depends on the speed of change of intensity of a current in time.

It is proved by that that fast change of a force of a current has a stimulation effect, while slow increase or falling of a current has no effect.

This law implies two conclusions:

1) direct current, when passing through tissue, does not have a stimulation action;
2) the time of passage of a current, after achieving a certain size, has no value for the occurrence of excitation impulses.

The first part of the law — absence of excitation when direct current is transmitted — was noted by Volt.

The second part, in which the time of action of a current is not taken into account, appeared incorrect. This law dominated for a long time, supported by Emil du Bois-Reymond's authority, even if it was wrong.

Experimental researches on slowly reacting tissues have shown, that the weaker a current, the longer it should act on a tissue to cause an effect. So, for example, the data from the experiences of a Swiss physiologist Fick (1863) on the ureter have shown a ratio between time of stimulation and the threshold of direct current (Table 1–1).

Fik found that beginning at some certain interval of time, the course of current duration matters. Further on, these data confirmed for all absolute creations.

For the first time, a quantitative estimation of the dependence of stimulation of a current on the time of the course was offered by Gorveg (1892) and Veis (1901), having presented it as an equation (Gorveg — Veis equation):

Table 1-1. Ratio between time of stimulation and the threshold of direct current

Time of stimulation, sec	The threshold of direct current, cm
0,1	10
0,5	10
0,25	15
0,1	25
1,5	100

$$i = \frac{a}{t_{cm}} + b$$

where: i — intensity of stimulation;
t_{cm} — duration of irritation; a, b — constants.

Lapik (1926) presented this dependence graphically, where on an axis abscissa there is the postponed duration of stimulus and on the ordinate axis — the intensity of a current.

The curve of dependence of time of action of a current received sort of an equipotential hyperbole. The curve shows — the smaller the time of action of a current, the larger the intensity necessary for achieving the threshold effect. It has an identical form for all living tissues.

The minimal voltage, which hardly is enough to cause excitation with a boundless action of direct current, is called rheobase.

The minimal time, during which the current with a voltage equal rheobase, should act on tissue, causing excitation, is called utilization time.

It is established, that in each case an establishment of all curve dependence demands spending significant time. It was important to choose on the curve a point which allowed to quickly and neatly judge in each tissue the speed of physiological processes. The definition of utilization of time cannot be exact. Behind these meditations of the characteristics of excitability from the side of time, L.Lapik suggested to determine the minimal duration, during which the current, equal double rheobase, should act on a tissue to cause an excitation effect. This value received the name chronaxy, and the technique of studying excitability of tissues by defining chronaxy — chronaxymetry

Chronaxy for the flexors of the hands of a person is — 0.08–0.16, the extensors — 0.16–0.32 milliseconds.

The faster a tissue reacts, the shorter the chronaxy. The bigger the speed of excitation propagation, the shorter the chronaxy. There is a rule — the speed of excitation distribution in a time interval is 1 cm, which is equal to the size of chronaxy of the given tissue.

It is determined by a chronaxymeter, constructed on the principle of condensers. It is known, that time of discharge of a condenser depends on its capacity and resistance through which the discharge passes. Resistance is known. Only capacity is taken into account.

The measurement of chronaxy is reduced to, using long enough stimulation, finding rheobase. The received rheobase is increased by the double and gradually increasing the capacity of the condenser, we achieve an effect. The time of the discharge of the condenser, which gives an effect, will be chronaxy.

Many authors in various times suggested different modifications to the Gorveg-Veis equation, which was later also called the characteristic equation of equal threshold stimulation.

IV. THE ACTION OF DIRECT CURRENT ON A LIVING TISSUE

In answering the question, what is excitation, it is necessary first to solve what is stimulation. A lot of scientists in the second half of the 20th century paid attention to the question — what is stimulation. Electrical stimulation, used for the stimulation of living tissues, was given an intensive study. Living tissues are very sensitive to an electric current. In studying the action of an electric current on tissue, direct current was applied, because an induction current for this purpose was not suitable. It was shown that using AP direct current to stimulate a nerve or a muscle, excitation arises at the moment of closing of a direct current only under the cathode (–), and at the moment of disconnection — only under the anode (+). These factors are united and called the polar law of stimulation, formulated by Pflüger (1859).

Position of the polar law:

1. Stimulating action of a direct current is shown at the poles.
2. During closing excitation occurs at the cathode (–), during disconnection — at the anode (–)
3. While closing on the cathode, stimulating action is stronger than at the disconnection on the anode.

There is an exception to this law: a very strong current can excite during the passage of a current (remains as an aftereffect).

There are some experimental models for researching the polar law and the finishing of the polar law:

1. On a preparation of two frog legs.
2. On a nervous muscular preparation with affected conductivity of a nerve.
3. On a damaged muscle.
4. On the example of idiomuscular contraction.

The mechanism of polar action of a direct current

It was investigated with the help of two pares of microelectrodes: one — deflecting, the second — stimulating. It appeared that the action potential (and so excitation) arises only if the cathode collides with the external surface of a membrane, and the anode is inside the cell. If vice versa — excitation does not arise.

Changes in membrane potential

While applying direct current to the area of the cathode, on the external surface of a membrane the positive potential increases, the following occurs:

1) depolarization of the membrane under the cathode;
2) hyperpolarization under the anode while passing through a nervous fiber with a weak subthreshold current.

These changes in membrane potential are observed at some distance from the poles of direct current. Shifts of membrane potential closer to the poles is referred to as electrotonometry.

They distinguished catelectronic and anelectronic changes. These changes are not connected to the active answer of excitable creation and consequently are referred to as passive.

The increase of membrane potential under the anode (passive hyperpolarization) is not accompanied by a change in ionic permeability of membranes, therefore, when closing a direct current, excitation does not occur.

Decrease of membrane potential under the cathode (passive depolarization) in comparison with anaelectrical current occurs first with a short-term increase of permeability of membranes for Na^+ ions and then slowly, but nonperishable, increase of permeability for K^+ ions. An increase of Na^+ entering the protoplasm strengthens membrane depolarization, and it leads to new substantial increasing in sodium permeability into membranes, and so, for further depolarization which in turn predetermines an even greater increase in sodium permeability, etc. This process is called regenerative depolarization. When the critical level of depolarization of membranes is achieved, which is determined exclusively by the properties of a membrane (instead of character of stimulation, an arrangement of electrodes, etc.), action potential occurs which characterizes excitation.

Membranes will play a role in the critical depolarization mechanism, and active subthreshold changes of membrane potential which appear as the local answer.

So, the action potential at the cathode area originates as a result of membrane depolarization, which is predetermined by the sum catelectrical potential and the local answer.

Electrodiagnostic law of Pflüger (1862–1863):

CI > AI > AI > CD

With an increase of current force a normal tissue reacts, if the active electrode is the cathode. The presence of the answer affects cathodic intensity (CI). With amplification of a current the nerve answers with excitation, not only closing of the cathode, but also works on closing of the anode, but CI > anodic intensity (AI). The underthreshold current detects excitation and on disconnection of the anode: CI > AI > anodic depression (AD). If a current decreases below threshold even more, we shall receive cathodic depression (CD) and the following equation:

CI > AI > AI > CD.

This is what a St. Petersburg scientist Pflüger studied. It appeared that this was of great importance. This is how a normal nerve reacts. If it is damaged or exposed to depolarization, then it reacts to an electrocurrent differently: a weak current — answers АЗС, and then on КЗС. This disfigured reaction of a

nerve has found a wide application in practice. This is the electrodiagnostic law of contraction and this method has received the name electrodiagnostics. It is applied in neurologic and neurosurgical clinics.

Physical electrotone occurs in a nerve in connection with the action of an electric current. These currents are distributed faster than excitation and do not display any physiological processes. An increase in excitability on the anode and a decrease on the cathode with the disconnection of direct current is connected with the occurrence of a polarizing current which has a direction opposite to the polarized current.

Chapter 2. Physiology of muscles. Skeletal and smooth muscles

I. GENERAL FEATURES OF MUSCLES AND CLASSIFICATION

Features of an animal organism — the ability to move, whether bodies move by one relating to the second, or whether the whole organism moves. This feature arose at early stages of evolution. The difference between movements of a person lies in their consciousness and laboriousness. The ability to move — an important feature of the living. An animal with the help of this feature finds food, is protected, and without this movement there is no breath, blood circulation, etc.

For the accomplishment of movement a specialized tissue, muscular tissue, came to pass. People and animals with spines have 3 kinds of muscles: skeletal muscles — which are attached to the bones of the skeleton, their function is to move the human body; smooth muscle of the internal organs, vessels and skin; cardiac muscles. Skeletal and cardiac muscles are striated muscles. The actin and myosin filaments are arranged in large parallel arrays in bundles, giving the muscle fibers a striped or striated appearance when they are viewed through a microscope.

Skeletal muscle is the most abundant tissue in the body, accounting for 40 % to 45 % of the total body weight. Most skeletal muscles are attached to bones, and their contractions are responsible for movements of the skeleton. Cardiac muscle, which is found in the heart, is designed to pump blood continuously. Smooth muscle is found in the iris of the eye, the walls of blood vessels, hollow organs such as the stomach and urinary bladder, and hollow tubes, such as the ureters, that connect internal organs.

The difference between these muscles is not only in their arrangement, but also in the structure, physiological properties, innervation, etc. For example, skeletal muscles contract quickly, smooth — slowly, and contraction continues for a long time. All skeletal muscles have a striated appearance when viewed with a light microscope or an electron microscope (Fig. 2–1). The regular and periodic pattern of the cross-striations of skeletal muscle relates closely to the way it functions at a cellular level.

Functions of the skeletal muscles:

1. Moving an organism through space.
2. Moving parts of the body in relation to one another.
3. Keeping a steady posture.
4. Assist in the movement of blood and lymph.

In fact, we can tell, that the basic physiological function of muscles is movement. An important secondary function of skeletal muscle is the production of body heat. In addition, muscles are the place where glycogen is stored, etc.

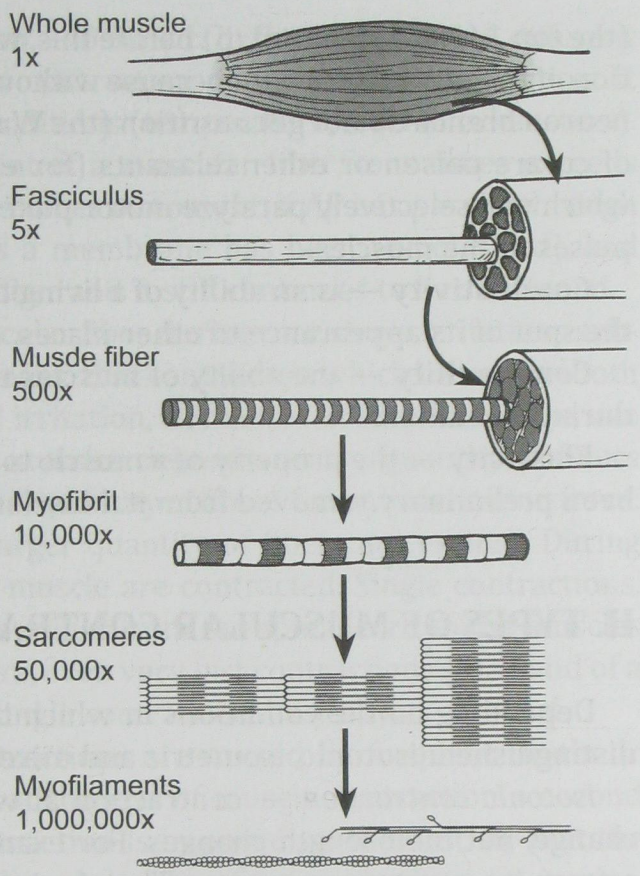


Fig. 2–1. Levels of complexity in the organization of skeletal muscle.

Physiological properties of muscles: excitability, conductivity, contractility, elasticity

Excitability — the ability of excitable tissue to answer by an action of a stimulus with the generation of AP. Irritability is the property of any living substance and arises together with it, but in different tissues it expresses differently: in natural conditions of excitation of muscles it will be caused by nervous pulses which come from the CNS. To excite a muscle in an experiment or at clinical research of a person, he is given artificial irritation with an electric current.

Indirect and direct irritations are distinguished. Indirect irritation — irritation of the muscles through a nerve. Direct irritation — direct irritation of the muscles. Because excitability of a muscular tissue is smaller than that of a nervous one with direct irritation the threshold raises. With direct irritation the current, being distributed on a muscular tissue, operates first of all on the end of motorial nerves which are located in it and excite it.

For the proof of direct irritation, it is necessary to lose nerves; it is provided in this way: a) the muscle in the top (proximal) part has no nervous endings

(the top 1/4 or 1/5 parts); b) before this, we cut a nerve, then comes degeneration of the nervous endings because without connection with a nervous cell, the neuron branch do not get nutrition (the Waller law or degeneration); c) the use of curare poison or other relaxants (for example, Lysthenonum, Diplacinum) which will selectively paralyze motor plates and prevents the transfer of nerve pulses to the muscles.

Conductivity — is an ability of a living tissue to conduct the excitation from the spot of its appearance to other places.

Contractility — the ability of muscles to be contracted or change pressure during excitation.

Elasticity — the property of a muscle to return to its prime position, if it has been preliminary removed from it. Muscles have small, but perfect elasticity.

II. TYPES OF MUSCULAR CONTRACTIONS

Depending on the conditions in which there is contraction, three types are distinguished: isotonic, isometric and mixed contractions.

Isotonic contraction — contraction at which the tone of muscles does not change, but their length changes. For example, a muscle raising a load which causes its constant pressure. That is, isotonic contraction is shortening of muscles with a constant tone thus one end of the muscle is fixed.

Isometric contraction — contraction at which the tone changes, and the length remains without changes, i.e. such a contraction at which the muscle cannot be shortened, for example, if both ends of it are motionlessly fixed with the attempts to lift an excessive cargo. Natural contractions of muscles in an organism never are cleanly isotonic or only isometric, like muscles lifting a load are shortened and together change the pressure.

Depending on the character of stimulation, the following kinds of contractions are distinguished: single, tetanic, tonic and rigor contraction.

1. Single contraction — (twitch) the answer of a muscle to a single irritation or to one nervous pulse. This elementary answer of a muscle consists of three periods: latent, contraction and relaxation. The period of time from the moment of irritation to the beginning of contraction is referred to as the concealed or latent period. Its continuation in different muscles is different (skeletal muscle — 0.1 sec., smooth muscle up to 1 sec). At this time in a muscle there are biophysical and biochemical processes, one of which is AP. At the time of single contraction m. gastrocnemius in frogs has the duration of 0.1 sec., during con-

traction is 0.05 sec., during relaxation — 0.05 sec. In the same animal, different muscles have different durations of contraction depending on its function. It depends also on the surrounding conditions (temperature, level of exhaustion). Duration of contraction in each point of a muscular fiber, tens times exceeds the duration of AP. Therefore, there comes a moment, if AP has passed through all the fiber and has ended (that is a membrane has been re-polarized), and contraction has captured all of the fiber and it still continues to be shortened.

Contraction of each separate muscular fiber during rare single irritations submits to the law "all or none ". This means, that contraction which has arisen both to threshold and to above threshold irritation, has maximal amplitude. The size of single contraction of all skeletal muscles depends on the quantity of fibers which take part in contraction, and the quantity of fibers from the irritation force. The stronger the stimulation, the larger quantity of fibers are excited. During maximal contraction, all fibers of a muscle are contracted. Single contractions, in natural conditions, do not meet, because from the spinal cord to the muscles departs a "volley" of pulses. Therefore, even very fast contractions (the hand of a pianist) are not a single twitch. So, single contraction, basically, is artificial.

2. Tetany. The long muscle contractions are caused by frequent rhythmic irritations referred to as tetany. If at the end of muscle contraction a second pulse is submitted, the second contraction is imposed on the first so, that the common pressure will be greater, than during the first contraction, i.e. mechanical stimulation occurs. If a stimulus is repeated with short intervals, single contractions merge into tetany.

Up until now, we have not received the standard explanation of that fact, that the pressure achieved during the time of tetany or super-positions of single contractions, is larger, than the force of single contractions. Kinds of tetany: incomplete and complete. With rather small frequencies of irritations, comes incomplete tetany. For example, if a muscle is stimulated 5 times in 1 sec. there will be 5 single contractions because for one contraction it is necessary 0.1 sec., then for 5 contractions — 0.5 sec. For others 0.5 sec. — pause time. Stimulating 10 times in 1 sec. — we receive 10 single contractions, but without a pause because for contraction there was all sec. $0{,}1 \cdot 10 = 1$ sec. Stimulating 12 times in 1 sec., each repeated contraction comes to a phase of relaxation because for one contraction, it will be used here 1 : 12, 0.08 sec., instead of 0.1 sec. During the stimulation of 15 times in 1 sec., the relaxation will not longer reach the 0 line.

That is why during incomplete tetany there are still elements of single contraction, therefore it is called incomplete. So, a muscle whose single contraction equals 0.1 sec. will be stimulated 20 times in 1 sec. During irritation more than

20 times in 1 sec. full tetany occurs because at such quantity of irritations the muscle will not have time to relax, 1 : 20 = 0.05 sec. (all the time which is necessary for contraction).

It is important that during tetany muscle contraction is summated, and its frequency answers with the frequency of rhythmic stimulation, which will cause tetany. That is, if irritation is 50 times in 1 sec. and causes tetanization of a muscle in it, we shall register 50 AP. It is proven that tetanization develops or superpositioned from a single contraction.That was specified for the first time by Gelmgolts (1847, the theory of superposition).

3. Tonic contraction. If AP is registered, we will have only 1 AP: this continuous, but not rhythmic contraction caused by a single irritation, is referred to as tonic contraction. In tonic contraction a number of muscles sustain life, for example, muscles of blood vessels, muscles of the digestive tract, the sphincter muscles of the bladder. Tonic contraction is greater characterized for smooth muscles. Skeletal muscles also can give tonic contraction, for example, the posture of a person sitting, the posture of a person standing. Though for skeletal muscles tetany is mainly characterized, a number of skeletal muscles are in tonic reduction.

III. THE DIFFERENCES BETWEEN TETANIZATION AND TONIC CONTRACTIONS

1. Tetany — contraction by nature is rhythmic, tonic — not rhythmic. It proves to be true because tetany is accompanied by rhythmic AP, the frequency which completely answers the frequency of rhythmic irritation. Tonic contractions are not accompanied by rhythmic AP. AP occurs only at the moment of occurrence or amplification of contraction. This is the main difference.
2. During registration of tetanization, skeletal muscles lift big (greater) loads, than during tonic contraction. During tonic contraction, muscles develop a smaller force, than during tetanization. Tonic contraction is longer, but weaker than tetanization.
3. Mechanical work during tetany is accompanied by a large expenditure of energy, increase of metabolism, increase in the need of oxygen (10 times greater than during rest). Tonic contraction is accompanied only by an insignificant contraction of a muscle. For a long time it was considered, that during tonic contraction the metabolism of a muscle does not

increase. It is established that during tonic contraction not enough energy is spent. It is said, that if the muscles of the walls of blood vessels were tetanizied, it would be necessary to spend all of one's daily diet. Only because they are in the condition of tonic contraction there is an economy of energy. During evolution where there was a necessity for long contraction tonic contraction was used, and where there was a necessity for stronger, but periodic, contraction, tetanization was used. During tetanization, exhaustion quickly develops, during tonic contraction, exhaustion does not come.

4. The fourth kind of contraction — contracture (rigor contraction).

Contracture or *rigor contraction* is a long contraction of a spasmodic nature, a pathological condition of return by stationary contraction which arises under the influence of long working rhythmic stimulations of high intensity. It differs from tetany because of the absence of AP distribution. Thus, a long local depolarization of the muscular membrane occurs. For example, if with the help of caffeine causing rigor contraction, MP can be close to AP level.

Rigor contraction is a contraction which stands on border of pathology and arises under abnormal conditions. This long contraction is not distributed, not of the rhythmic nature during which the muscle for a long time does not come back to its normal length and arises under adverse conditions. This is found: if a big irritation is applied to a muscle; during long-term work of a muscle, for example, rigor contraction from weariness; rigor contraction from the action of chemical substances, for example, increases the concentration of K + ions, from the action of nicotine (nicotinic rigor contraction), from the action of acids, caffeine, quinine; from different diseases. For example, if a person, as a result of a wound, kept a long compelled position, rigor contraction develops.

IV. THE MECHANISM OF CONTRACTION

Structure. Skeletal muscle is a highly organized tissue (Fig. 2–2). As you know, the structural unit of skeletal muscles is muscular fiber (diameter from 10 up to 100 micron, length from several mm up to several cm). The contraction device of a muscular fiber is myofibril — thin strings with a diameter of 0.5–2 microns, length is the same length of the muscular fiber. Due to the contraction of myofibril, there is a contraction of muscular fiber and from this movement.

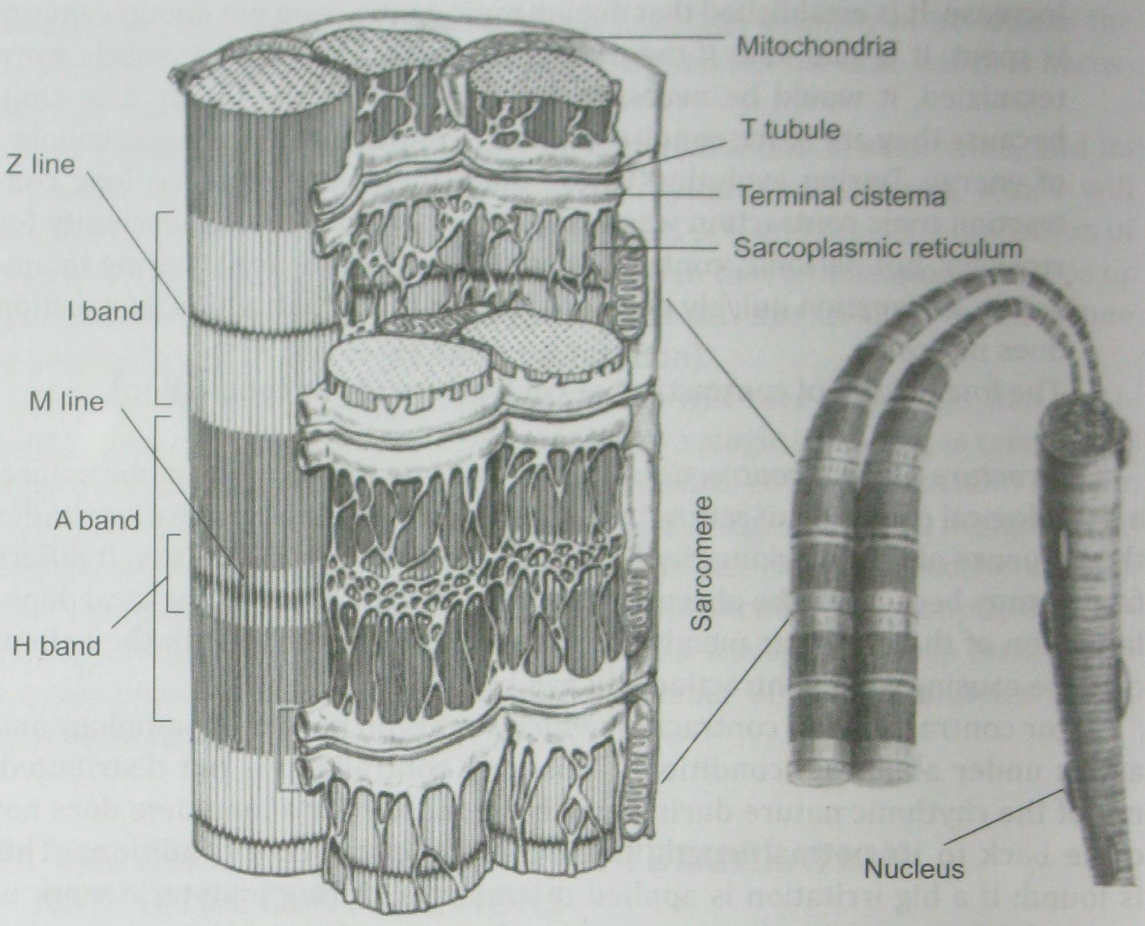


Fig. 2-2. The ultrastructure of skeletal muscle.

Myofibril consists of thick and thin filaments, formed of retractile fibers — actin and myosin. 1 g of skeletal muscle tissue has about 100 mg of retractile fibers — actin (molecular weight 42 000) and myosin (molecular weight — 500 000).

The mechanism of interaction between these fibers during the elementary act of muscular contraction is explained by the theory of sliding, developed by Haxly and Hanson. The septum called plate divides myofibril into some compartments with a length of approximately 2.5 microns, referred to as sarcomeres. Luminous microscope shows that sarcomeres contain light and dark striated strips and disks. According to the Haxly and Hanson theory (1954), this transverse striation of myofibrils is caused by a special organization of filaments of actin and myosin. In the middle of any sarcomere, a few thousand

thick filaments of myosin are located, each one with a diameter approximately of 10 nanometers, on both ends of a sarcomere there are located about 2000 thin (thickness of 5 nanometers) filaments of actin, attached to 2-plates similar to bristles in a brush. The filaments of myosin, length of 1.6 microns, in the middle of a sarcomere look like a dark filament, due to the property of double refractions in polarized light (that is anisotropies), they are referred to as the A-band. On both sides of the A-disk are fields which have only thin filaments and consequently are light, these are isotope I-bands which are pulled toward Z-lines (Fig. 2–2). Due to such periodic changes of light and dark strips in sarcomeres, those indefinitely repeating myofibril fibers of cardiac and skeletal muscles appear are striated. In a muscle which is in relaxation, the ends of the thick and thin filaments are only insignificantly blocked on the border between A- and I-bands.

The muscle is shortened as a result of the contraction of many sarcomeres connected consistently in myofibrils. During shortening, the thin actin filaments slide across the thick myosin filaments (Fig. 2–3), moving between them to the middle of the bunch and sarcomere. During sliding the filaments of actin and myosin are not shortened. This substantive provision of the theory of sliding filaments. The length of filaments does not change during contraction and during muscles stretching. Instead, the bunches of thin sliding filaments leave intervals between thick filaments so that the degree of their overlapping

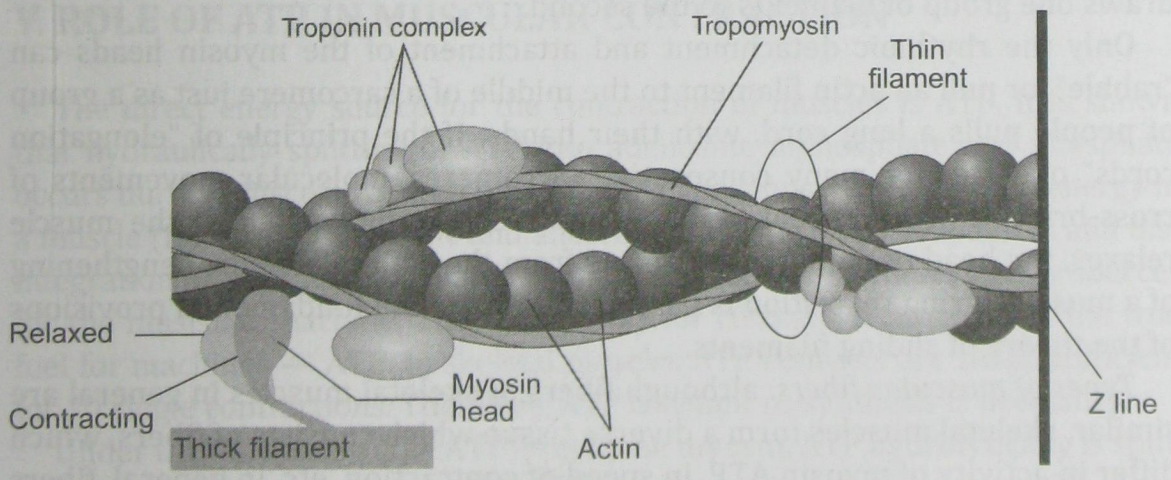


Fig. 2–3. Molecular structure of the thin actin and thicker myosin filament of striated muscle.

decreases. At which level does "sliding directed in different direction" of actin filaments occur in the next half sarcomere?

Myosin filaments have cross-section appearances, in length close to 20 cross-bridges, with the head approximately from 150 molecules of myosin; they come from a bipolar filament. On the actin filaments there are active centers which in a condition of rest are closed by tropomyosin fiber between molecules where there are troponin fibers, having huge resources of Ca^{2+}, in a condition of rest the bridge cannot join with the actin. With an increase of concentration of Ca^{2+} ions, and in the presence ATP troponin cooperates with Ca^{2+}, changes the configuration and moves the tropomyosin molecule, creating conditions for connection of the head of the bridge with the active actin centers. As a result, it is shown by ATP activity with myosin. It is accompanied by changes of position of the head and the moving of an actin filament lengthways to myosin filament.

Thus, during contraction of each head of a myosin (or the cross-bridge) can connect a filament of myosin with the next — actin. Inclinations of the head form incorporated efforts and occurs a "rabble" that moves actin filaments to the middle of the sarcomere. The bipolar organization of myosin molecules in two halves of a sarcomere already provides an opportunity of sliding for actin filaments in the opposite direction in the left and right half of a sarcomere. During isotonic contraction of a muscle of a frog, sarcomeres are shortened by 1 micron, i.e. 50 % of length for 0.1 sec. For this purpose, cross-bridges should carry out the just described rabble, movements not once for such time interval, but 50 times. Cross-bridges will play a role of some kind of "cogwheel" which draws one group of filaments to the second.

Only the rhythmic detachment and attachment of the myosin heads can "rabble", or pull an actin filament to the middle of a sarcomere just as a group of people pulls a long cord, with their hands. If the principle of "elongation cords" operates for many consecutive sarcomeres, molecular movements of cross-bridges which repeat result in macroscopic movement. If the muscle relaxes, the head of the myosin departs from the actin filaments. Lengthening of a muscle during relaxation is passive. These are the fundamental provisions of the theory of sliding filaments.

Types of muscular fibers: although fibers of skeletal muscles in general are similar, skeletal muscles form a diverse tissue which consists of fibers, which differ in activity of myosin ATP, in speed of contraction, etc. In general, fibers are divided into 2 types (I and II). Muscles which contain many fibers of type I are called red muscles because they look darker than others. They have a latent period, slowly react, and are intended for long, slow contractions.

Their main function is supporting a constant position of the body. White muscles consist mainly of fibers of type II; they have short contraction duration and are designated for exact coordinated of movements. An example of such white muscles can be muscles of the eyeball, some muscles of the hand and others.

The basic difference between muscular fibers is the difference between proteins from which they develop. The majority of fibers are multigenes coded. Ten different isoforms of heavy myosin circuits (HMC) are allocated. They have different amino acid structures. In any of the two types of myosin circuits slow and fast isoforms are distinguish. Consider, that there is only one form of actin, three — troponin, and plenty of forms of tropomyosin.

Expression of HMC is very precisely regulated in the development of an organism and difference in expression of HMC will play an outstanding role in the formation of the muscular constitution of the person.

In adults, changes of physical activity can influence function of muscles, innervation and a hormonal background. These changes, as usual, are caused by infringement of transcriptions of HMC genes.

Muscles are machines which transform chemical energy directly into mechanical energy, i.e. work into heat. What level of muscle transforms chemical energy into mechanical? Today this is a hot question in modern molecular arguments.

V. ROLE OF ATP IN MUSCULAR CONTRACTION

The direct energy source for the contraction of muscles is ATP. It is shown that hydraulically splitting of ATP into adenosine diphosphate and phosphate occurs during muscular contraction. All other reactions, which provide energy in a muscle (for example, aerobic and anaerobic splitting of carbohydrates and disintegration of creatine phosphate) cannot be considered as direct energy sources for the muscular machine. They serve only for constant restoration of the true fuel for machines — ATP. In skeletal muscles ATP contents are small, sufficient for 10 single contractions. Therefore, ATP constant re-synthesis is necessary.

Under the influence of the ATP — enzyme myosin, ATP hydrolytically is split. This process is activated by actin. ATP substances in muscles (exception make rare nucleoside triphosphate) can direct the utilization of fiber contraction. The ATP mechanism, with the help of an energy donor, provides the moving cross-bridges today to be intensively studied.

Probably a molecule of ATP connects with cross-bridges after its "rabble" movements. And this provides energy for distribution, splitting of components which take part in the reaction — actin and myosin. Almost at once after that, the myosin heads are separated from the actin. Then, ATP is split into ADP and phosphate with an intermediate formation of an enzyme-product complex. Splitting is an obligatory condition for the next attachment of the cross-bridge to actin with the dismissal of ADP and phosphate and rabble movement. If the movement of the bridge comes to the end, with it a new molecule of ATP connects and a new cycle begins (Fig. 2–5).

Cyclic activity of cross-bridges, i.e. a rhythmic attachment and detachment of bridges which provides muscular contraction, is possible only when ATP hydrolysis proceeds, i.e. during the ATP activation. If the splitting of ATP is blocked, bridges cannot be attached repeatedly; the force of a muscular fiber falls to zero and the muscle relaxes. After death, ATP contents in muscular cells

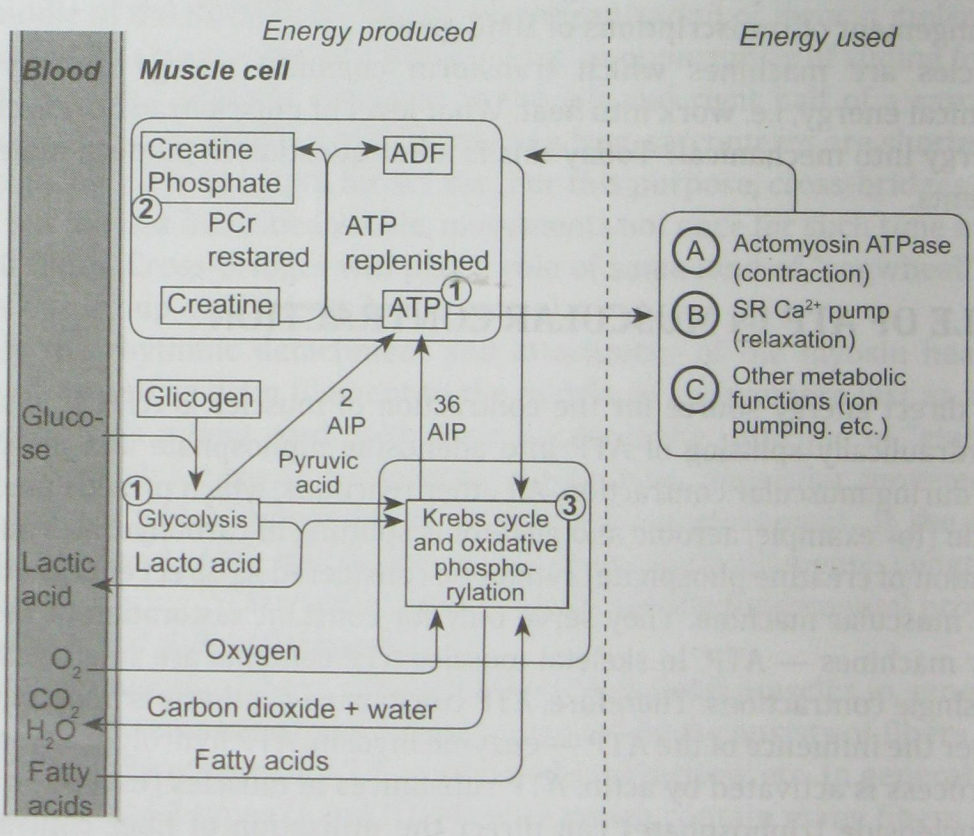


Fig. 2–5. Main metabolic processes in the skeletal muscle.

go down; if it passes critical level, cross-bridges appear steadily attached to actin filaments (yet will not pass autolysis). In such condition, actin and myosin filaments are very strongly connected to one another — the muscle is in a condition of rigor mortis.

VI. REGULATION OF MUSCULAR CONTRACTION

Excitation of muscles frequently passes during the passage of AP from nervous motoneurons through nervous-muscular synapses. Signaling about contraction from the excited cellular membrane to myofibrils in the depth of a cell is referred to as electromechanical coupling. It consists of several consecutive processes, a key role in which Ca^{2+} ions will play.

The mechanism of Ca^{2+}: the endocellular injection of calcium will cause the contraction of muscular fibers. However, muscular fibers are not suitable objects for the demonstration of direct influence of calcium on myofibrils. Whole fibers are more favorable to this with or without the mentioned cellular membrane. For reception of such fibers, "stripping" a membrane in the mechanical way or operate detergents; such de-membrane fibers are contracted only by immersing in a solution which has ATP and the ionized calcium for ATP activation. If the active agent, i.e. ions of calcium, is taken out from the environment (for example, by way of adding chelato-creating agents), myofibrils relax, because the interaction between cross-bridges, and actin comes to an end and as consequence the ATP activity is depressed (Fig. 2–6).

Calcium ions via the troponin-tropomyosin complex regulate inter-action between myosin heads and actin active sites.

The mechanisms, which help Ca^{2+} activate a fiber, are easier to understand when examining the structure of actin filaments. Actin filaments consist of two twirled one around the other actin monomers, thickness up to 5 nanometers. A similar structure is possible to be received if 2 strings of a necklace are taken and braided as a spiral of 14 beads in the filament. On actin circuits with regular intervals (approximately 40 nanometers) spherical molecules of troponin are placed, and in grooves between two circuits of actin filaments tropomyosin lay. With the absence of Ca^{2+}, that is during the relaxed condition of myofibril, long molecules of tropomyosin settle down so that the attachments of myosin cross-bridges to actin circuits are blocked. Under the influence of active Ca^{2+} ions molecules tropomyosin go down deeper into grooves between actin monomers, opening a site of attachment for myosin cross-bridges. In result, the

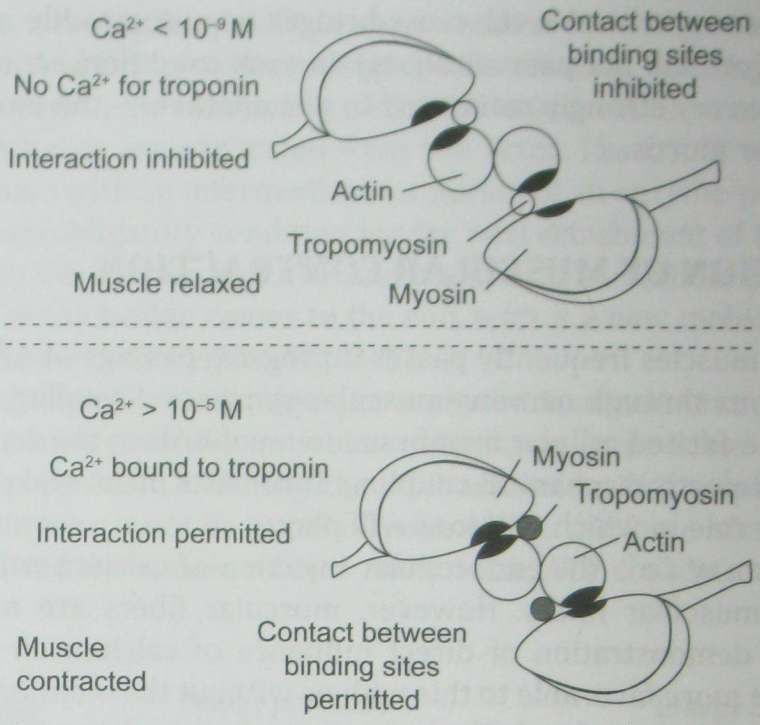


Fig. 2–6. Calcium switch for the muscle contraction providing.

myosin bridges are attached to actin filaments; ATP is split and muscular force develops. Thus, activation effects of Ca^{2+} are caused by its action on troponin C. During this troponin C works as "calcium switches", that is during the linkage with Ca^{2+} the troponin C molecule is deformed so, that it pushes tropomyosin in the grooves between two actin circuits — in "active position" and opens a site of attachment for cross-bridges (Fig. 2–6).

If Ca^{2+} salts were not isolated in special endocellular storehouses, enriched Ca^{2+} muscular fibers would be in a condition of continuous contraction. The structure of endocellular systems of Ca^{2+} preservation differs a little in different muscles. In many sites, the membrane of a muscular cell goes deep into the fiber, perpendicular to its axis forming tubes. This system of transverse tubules (so-called T-system) incorporates with the exocellular environment. Tubules (diameters of 50 nanometers) surround any myofibril on equal 2-plates or around the I-disks. Perpendicular to the transverse system, i.e. parallel to the myofibrils, the system of longitudinal l-tubules is located. Bubbles on the ends of these tubules — terminal sacs, which are the place of preservation of endocellular Ca^{2+}.

Opposite to the transverse systems, longitudinal systems do not incorporate with exocellular environment.

Electromechanical junction occurs with the help of AP distribution on the membranes of T-tubules deep into the cells. Thus, excitation quickly passes into the depth of a fiber, passes to T-tubules and in the end causes the free Ca^{2+} ions to the endocellular liquid near the myofibril, resulting in contraction. Dihydropyridine receptors play the role of electric relays which adjust the receiving of Ca^{2+} with sarcoplasmic grids located beside. Dihydropyridine receptors received the name from a medicamentous preparation dihydropyridine, which actually blocks them and is widely applied in the practice of a doctor (Fig. 2–7).

Opposite to the membrane of T-tubules, Ca^{2+} channels of sarcoplasmatic grids, which opening provides the ejection of Ca^{2+} ions, are not potential-depended. It is named rianodin receptor as it blocks vegetative alkaloid rianodin.

Sarcoplasmatic grid, as a result of the depolarization of the membranes of T-tubules, is activated thanks to the dihydropyridine receptors, which are potential-depended Ca^{2+} — channels in the membrane of T-tubules.

During one-time contraction, the process of contracting soon comes to an end; if active Ca^{2+} ions return to the system of channels of sarcoplasmatic grids

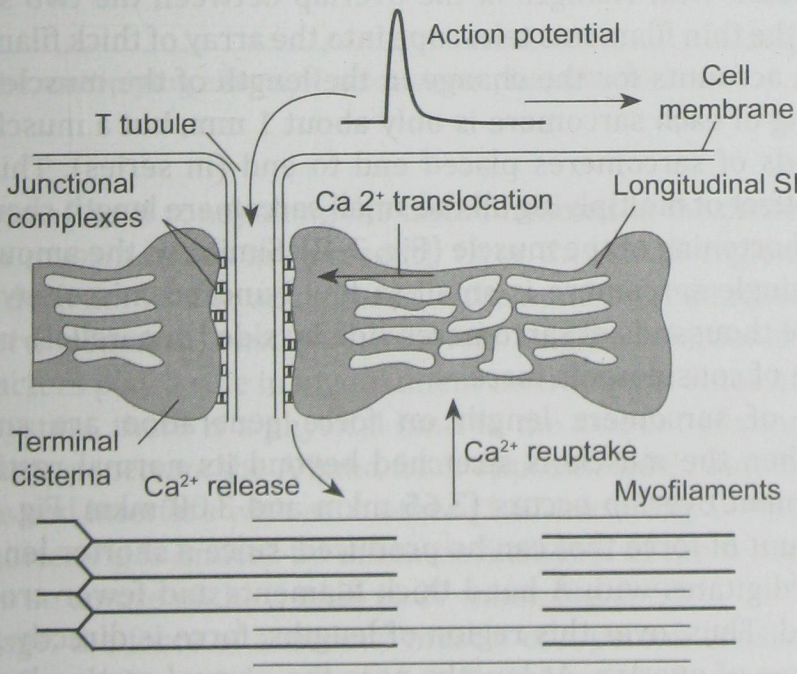


Fig. 2–7. Excitation-contraction coupling and the cyclic movement of calcium.

with the help of calcium pumps, there is a relaxation of muscles. This process goes against the gradient with the participation of active transport, which uses ATP energy. If stimulus act in a muscle with high frequency, 20 Hz or higher, the level of Ca^{2+} during the time of inter-stimuli intervals remains high because the calcium pump has not enough time to return all Ca^{2+} ions in the sarcoplasmic grid system. Thus, a condition of sustained contraction or tetanus occurs. It is frequently observed in the event that intervals between stimuli are smaller than 1/3 of the time of single contraction.

VII. FORCE AND WORK OF MUSCLES

Absolute and relative force is distinguished. Absolute force is defined by the maximal load, which a muscle is capable to lift, or maximal pressure which a muscle can develop. It depends on the structure of a muscle, its functional condition, influence of the CNS. The greater diameter (or physiological diameter) the more force of a muscle. For example, the big ischium muscle can develop a force up to 1 200 kg.

The *sliding filament theory* proposes that changes in overall fiber length are directly associated with changes in the overlap between the two sets of filaments; that is, the thin filaments telescope into the array of thick filaments. This interdigitation accounts for the change in the length of the muscle fiber. The total shortening of each sarcomere is only about 1 mm, but a muscle contains many thousands of sarcomeres placed end to end (in series). This arrangement has the effect of multiplying all the small sarcomere length changes into a large overall shortening of the muscle (Fig. 2–8). Similarly, the amount of force exerted by a single sarcomere is small (a few hundred micronewtons), but, again, there are thousands of sarcomeres side by side (in parallel), resulting in the production of considerable force.

The effects of sarcomere length on force generation are summarized in Fig. 2–9. When the muscle is stretched beyond its normal resting length, decreased filament overlap occurs (3.65 mkm and 3.00 mkm, Fig. 2–9). This limits the amount of force that can be produced, since a shorter length of thin filaments interdigitates with A band thick filaments and fewer cross-bridges can be attached. Thus, over this region of lengths, force is directly proportional to the degree of overlap. At lengths near the normal resting length of the muscle (i.e., the length usually found in the body), the amount of force does not vary with the degree of overlap (2.25 mkm and 1.95 mkm, Fig. 2-9) because of

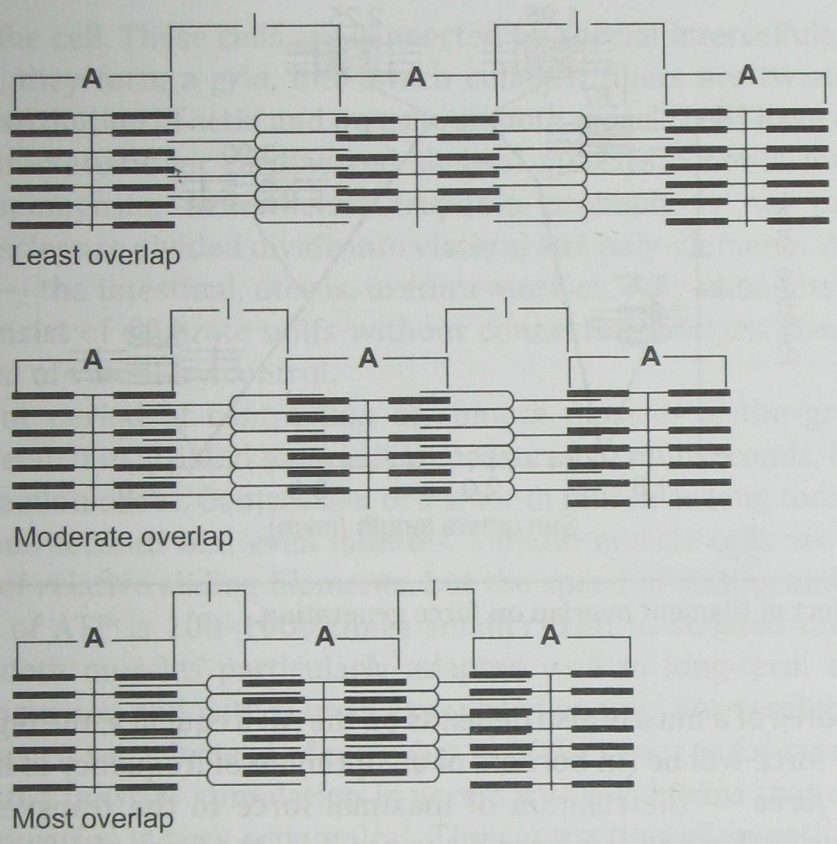


Fig. 2–8. The multiplying effects of sarcomeres placed in series.

the bare zone (the H zone) along the thick filaments at the center of the A band (where no myosin heads are present). Over this small region, further interdigitation does not lead to an increase in the number of attached cross-bridges and the force remains constant. At shorter lengths, additional geometric and physical factors play a role in myofilament interactions. Since muscle is a "telescoping" system, there is a physical limit to the amount of shortening. As thin myofilaments penetrate the A band from opposite sides, they begin to meet in the middle and interfere with each other (1.67 mkm, Fig. 2–9). At the extreme, further shortening is limited by the thick filaments of the A band being forced against the structure of the Z lines (1.27 mkm, Fig. 2–9).

In average, the force of a skeletal muscle of a person consists of 3–4 kg on 1 cm^2. The force of a muscle depends on the quantity of motor units excited at the present time. So, if the structure of a muscle includes 10 motor units, but only one is active and excited the muscle will create a pressure 0.1 from the maximal

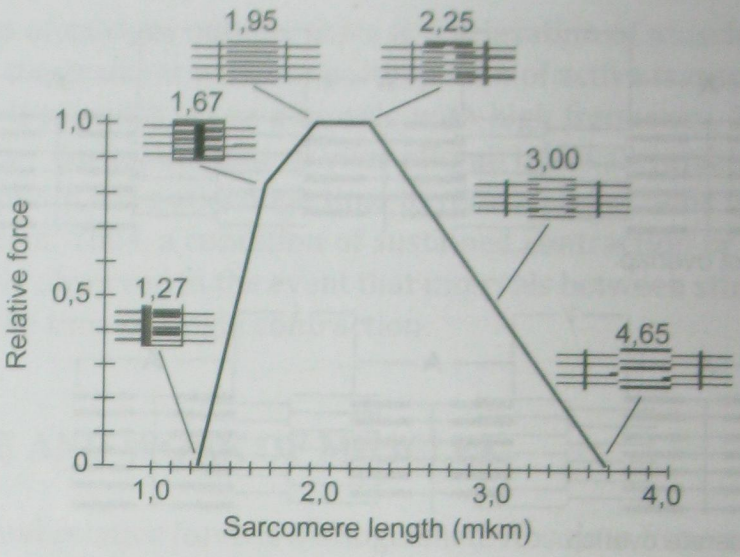


Fig. 2–9. Effect of filament overlap on force generation.

force. The force of a muscle also depends on the AP frequency; the bigger it is the greater the force will be (in borders of an optimum of frequency of irritation).

Relative force — the relation of maximal force to the transverse area of muscles. Due to this size, it is possible to compare the force of different muscles of one organism, and also the force of muscles of different people. For example, relative force ulnar muscles makes 5,9 kg / cm^2, chewing muscle — up to 10 kg / cm^2, the three-head muscle of the shoulder — 16,8 kg / cm^2. A muscle can carry out both dynamic and static work. Dynamic work is defined by the size of the lifted load multiplied by the parameter of the muscle contraction. It is measured in kilograms, joules and calories. Maximal work happens with average loads. Work during which muscles are almost not contracted is referred to as static work. Static work is measured by the size of the load, multiplied by the period of its deduction. Weariness develops faster during static work than during dynamic.

VIII. PHYSIOLOGICAL PROPERTIES OF SMOOTH MUSCLES

In an organism apart from striated muscles, smooth muscles also exist.

Features of the smooth muscles structure. They represent separate tapered forms of a cell, length 50–400 microns, thickness 2–10 microns. They have one

nucleus in the cell. These cells are connected by special intercellular contacts. As a result, they form a grid, into which collagen fibers are twisted. Due to irregular distribution of actin and myosin, smooth muscle cells have no striated strings, characteristic for cardiac and skeletal muscles. Smooth muscles also contain tropomyosin; nevertheless troponin probably is not present. All smooth muscles are divided divide into visceral and poly-elements. An example of visceral — the intestinal, uterus, urethra muscles. Poly-elements of smooth muscles consist of separate units without connecting bridges, these muscles are deprived of voluntary control.

The latent period of contraction of smooth muscles is the greater than during skeletal. If in striated muscle it is measured in milliseconds, here it is in sec. or in fraction of sec. Contraction of a smooth muscle is long tonic. Contraction proceeds seconds and even minutes. Smooth muscle cells are shortened as a result of relative sliding filaments, but the speed of sliding and the speed of splitting of ATP is 100–1000 times smaller than in striated muscles. Due to this, smooth muscles particularly adapted well to long-term contraction without exhaustion and with a small expense of energy. For receiving tetanus, rare (1 time in a sec.), instead of often (up to 20 in 1 sec., as for receiving tetanus in striated muscle) stimulation is necessary. This means that contraction of smooth muscles is very economical. The contraction of smooth muscles is weaker. The metabolism during contraction is less amplified, than in striated muscles. Energy spent by smooth muscles is 100–500 times smaller if estimated by the use of oxygen. Exhaustion is very weak.

Opposite to skeletal muscle, smooth muscle of the stomach, intestines and uterus develops spontaneous tetanus contractions in conditions of its isolation or after blockade of ganglion neurons. That is AP is not caused by the transfer of nervous pulses to muscles. In other words, they have no neurogenetic and myogenetic (as well as heart) origin. Myogenetic excitation arises in pacemaker cells, which are identical to other muscular cells in structure, but differ in electrophysiological properties, namely: local potentials, pre-potential, or pacemaker potentials spontaneously depolarize a membrane to threshold level so the AP arises. Due to an input of cations (mainly Ca^{2+}) membranes depolarize to the zero level, and then during several milliseconds overshoot is found at +20 mV. Following repolarization, comes a new potential which causes one more AP. Excitation is distributed in smooth muscles through special hard contacts between plasmatic membranes of the next muscular cells. These contact areas provide electrical distribution of depolarization from excited to not excited cells. Opposite to skeletal muscles, the majority of smooth muscles during

stretching are more pliable to behave not as elastic structures, but as plastic formations. Due to the plasticity, smooth muscle can be completely weakened, both in short and in the stretched condition. So, for example, elasticity of the bladder with a measure of its accumulation keeps it from increasing to intramural pressure. In many cases, the strong stretching results in the activation of contraction. It is caused by depolarization during the stretching of pacemaker muscle cells which are accompanied by the increase of AP frequency.

Smooth muscles of the arterioles, arteries, muscles of iris, and also muscles of the neck show weak spontaneous activity or do not show it at all. Their activity frequently does not have myogenetic, but neurogenetic nature, caused by pulses which act on the vegetative nerves. The differences are caused by the structural organization of the muscles. The contacts in these muscles are submitted in very little quantities.

Features of contraction. As well as in skeletal muscles, initiation of contraction in smooth muscles occurs because of Ca^{2+}. Nevertheless, smooth muscles have badly advanced sarcoplasmic grids. Therefore, the increase in concentration of Ca^{2+} inside a cell arises owing to the reception of Ca^{2+} from an extracellular solution through potential depended Ca^{2+}-channels. For the activation of myosin ATP, phosphorylation is necessary for a myosin molecule. Ca^{2+} connects with calmodulin and a complex of Ca^{2+}- calmodulin activates kinase of myosin circuits. This enzyme is the catalyst of phosphorylation reaction of easy myosin circuits. Phosphorylation provides further activation of ATP, and actin molecules start to slide in myosin circuits, leading to the contraction of muscles.

Skeletal and smooth muscles always relax if the concentration of Ca^{2+} falls below 10 mole/l. However, their relaxation is slower because the process of reception of Ca^{2+} in a poorly advanced sarcoplasmic grid goes slowly. Whether other mechanism participates in contraction — similar to the connection of Ca^{2+} with troponin — it is not known. ATP plays an important role in the decreasing of the tone of smooth muscles.

Chapter 3. Physiology of synaptic transmission

With the help of electronic microscopy and micro-electrophysiological research, it is established that neurons do not pass directly one into another, but only have contacts between themselves through synapses, which are a thousand greater in amount, than nervous cells. For example, the number of synapses on soma of the big motoneuron of the spinal cord can reach 2000, and on the body of a pyramidal cell of the cortical substance of the big hemispheres — several thousand. Big neurons, as a rule, have greater amount of synapses. Synapses are also located between the branching of axons and receptors or effectors (for example, muscular cells).

Two types of synapses have been identified: *electrical* and *chemical*. At electrical synapses, passageways known as gap junctions connect the cytoplasm of adjacent neurons and permit the bidirectional passage of ions from one cell to another. Electrical synapses are found at dendro-dendritic sites of contact; they are thought to synchronize the activity of neuronal populations.

At *electrical synapses*, passageways known as *gap junctions* connect the cytoplasm of adjacent neurons (Fig. 3–1) and permit the bidirectional passage of ions from one cell to another. Electrical synapses are uncommon in the adult mammalian nervous system. Typically, they are found at dendrodendritic sites of contact; they are thought to synchronize the activity of neuronal populations.

Gap junctions are more common in the embryonic nervous system, where they may act to aid the development of appropriate synaptic connections based on synchronous firing of neuronal populations.

The most common type of synapse is the *chemical synapse*. The following kinds of chemical synapses are distinguished: axo-somatical, axo-dendritical, axo-axonal, and dendrite-dendritical. The majority of synaptic contacts belong to axo-dendritical. Besides, behind the functional features central synapses divide into exciting and inhibiting.

For the corresponding departments of synaptic connections at present such terms are accepted: the ending of a branch of one neuron is called the *presynaptic part* of the synapse (presynaptic membrane), space between membranes which cover pre- and postsynaptic cells, — *synaptic clefts* and that part of the

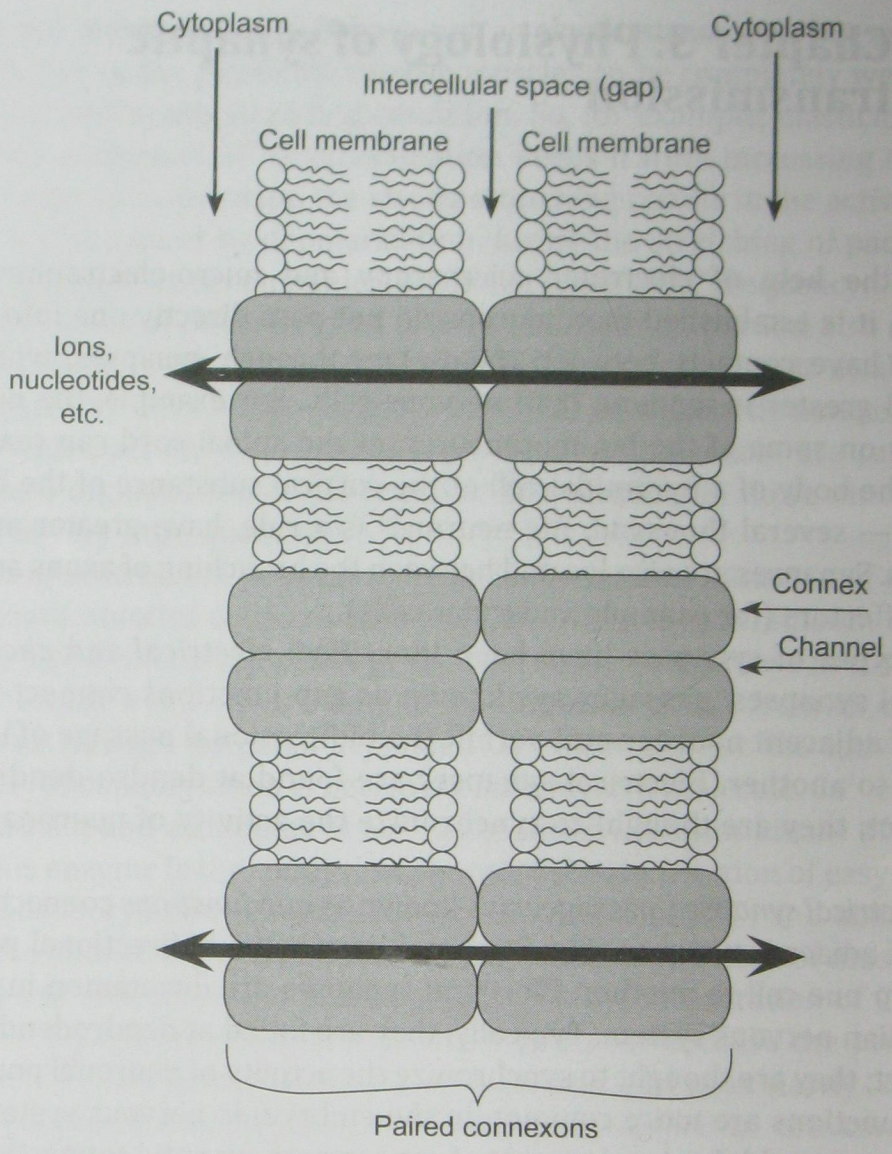


Fig. 3–1. The structure of gap-junction.

second cell to which the ending close approaches, — the *postsynaptic membrane* (Fig. 3–2). The form of presynaptic connections can be different — as small buttons, "final legs", irregular-shaped expansions which contain a net of neurofibrilles, and so on. Most often the small button type is found.

The process of neurotransmission involves the synthesis, storage, and release of a neurotransmitter; the reaction of the neurotransmitter with a receptor; and termination of the receptor action (Fig. 3–3). One of the main

functions of presynaptic membrane will consist in allocation mediators under the influence of AP.

That is distributed by a nervous fiber. In the presynaptic endings spherical formations 20–50 microns in diameter are noticeable. It is vesicles, or bubbles which in different amounts "accumulate" in the nervous endings. Vesicles are distributed on all presynaptic endings, and their distribution and amount can change depending on nervous activity. They gather near the presynaptic membranes. Vesicles contain thousand of molecules of chemical mediators. It is considered that vesicles burst and their contents are poured into the synaptic cleft, carried out by "quantum" dismissal of mediators and transition of excitation from one cell into another as it is described above. There are data that the moving of vesicles to the side of the synaptic cleft and excretion of the mediator occurs under the influence of Ca^{2+}, that enters into the nerve ending. The vesicle protects the neurotransmitters from enzyme destruction in the nerve terminal. The arrival of an impulse at a nerve terminal causes the vesicles to move

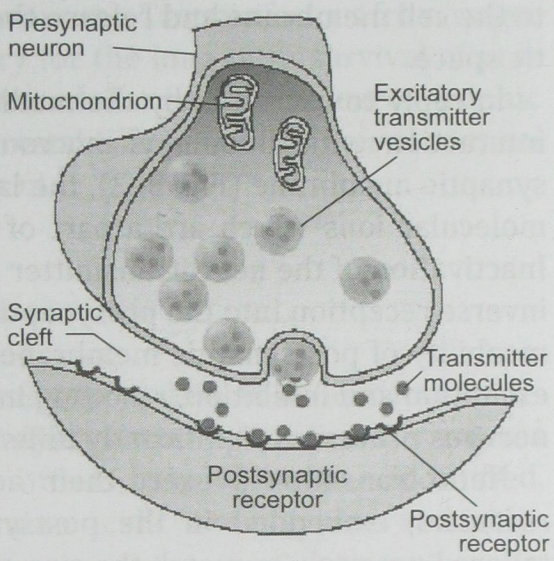


Fig. 3–2. Schematic illustration of the chemical synapse.

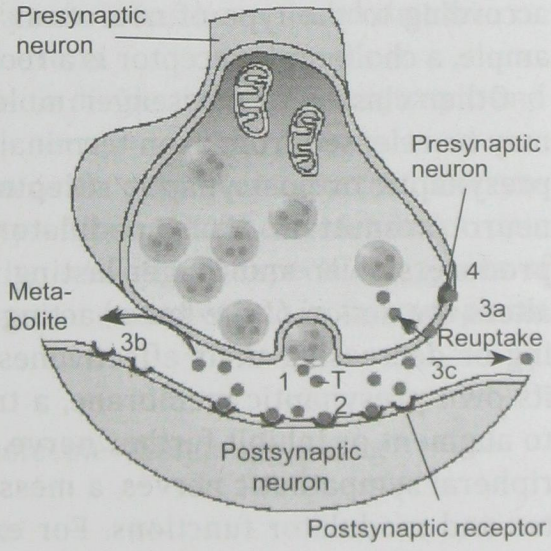


Fig. 3–3. Schematic illustration of (1) neurotransmitter (T) release; its binding to postsynaptic receptor (2); termination of T action by (3a) reuptake of transmitter into the presynaptic terminal, (3b) enzymatic degradation, or (3c) diffusion away from the synapse; and (4) binding of transmitter to presynaptic receptors for feedback regulation of transmitter release.

to the cell membrane and release their transmitter molecules into the synaptic space.

In reply to excretion by the mediator with the presynaptic ending and its interaction with aluminous macromolecules (chemoreceptors) of the postsynaptic membrane (Fig. 3–3), the latter changes its permeability for all low-molecular ions which are a part of the exo- and endocellular environment. Inactivation of the neurotransmitter in the synaptic clefts is carried out by the inverse reception into the presynaptic endings or enzymatic destructions. Permeability of postsynaptic membranes changes as an unequal measure during excitation and inhibition, resulting in the ionic mechanisms of these two basic nervous processes significantly differ.

Neurotransmitters exert their actions through specific proteins, called *receptors*, embedded in the postsynaptic membrane. These receptors are tailored precisely to match the size and shape of the transmitter. In each case, the interaction between a transmitter and receptor results in a specific physiologic response. The action of a transmitter is determined by the type of receptor (excitatory or inhibitory) to which it binds. For example, acetylcholine is excitatory when it is released at a myoneural junction, and it is inhibitory when it is released at the sinoatrial node in the heart. Receptors are named according to the type of neurotransmitter with which they interact. For example, a cholinergic receptor is a receptor that binds acetylcholine.

Other classes of messenger molecules, known as *neuromodulators*, also may be released from axon terminals. Neuromodulator molecules react with presynaptic or postsynaptic receptors to alter the release of or response to neurotransmitters. Neuromodulators may act on postsynaptic receptors to produce slower and longer-lasting changes in membrane excitability. This alters the action of the faster-acting neurotransmitter molecules by enhancing or decreasing their effectiveness. By combining with autoreceptors on its own presynaptic membrane, a transmitter can act as a neuromodulator to augment or inhibit further nerve activity. In some nerves, such as the peripheral sympathetic nerves, a messenger molecule can have both transmitter and modulator functions. For example, norepinephrine can activate an α_1-adrenergic postsynaptic receptor to produce vasoconstriction or stimulate an α_2-adrenergic presynaptic receptor to inhibit further norepinephrine release.

Neurotrophic or *nerve growth factors* are required to maintain the long-term survival of the postsynaptic cell and are secreted by axon terminals independent of action potentials. Examples include neuron-to-neuron trophic factors

in the sequential synapses of CNS sensory neurons. Trophic factors from target cells that enter the axon and are necessary for the long-term survival of presynaptic neurons also have been demonstrated. Target cell-to-neuron trophic factors probably have great significance in establishing specific neural connections during normal embryonic development.

Transmitters within CNS

There are several classes of neurotransmitters The first neurotransmitters described were acetylcholine and norepinephrine, identified at synapses in the peripheral nervous system. Many others have since been identified, and they fall into three main classes: amino acids, monoamines, and polypeptides.

Mediators or neurotransmitters, of CNS neurons are different biologically active substances. More than 50 transmitting substances have been discovered so far. Amino acids and monoamines are collectively termed *small-molecule transmitters*. Most of them are listed in tables 3–1 and 3–2.

Earlier it was considered that in all endings of one neuron, one mediator is excreted (by Dale's principle). In the last years it has been found that two or more mediators can be located in many neurons.

By the action of mediators it is possible to divide them into ionotropic and metabolotropic.

Ionotropic mediators after interaction with cytoreceptors, the postsynaptic membranes change the permeability of ionic channels. Metabolotropic mediators show postsynaptic action by the activation of specific enzymes of the membrane. As a result in the membrane or (most often) in the cytoplasm

Table 3–1. (from Guyton, 1996). Small molecules frequently acting as transmitters

Class I	Acetylcholine
Class II	The amines: norepinephrine, epinephrine, dopamine, serotonin, histamine
Class III	The amino-acids: gama-aminobutyric acid (GABA), glycine, glutamate, aspartate
Class IV	Nitric oxide (NO)

Table 3–2. Neuropeptides, slowly acting transmitters

A. Hypothalamic-releasing hormones	Thyrotropin-releasing hormone Luteininzing hormone-releasing hormone Somatostatin (growth hormone inhibitory factor)
B. Pituitary peptides	ACTH, α-melanocyte-stimulating hormone Prolactin, luteininzing hormone Thyrotropine, growth hormone Vasopressin, oxytocin
C. Opioids	Leucine-enkephalin, Methionine-enkephalin, Dynorphin, β-endorphin
C. Peptides that act on gut and brain	Substance P, gastrine, Cholecystokinine Vasoactive intestinal peptide (VIP) Neurotensin, Insulin Glucagon
D. From other tissues	Angiotensin II Bradykinin, Carnosine Sleep peptides, Calcitonin

of a cell the so-called secondary intermediaries (secondary messengers) in turn activate cascades of endocellular processes, thus influencing the function of cells. One of the most prevailing types in neurons uses a group of proteins called G-proteins. Depending on the specific characteristic of each type of neurons four subsequent effects of the G-proteins can occur:

Opening specific ion channels through the postsynaptic cell membrane (for example potassium channels often stay open for a prolonged time).

Activation of cyclic adenosine mono-phosphate (AMP) or cyclic guanosine mono-phosphate (GMP) in the neuronal cell.

Activation of one or more intracellular enzymes.In turn, the enzymes can cause any one of many specific chemical functions in the cell.

Activation of gene transcription, which can cause the formation of new proteins within the neuron, and this can change the metabolic machinery of the cell or its structure.

By the functional properties mediators of the CNS are divided into excitative, inhibitive and modulative.

Excitative mediators can be different substances which serve as the reason for depolarisation of postsynaptic membranes. The derivative of glutamine

acids (glutamates) have the most value, substance P. Some central neurons have cholinoreceptors, i.e. contain on the postsynaptic membrane receptors which react with choline connections, for example, acetylcholine in cells. Excitative mediators can also be monoamines (norepinephrine, dopamine, and serotonin). There are bases to consider, that the type of mediator which is received in synapses, is caused not only by the properties of the endings, but also the general direction of the biochemical processes in the whole neuron.

The nature of the *inhibitive mediator* is not fully established. It is considered, that in the synapses of different nervous structures, this function can be carried out by aminoacids — glycine and GABA. GABA-ergic neurons represent the major inhibitory neurons of the CNS, whereas glycinergic neurons are found in limited numbers, restricted only to the spinal cord and brainstem. Glycinergic transmission has not been as well characterized as transmission using GABA.

Nitric Oxide and Arachidonic Acid. Recently a novel type of neurotransmission has been identified. In this case, membrane-soluble molecules diffuse through neuronal membranes and activate "postsynaptic" cells via second messenger pathways. Nitric oxide (NO) is a labile free-radical gas that is synthesized on demand from its precursor, L-arginine, by nitric oxide synthase (NOS). Because NOS activity is exquisitely regulated by Ca^{2+}, the release of NO iscalcium-dependent even though it is not packaged into synaptic vesicles.

Nitric oxide was first identified as the substance formed by macrophages that allow them to kill tumor cells. NO was also identified as the endothelial-derived relaxing factor in blood vessels before it was known to be a neurotransmitter. It is a relatively common neurotransmitter in peripheral autonomic pathways and nitrergic neurons are also found throughout the brain, where the NO they produce may be involved in damage associated with hypoxia. The effects of NO are mediated through its activation of second messengers, particularly guanylyl cyclase.

Arachidonic acid is a fatty acid released from phospholipids in the membrane when phospholipase A_2 is activated by ligand-gated receptors. The arachidonic acid then diffuses retrogradely to affect the presynaptic cell by activating second messenger systems. Nitric oxide can also act in this retrograde fashion as a signaling molecule.

Unit II GENERAL PHYSIOLOGY OF CENTRAL NERVOUS SYSTEM

Chapter 4. General principles of biologic regulation. Neural regulation. The principles of reflectory regulation

Our bodies are made up of incredibly complex and delicate materials, and we are constantly subjected to all kinds of disturbances, yet we keep going for a lifetime. It is clear that conditions and processes in the body must be closely controlled and regulated, i.e. kept at appropriate values. The types and mechanisms of physiological regulation in the body will be discussed below.

For optimal cell, tissue, and organ function in animals, several conditions in the internal environment must be maintained within narrow limits. These include but are not limited to (1) oxygen and carbon dioxide tensions, (2) concentrations of glucose and other metabolites, (3) osmotic pressure, (4) concentrations of hydrogen, potassium, calcium, and magnesium ions, and (5) temperature. Departures from optimal conditions may result in disordered functions, disease, or death.

The concept of *homeostasis* is helpful in understanding and analyzing conditions in the body. The existence of steady conditions is evidence of regulatory mechanisms in the body that maintain stability. To function optimally under a variety of conditions, the body must sense departures from normal and must engage mechanisms for restoring conditions to normal. Departures from normal may be in the direction of too little or too much, so mechanisms exist for opposing changes in either direction.

Homeostatic regulation of a physiological variable often involves several cooperating mechanisms activated at the same time or in succession. The more important a variable, the more numerous and complicated are the mechanisms

that keep it at the desired value. Disease or death is often the result of dysfunction of homeostatic mechanisms. The effectiveness of homeostatic mechanisms varies over a person's lifetime. Some homeostatic mechanisms are not fully developed at the time of birth. Homeostatic mechanisms gradually become less efficient as people age. For example, older adults are less able to tolerate stresses, such as exercise or changing weather, than are younger adults.

One of important regulative mechanism that can maintain the stable conditions is the negative-feedback control systems (Fig. 4-1). Feedback is a flow of information along a closed loop. The components of a simple negative-feedback control system include a regulated variable, sensor (or detector), controller (or comparator), and effector. Each component controls the next component. Various disturbances may arise within or outside the system and cause undesired changes in the regulated variable. With ***negative feedback***, a regulated variable is sensed, information is fed back to the controller, and the effector acts to oppose change.

Although homeostatic control mechanisms usually act for the good of the body, they are sometimes deficient, inappropriate, or excessive. Many diseases, such as cancer, diabetes, and hypertension, develop because of a defective control mechanism. Homeostatic mechanisms may also result in inappropriate actions, such as autoimmune diseases, in which the immune system attacks the body's own tissue.

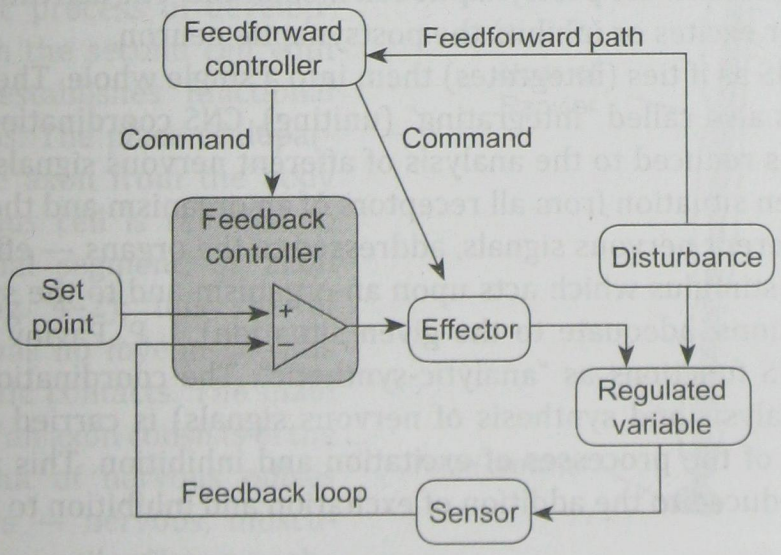


Fig. 4–1. Elements of the negative feedback control system.

I. NEURONAL REGULATION

The nervous system coordinates the activities of many other organ systems. It activates muscles for movement, controls the secretion of hormones from glands, regulates the rate and depth of breathing, and is involved in modulating and regulating a multitude of other physiological processes. To perform these functions, the nervous system relies on neurons, which are designed for the rapid transmission of information from one cell to another by conducting electrical impulses and secreting chemical neurotransmitters.

The CNS activity is most often referred to coordination or conciliatory. It is explained in that CNS was generated during evolution as the mechanism which provides the coordination of functions of an organism with the environment and among itself. It activates muscles for movement, controls the secretion of hormones from glands, regulates the rate and depth of breathing, and is involved in modulating and regulating a multitude of other physiological processes. To perform these functions, the nervous system relies on neurons which are designed for the rapid transmission of information from one cell to another by conducting electrical impulses and secreting chemical neurotransmitters. The electrical impulses propagate along the length of nerve fiber processes to their terminals, where they initiate a series of events that cause the release of chemical neurotransmitters. The release of neurotransmitters occurs at sites of synaptic contact between two nerve cells. Released neurotransmitters bind with their receptors on the postsynaptic cell membrane. The activation of these receptors either excites or inhibits the postsynaptic neuron.

Thus, the CNS as if ties (integrates) them into a single whole. Therefore, the CNS function is also called "integrating" (uniting). CNS coordination (integration) function is reduced to the analysis of afferent nervous signals which act on it in the given situation from all receptors of an organism and the synthesis of adequate efferent nervous signals, addressed to the organs — effectors (i.e. to the analysis stimulus which acts upon an organism and to the synthesis of organism reactions, adequate to the given situation). I. P. Pavlov frequently defined the CNS functions as "analytic-synthetic". The coordination function of the CNS (analysis and synthesis of nervous signals) is carried out due to the interaction of the processes of excitation and inhibition. This interaction eventually is reduced to the addition of excitation and inhibition to the central neurons.

1. Structure and functions of the neurons. The CNS has a neuronal type of structure that means that it consists of separate nervous cells or neurons. The

brain of a person contains approximately to 100 billion neurons, nearly 25 million of them are located on the periphery or connect the periphery with CNS.

Neurons communicate by a combination of electrical and chemical signaling. Generally, information is integrated and transmitted along the processes of a single neuron electrically and then transmitted to a target cell chemically. The chemical signal then initiates an electrical change in the target cell.

The neuron is the basic structural and functional unit of the CNS. It consists of bodies (soma) and plenty of branches which have a surpassing direction and specialization (Fig. 4–2). The long branches (axons) in the ontogenetic process of development reach the second cell with which it establishes functional connections. The place of departure of the axon from the body of a nervous cell is referred to as an initial segment, or axon hillock (Fig. 4–2); this site of the axon has no myelin sheaths and synaptic contacts. The main function of an axon consists of the carrying out of nervous pulses to the cells — nervous, muscular, secretory cells. Closer to the end of the axon, a thin cluster of final branches is formed — axon

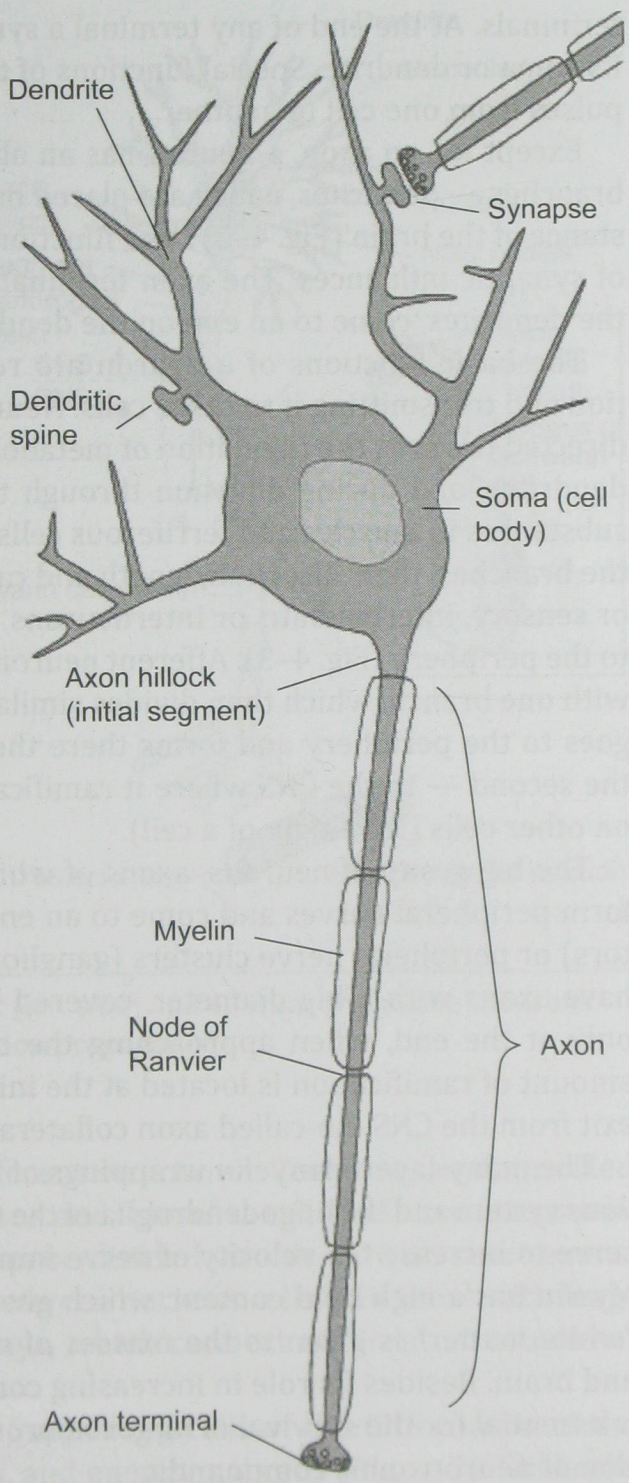


Fig. 4–2. Structural components of a neuron and a synapse.

terminals. At the end of any terminal a synapse forms from postsynaptic cells, its soma or dendrite. Special functions of the synapse consist of the transfer of pulses from one cell to another.

Except for an axon, a neuron has an abundance of short tree-like ramified branches — dendrites, which are placed mainly on the borders of the grey substance of the brain (Fig. 4–2) . The functions of dendrites consist in recognition of synaptic influences. The axon terminals, which cover the whole surface of the dendrites, come to an end on the dendrites.

The basic functions of a neuron are recognizing and processing information and transmitting it to other cells. Neurons also carry out trophic functions directed towards the regulation of metabolism and nutrition of both axons and dendrites, and during diffusion through the synapses of physiological active substances in muscles and ferriferous cells. Neurons, depending on the form of the branches, their direction, length and ramification, are classified as afferent, or sensory, intermediate, or interneurons, and efferent, which transmit pulses to the periphery (Fig. 4–3). Afferent neurons have a simple, round form of soma with one branch, which then divides similarly: one branch (modified dendrite) goes to the periphery and forms there the sensitive endings (receptors), and the second — to the CNS where it ramificates on fibers which come to an end on other cells (true axon of a cell).

The big group of neurons, axons of which fall outside the limits of the CNS form peripheral nerves and come to an end at the executive structures (effectors) or peripheral nerve clusters (ganglions), affect as efferent neurons. They have axons with a big diameter, covered by a myelin sheath and ramificates only at the end, when approaching the organ, which it innervates. A small amount of ramification is located at the initial part of the axon and it forms an exit from the CNS (so-called axon collateral).

The many-layered myelin wrappings of Schwann cells of the peripheral nervous system and the oligodendroglia of the CNS produce the myelin sheaths that serve to increase the velocity of nerve impulse conduction in axons (Fig. 4–4). Myelin has a high lipid content, which gives it a whitish colour, and the name "white matter" is given to the masses of myelinated fibers of the spinal cord and brain. Besides its role in increasing conduction velocity, the myelin sheath is essential for the survival of larger neuronal processes, perhaps by the secretion of neurotrophic compounds.

There is also a large amount of neurons In the CNS, which are characterized by soma located in the middle of the CNS and the branches do not leave it. These neurons establish a connection only with other nervous cells of the

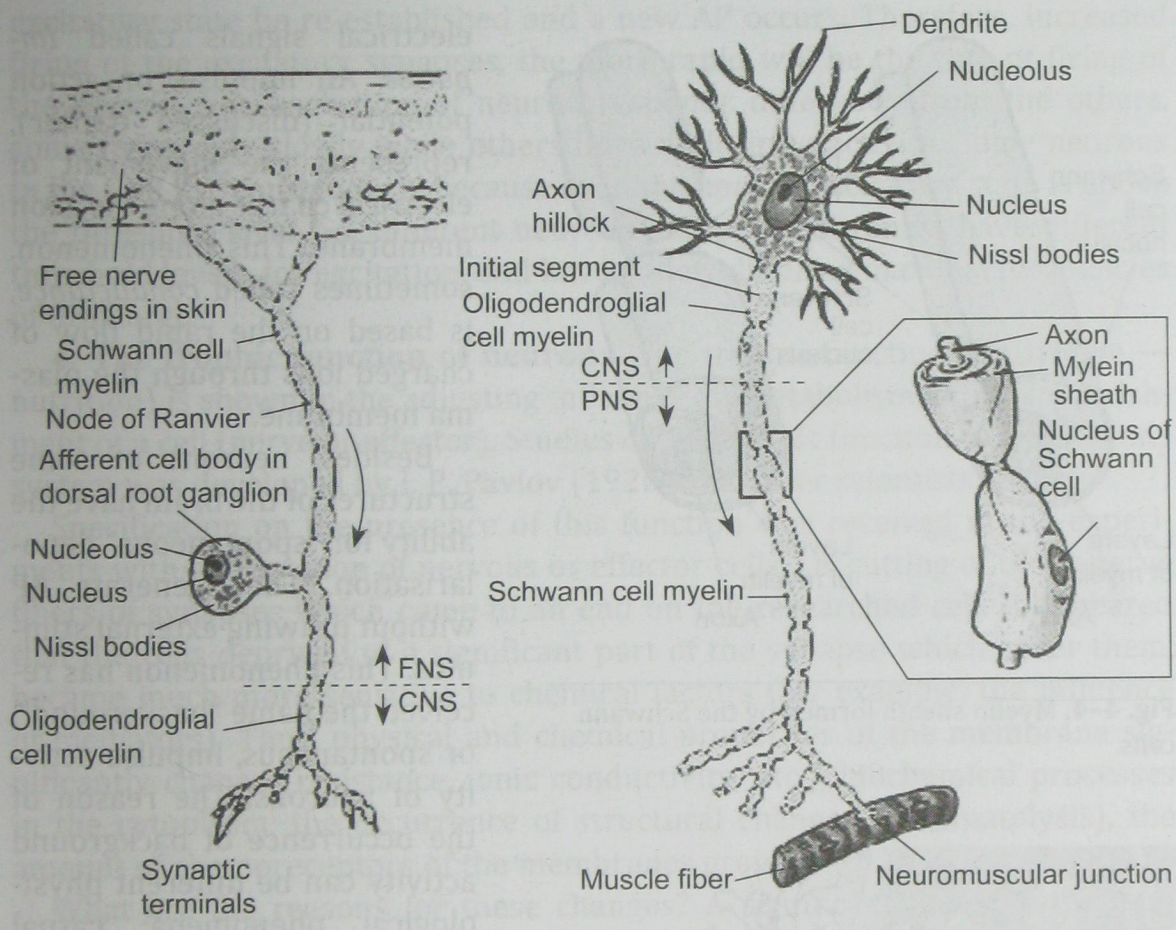


Fig. 4–3. Schematic draw of the afferent (left side) and efferent (right side) neurons. Arrows indicate the direction for excitation propagation.

CNS, instead of with sensory or efferent structures. They are, as if, inserted between afferent and efferent neurons and "close" them. These are intermediate neurons (interneurons). They can be divided into those with short-axons, which establish short connections between nervous cells, and those with long-axons — neurons of the leading ways, which connect different structures of the CNS.

2. Supporting cells of the CNS. Supporting cells of the CNS consist of the oligodendroglia, astroglia, microglia, and ependymal cells (Fig. 4–5). Glial cells look like just as nervous cells come from the primary ectoderm.

3. Background activity of a neuron. Neurons are characterized by the ability to communicate with other neurons and body cells through pulsed

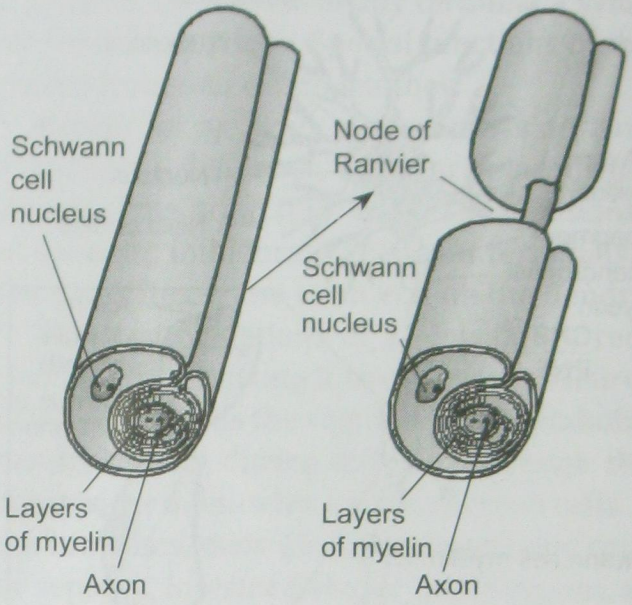


Fig. 4–4. Myelin sheath formed by the Schwann cells.

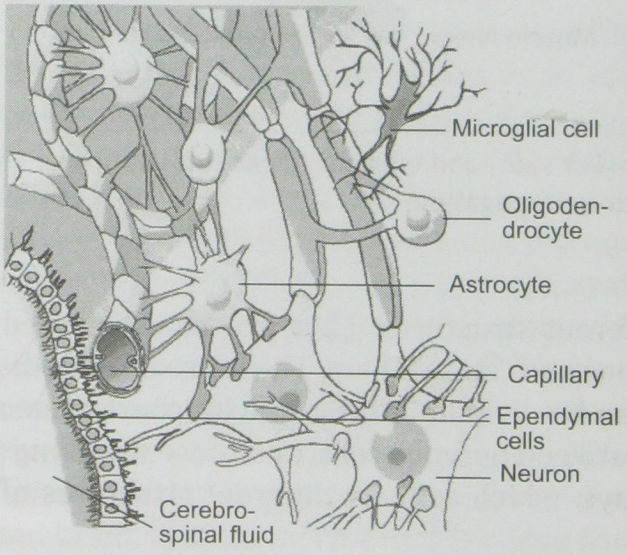


Fig. 4–5. Schematic view of relationships between the glial elements, the capillaries, the cerebral spinal fluid and the cell bodies of CNS neurons.

electrical signals called *impulses.* An impulse, or action potential (discussed earlier), represents the movement of electrical charge along the axon membrane. This phenomenon, sometimes called *conductance*, is based on the rapid flow of charged ions through the plasma membrane.

Besides, neurons of some structures of the brain have the ability for "spontaneous" depolarisation. They generate AP without drawing external stimulus. This phenomenon has received the name "background", or spontaneous, impulse activity of neurons. The reason of the occurrence of background activity can be different physiological phenomena: casual distribution of quantum by mediators into the synaptic clefts, chaotic infringement of ionic streams and metabolic processes, changes in the level of MP, etc. Background activity enables neurons both to strengthen and to inhibit excitation. For example, during a deep narcosis the level of background activity of the brain is reduced, and during an epileptic spell it sharply amplifies. The greater the number of excitatory synapses being stimulated, the more rapidly will the

excitatory state be re-established and a new AP occurs. Therefore, increased firing of the excitatory synapses, the more rapid will be the rate of firing of the neuron. Yet, each type of neuron responds differently from the others, some firing only slowly while others burn with firing activity. Some neurons in the CNS fire continuously because even the normal excitatory state is above the threshold level. So, different neurons respond differently, have different threshold levels, for excitation, and have widely differing maximal frequencies of discharge.

4. The trophic function of neurons. The trophic function (Gr. tropne — nutrition) is shown in the adjusting influence on metabolism and the nourishment of a cell (nerve or effector). Studies of the trophic function of the nervous system was developed by I. P. Pavlov (1920) and other scientists.

Specification on the presence of this function was received in the experiments with denervation of nervous or effector cells, i.e. cutting of those nerve fibers of synapses which come to an end on the researched cell. It appeared that the cells deprived of a significant part of the synapse which cover them, became much more sensitive to chemical factors (for example, the influence of mediators). Thus, physical and chemical properties of the membrane significantly change (resistance, ionic conductivity, etc.), biochemical processes in the cytoplasm, the occurrence of structural changes (chromatolysis), the amount of chemoreceptors of the membranes grows.

What are the reasons for these changes? A significant factor is the constant arrival (including spontaneous) of mediators to a cell, which adjusts the membrane processes in the postsynaptic structure; raises the sensitivity of receptors to chemical stimuli. The reason of changes can be the excretion of substances ("trophic" factors) with synaptic endings, which penetrate into the postsynaptic structure and influence it.

There are data on the moving of some substances by axons (axon transport). Protein, which is synthesized in the cell body, products of metabolism, nuclear acids, neurotransmitters and other substances, are moved by the axon to the nerve ending together with cellular organelles, in particular mitochondrion, obviously, with a full set of enzymes. It is experimentally proven that fast (410 gg in 1 period) and slow (175–230 gg in 1 period) axon transport is an active process, which demands an expense of metabolism energy. Suppose that the transport mechanism is carried out with the help of microtubules and neurofilaments of the axon with which a sliding of actinic transport of filament occurs. Thus, the decomposition of ATP, from which energy for transport is provided.

Retrograde axon transport (from periphery to the cell body) is also revealed. Viruses and bacterial toxins can penetrate the axon on the periphery and move to the cell body. For example, tetanic toxin which produces bacteria which are in a wound on the skin, gets in the organism by retrogradative axon transport in the CNS and becomes the reason of muscular spasms which can cause death. The injection to cut axons of some substances (for example, peroxidase enzyme) is accompanied by their entering through the axon and their distribution down to the neuron soma.

The solution to the problem of trophic influence of the nervous system is very important for understanding the mechanism of some trophic dissonances (trophic ulcers, hair loss, nail fragility, etc.) which are quite often observed in clinical practice.

II. THE GENERAL PRINCIPLES OF REFLECTORY ACTIVITY

The basic and specific display of the nervous system activity is the realization of reflexes. I. P. Pavlov stated that in the life of a complex organism, reflexes are the essential and most often met by nervous phenomena. With its help constant, correct and exact ratios of parts of an organism between themselves and the relation of the whole organism to its surrounding conditions are established.

The reflex is a natural reaction of an organism to internal or environmental changes, which are carried out with the participation of the CNS in reply to irritation of receptors. Due to reflexes, the organism is capable of quickly and precisely reacting to changes of external and internal environments and adapting to these changes.

1. Reflectory arch

The reflectory arch (reflex path) is a neuron circuit from the peripheral receptor through the CNS to the peripheral effector (working body).

Components of the reflectory arch are receptive receptors, afferent pathways, the "nervous centre" (central neurons), efferent pathway and effector. You could see the reflectory arch of one of the simplest reflex (myotatic reflex) on Fig. 4–6. Muscle spindle here serves as receptor from which afferent excitative pathways are started and contacted with the sensitive nuclei of the

spinal cord back horns. The presented reflectory pathway is considered to be monosynaptic.

Receptors make themselves sensor controls which perceive various changes which occur in an organism or in the environment. Distinguish extero-, proprio- and viscероreceptors. The combination of receptors, irritation which predetermines the given type of reflex, is called receptive field of reflex (reflex zone). Receptors, identical in structure, can belong to different receptor fields. Receptor fields — a set of receptors, which belong to one sensory neuron. The reflex zone contains not one but some receptor fields.

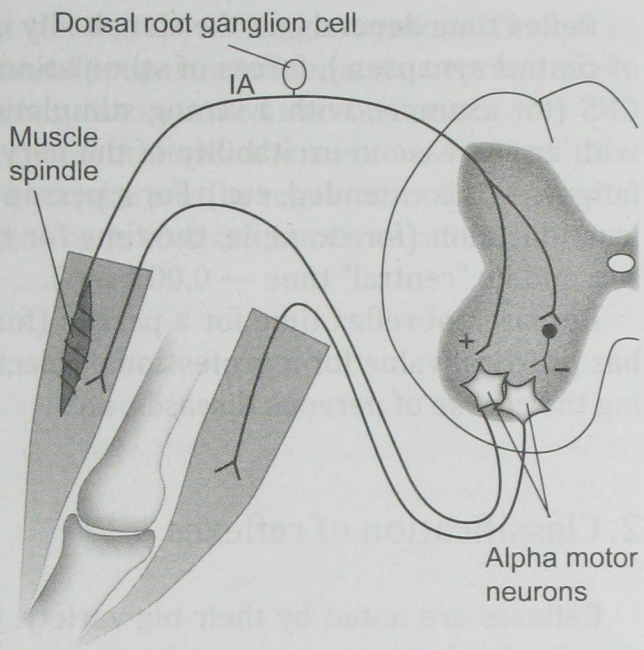


Fig. 4–6. Reflectory pathway of the myotatic reflex.

On the other hand, in the reflex zone of a certain reflex different receptors with different structures can enter. For example, irritation of the same receptors located on different sites of the skin of a frog, can result in the appearance of different reflexes (bending, extending, scratching).

Sensitive (centre-seeking) nerves, which bring excitation to the CNS, make up the afferent pathways. The centre of a reflectory arch are the structures located in the CNS, which receives, processes information, which arrives and transfers it to the periphery. Efferent pathways consist of motor, or vegetative, fibers of the nervous system. The effectors are different organs (skeletal and smooth muscles, glands, heart, etc.). Because of the presence of a feedback (inverse afferent) reflectory arches "become isolated" in a ring, therefore sometimes the term "reflex ring" is used.

Reflex reaction occurs in time: the reflex time (latent period) is the time from the beginning of receptor stimulation till the approach of the corresponding effector reaction. It is determined at times of excitation conduction by afferent and efferent pathways and in the central part of the reflectory arch (so-called central time of reflex). Thus, reflex time is the total expression of duration of these intervals.

Reflex time depends on the complexity of the reflectory arch (the amounts of central synapses), forces of stimulation and the level of excitability of the CNS (for example, with a strong stimulation it is shorter, than with a weak, with an increase in excitability of the nervous centres it is reduced, and with fatigue — it is extended, etc.). For a person the time of tendon reflexes has the least duration (for example, the time for the tendon knee reflex is 0.01-0.02 sec. and its "central" time — 0,003 sec.).

Research of reflex time for a person (for example, sensorimotor reactions) has practical value for a professional selection of transport drivers, for studying the course of nervous diseases, etc.

2. Classification of reflexes

Reflexes are noted by their big variety. By their attributes, they are conditionally divided into groups and types.

The reflex, which arch has only one central synapse (between the afferent and efferent neurons), is referred to as monosynaptic. It is an elementary kind of reflectory arch. Such reflexes are muscular extension reflexes. Tendon reflexes (knee, ulnaris, etc.) which have an important clinical value belong to this group. Apart from monosynaptic reflexes of extension, all reflectory arches include some consecutive central neurons. Such reflexes are referred to as polysynaptic. They can occur with the stimulation of muscular, skin or visceral receptors.

By the character of reaction, depending on what organs took part in it, reflexes are divided into motor (if their effector is skeletal muscles) and vegetative (if their effector is glands, blood vessels, internal organs). Reflexes are divided also into extero-, intero- and proprioreceptors — depending on where the receptors are located, stimulation which predetermines the given reflex (on the external surface of the body or in the internal organs, in the skeletal muscles, the tendons, or the joints).

All the above-named reflexes are born. They are caused by the (genetic) structural organization of the nervous system, their reflectory arches become isolated usually in phylogenetically old departments of the CNS, spinal cord, brain stem and subcortical structures. Near these reflexes each organism is capable of developing reflex reactions during its life. It helps him better and more precisely adapt to the changeable conditions of the environment. Such reflexes are called conditional. Arches of these reflexes usually become isolated in the supreme departments of the CNS, mainly in the cortical substance of the big brain.

Till now, the question was about the activity of single neurons. Now it is expedient to consider how the functional properties of separate neurons affect the activity of a combination of neurons, the so-called nervous centre.

Proceeding from physiological representations, the nervous "centre" can settle at different levels of the CNS and take part in regulation of any physiological function (breath, digestion, etc.) or in realization of any reflex.

Chapter 5. Excitation and inhibition in central nervous system. The principles of reflectory activity coordination

I. EXCITATION IN THE CNS

Excitation is connected to the development a temporary depolarisation in neurons. The depolarisation answer of a neuron to excitation has received the name excitatory postsynaptic potential (EPSP). Endocellular EPSP in neurons of mammals for the first time has been registered by J. Eccles and his co-workers (1966) with the help of microelectrode technical equipment.

Since depolarisations can lead to the excitation and activation of a neuron, they are commonly called *excitatory postsynaptic potentials* (EPSPs). In contrast, hyperpolarisations of the membrane prevent the cell from becoming activated and are called *inhibitory postsynaptic potentials* (IPSPs). These membrane potential changes are caused by the influx or efflux of specific ions (Fig. 5–1).

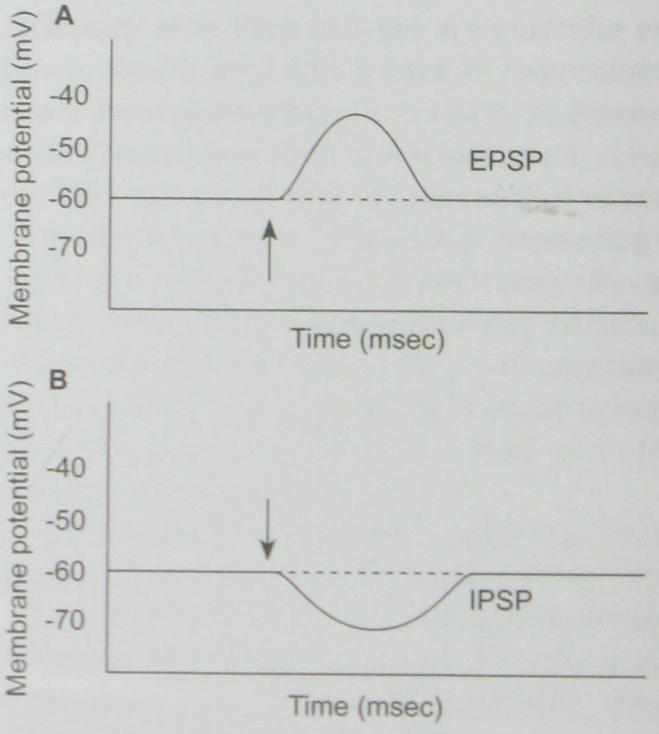


Fig. 5–1. Excitatory (A) and inhibitory (B) postsynaptic potentials.

Opposite of excitation which is distributed (AP), EPSP is local. It is local, graduation (depends on the force of inhibition, i.e. does not submit to the law "all or none"), capable of summation with previous depolarisation, does not leave refractory. Apart from its characteristics, EPSP is identical to the end-plate potential (EPP) in neuromuscular junction. EPSP reaches a maximum during

1.5-2 milliseconds, and then slowly fades behind the exponential curve. General duration of EPSP in many neurons of the CNS is close to 15 milliseconds, and the amplitude can reach — 20-30 mV with the size of membrane potential — 70 mV. The place of primary generation of AP in neurons, in the majority of cases, is the initial part of the axon (axon hillock). High excitability of this hillock is explained in that its threshold depolarisation is twice as small (15 mV) as in the soma of a neuron (30 mV). It is explained in that in a membrane of the axon hillock there are much more ionic channels. It is the most excitable site of a neuron, and only after its excitation the pulse is distributed, from one side, to the axon, and from the second — to the soma and dendrites. AP of the neuron, which is registered endocellular, has two components: first, the potential of an initial segment, or IS-potential, and then potential of the soma and dendrites — SD-potential is registered.

At the basis of the occurrence of excitation lies trans-membrane movement of ions and include the following mechanisms:

1. Opening of sodium channels to allow large amount of positive electrical to flow to the interior of the postsynaptic cell. This raises the membrane potential in the positive direction up towards the threshold level for excitation.
2. Depressed conduction through chloride or potassium channels or both. This decreases the diffusion of negatively charged chloride ions to the inside of the postsynaptic neuron or decreases the diffusion of positively charged ions to the outside. In either case, the effect is to make the internal membrane potential more positive than normally.
3. Various changes in the internal metabolism of the cell to excite cell activity or in some instances changes of the amount of excitatory.

II. PROPERTIES OF EXCITATION PROPAGATION IN THE NEURAL CENTRES

1. One way conduction. Opposite to the nervous fiber, capable of conducting excitation in both directions, central synapses distribute only in one: from the receptor through inserted neurons to the effectors. It is caused by the mediator, released by the presynaptic ending, entering the synapse clefts, operating on the receptors of the postsynaptic membranes, causing EPSP there, and then AP in the postsynaptic structure, i.e. the synapse functions as a valve.

2. Synaptic delay. In the synapses of the CNS, excitation is conducted slower than in nervous fibers. It is explained by the features of synaptic transfers (so-called synaptic decremention), namely diffusion of the mediator to the postsynaptic membrane, occurrence EPSP, its increase to critical size, generation of AP. Exact measurements have allowed the establishment, that for the conducting of excitation through one synapse elapsed time is 2-3 milliseconds. Proceeding from this, it is possible to determine the quantity of synapses in the neuron grid by which a pulse in CNS passes.

3. Transformation of the excitation rhythm. The ability of the nervous centres to change (to process) the rhythm of impulses, which enter them, is referred to as rhythm transformation. More often rhythm transformation is shown in that in reply to a single stimulus — irritation by the nervous centres send to an agency (effector) a lot of impulses which act one on one with a certain frequency. It explains also the occurrence of tetanic reductions of skeletal muscle in reply to a single stimulation of a corresponding afferent nerve. Figuratively speaking, in reply to a single shot the centre answers with a machine gun firing . The basis of this phenomenon lies mainly in traces of fluctuation of MP, in particular duration of the following negative potential. If the following negative potential which accompanies AP is big, it can reach a critical level of membrane depolarisation and cause the occurrence of the second pulse. In some cases transformation increases in duration of EPSP, and after the termination of the first AP the following develops.

4. Summation of excitation. Summation of excitation on the body of a neuron is understood as the summation of EPSP as local potential. Addition — the process of adding two or more identical reactions, in which the resulting reaction will be greater in size, than any of its compounds.

Intensity of summation of EPSP on the body of a neuron determines whether this neuron will be excited, whether facilitated (facilitation — continuous increase in excitability of a neuron in the cortical substance of the brain), and also frequencies of discharge of an excited neuron. The frequency of discharge of a neuron, i.e. AP frequency in a series which is generated by a neuron, is its major functional characteristic. It allows to judge the intensity of excitation in a certain neuron. The larger the frequency of discharges is, the stronger the excited neuron.

Temporal and spatial summations of excitation are distinguished (Fig. 5–2). If EPSP is faster (with an interval close to 15 milliseconds) acting one after another axon, then they increase in the neuron, reaching eventually a threshold level of depolarisation, necessary for generation of AP. It is referred to as temporal or sequential. If separate the stimulations of the two axons, then only

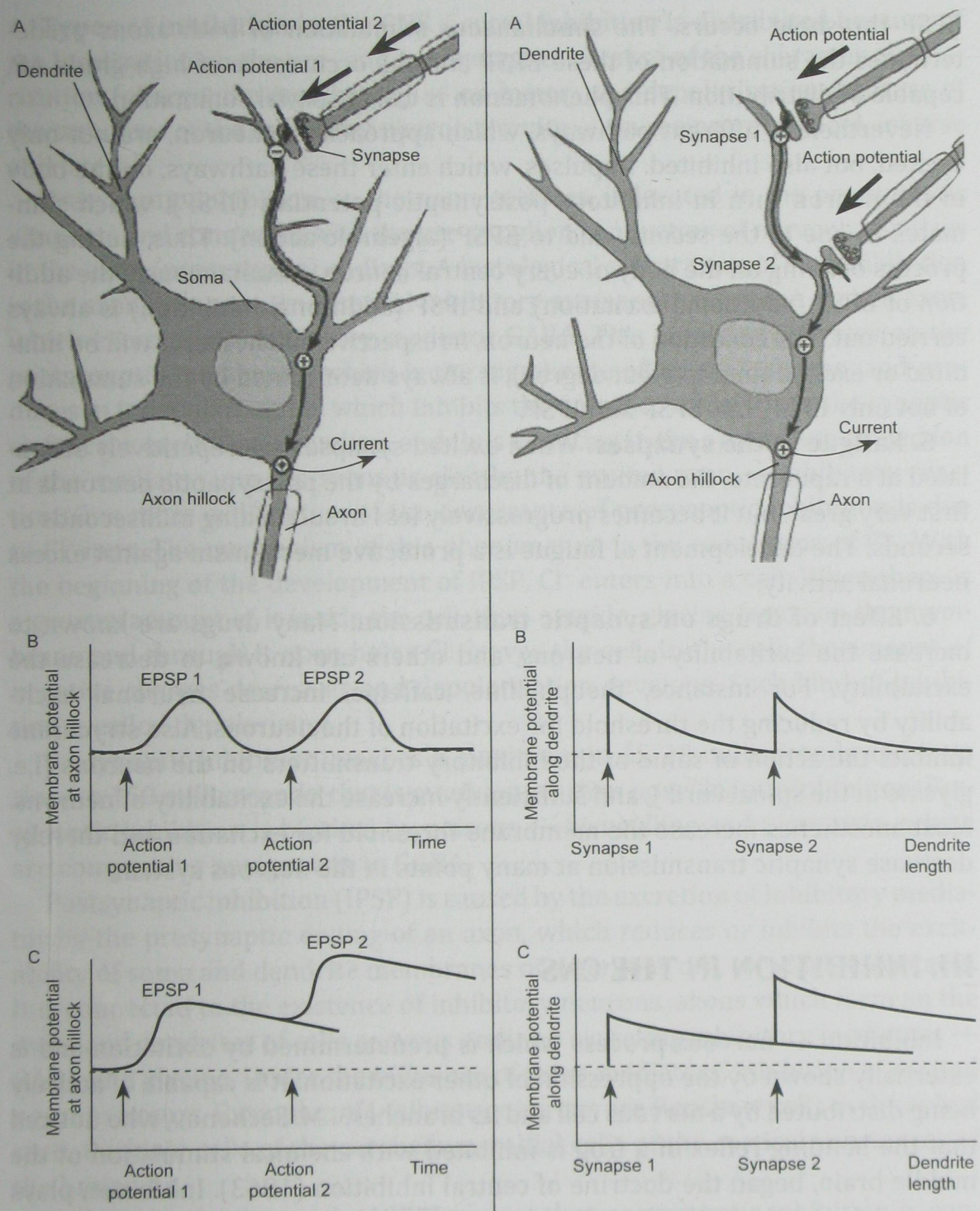


Fig. 5–2. Schematical models of the temporal (left) and spatial (right) summations.

EPSP threshold occurs. The simultaneous stimulation of both axons predetermines the summation of these EPSP and the occurrence of high-grade AP, capable of distribution. This phenomenon is called spatial summation.

Nevertheless, afferent pathways, which approach the neuron, are not only excited, but also inhibited. Impulses, which enter these pathways, on the body of the neuron turn in inhibitory postsynaptic potentials (IPSP), which summates as one to the second, and to EPSP (algebraic action). Thus, during the process of living on the body of every central neuron, simultaneously the addition of EPSP (additional excitation) and IPSP (additional inhibition) is always carried out. The condition of the neuron, irrespective of whether it will be inhibited or excited (and to what degree), is always determined by the summation of not only EPSP, but EPSP and IPSP.

5. Fatigue in the synapses. When excited synapses are repetitively stimulated at a rapid rate, the amount of discharges by the postsynaptic neuron is at first very great, but it becomes progressively less in succeeding milliseconds or seconds. The development of fatigue is a protective mechanism against excess neuronal activity.

6. Effect of drugs on synaptic transmission. Many drugs are known to increase the excitability of neurons, and others are known to decrease the excitability. For instance, theophylline, caffeine, increase neuronal excitability by reducing the threshold for excitation of the neurons. Also strychnine inhibits the action of some of the inhibitory transmitters on the neurons (i.e. glycine in the spinal cord), and sufficiently increase the excitability of neurons. Most anesthetics increase the membrane threshold for excitation and thereby decrease synaptic transmission at many points in the nervous system.

III. INHIBITION IN THE CNS

Inhibition — nervous process which is predetermined by excitation and is externally shown by the oppression of other excitation. It is capable of actively being distributed by a nervous cell and its branches. I.M. Sechenov, who noticed that the bending reflex of a frog is inhibited with chemical stimulation of the middle brain, began the doctrine of central inhibition (1863). Inhibition plays an important role in the activity of the CNS, namely: in the coordination of reflexes; in the behaviour of a person and animals; in the regulation of activity of the internal organs and systems; in the realization of the protective function of nervous cells.

Types of inhibition in the CNS. Central inhibition is distributed because of the localization on the pre- and postsynaptic; because of the character of polarization (charge of a membrane) — on hyper- and depolarisation; because of the structure of the inhibitory neuron circuits — on reciprocal or pool, inverse and lateral.

Presynaptic inhibition, as the name testifies, is located in the pre-synaptic elements and connected with the pre-ganglion conduction of nervous impulses in the axon (presynaptic) endings. A histological substratum of such inhibition is the axon synapses. An inserted inhibitory axon approaches the excited axon, which excretes the inhibitory mediator GABA. This mediator operates on the postsynaptic membrane, which is the membrane of an excited axon, and produces in it depolarisation, which inhibits the entrance of Ca^{2+} from the synaptic cleft at the excited axon ending, and thus resulting in the oppression of ejection of the mediator into the synaptic cleft by the excited axon — inhibitory reaction. One more explanation of the occurrence of presynaptic inhibition is due to Cl^- ions. The mechanism of this phenomenon is the connection of Cl^-. With the beginning of the development of IPSP, Cl^- enters into a cell. When there is a greater amount of it inside the cell, than outside, glycine forms on the membrane and through it open holes Cl^- leaves the cell. In the cell, the amount of negative charges decreases and depolarisation develops. Such kind of inhibition is called depolarising.

Presynaptic inhibition reaches a maximum in 15-20 milliseconds and lasts close to 150 milliseconds, that is much longer than postsynaptic inhibition. Presynaptic inhibition is blocked by poisons — bicuculline and picrotoxin, which are competitive antagonists to GABA.

Postsynaptic inhibition (IPSP) is caused by the excretion of inhibitory mediator by the presynaptic ending of an axon, which reduces or inhibits the excitability of soma and dendrite membranes of a nervous cell, to which it contacts. It is connected to the existence of inhibitory neurons, axons which form on the soma and dendrites of cells nervous endings, excreting inhibitory mediators — GABA and glycine. Under the influence of these mediators, inhibition of excited neurons occurs. Examples of inhibitory neurons are Renshaw cells in the spinal cord, Purkinje cells of the cerebellum, radial cells of the cortical substances of the brain, etc.

P. G. Kostyuk's research (1977) proves that postsynaptic inhibition is connected to the primary hyperpolarization of the soma membrane of a neuron, on its basis lays the increase in permeability of postsynaptic membranes for K^+, and opening of chloride channels (and will allow negatively charged chloride

ions to move towards the interior), which will make the MP more negative than normal. Due to hyperpolarization, the level of MP keeps away from a critical (threshold) level. That is, it increases — the degree of intracellular negativity which is called hyper-polarization. This results in the inhibition of a neuron. Such kind of inhibition is referred to as hyperpolarizing.

The amplitude and polarity of IPSP depends on the ascending level of MP of the neuron.(more often ≈ –5 mV).

Postsynaptic inhibition is local. It develops gradually, capable of summation, and does not leave after itself refractory. There is a more operative, precisely addressed and universal inhibitory mechanism. "Central inhibition" was described by Ch. Sherringtone (1906). Depending on the structure of the inhibitory neuron circuits, such forms of postsynaptic inhibition were distinguished: reciprocal, reverse and lateral, which is a version of reverse.

Reciprocal inhibition is characterized by, in the event with the activation of afferent, for example, motoneurons of the flexor muscles are excited; simultaneously (on this side) motoneurons of the extensor muscles are inhibited, which operate on the same joint. This occurs because afferent from the muscular spindles form excited synapses on the motoneurons of the agonist muscles, and through the intermediary inserted inhibitory neuron — inhibitory synapses on the motoneurons the antagonist — agonist muscles. From the physiological point of view such inhibition is very advantageous as it facilitates the movement of a joint "automatically" without additional voluntary or not free control.

Inverse inhibition. In this case from the axon of the motoneurons departs one or several collaterals, which go to the inserted inhibitory neurons, for example, Renshaw cells. In turn, Renshaw cells form inhibitory synapses on the motoneurons. In case of excitation of the motoneuron and Renshaw cells, followed by hyperpolarization of the motoneurons membrane, its activity is inhibited. That which excites motoneurons better, inhibits the influences through Renshaw cells. Thus, inverse postsynaptic inhibition functions by the principle of negative inverse linkage. There is an assumption, that this kind of inhibition is necessary for self-regulation of excitation of neurons and also for the prevention of their over-excitation and reactions.

Lateral inhibition. The inhibitory circuit of neurons influences not only the excited cell, but also the next neurons in which excitation is weaker or absolutely absent. Such inhibition is referred to as lateral, because the site of inhibition is located to the side (lateral) of the excited neurons. It will play an important role in the sensory systems, creating the phenomenon of contrast.

Postsynaptic inhibition is easily removed at the introduction of strychnine, which competes with the inhibitory mediators (glycine) on the postsynaptic membrane. Tetanic toxin also depresses postsynaptic inhibition, increasing the release of mediators with postsynaptic inhibitory endings. Therefore, the introduction of strychnine or tetanic toxin is accompanied by a spasm which occurs as a result of a sharp increase of the excitation process in the CNS, motoneurons.

In connection with the disclosing of ionic mechanisms of postsynaptic inhibition, the opportunity has come up for an explanation of the action mechanism of Br. Sodium bromide in optimum dozes is widely applied in clinical practice as a sedative agent. It is proved that such effect of sodium bromide is connected with the increase of postsynaptic inhibition in the CNS.

The role of different kinds of central inhibitions. The leading role of the central inhibitions is that with the interaction of the central excitation providing the opportunity to analysis and synthesis of nervous signals in the CNS, and also, to coordinate all functions of an organism between itself and with the environment. This role of the central inhibitions is called coordination. Some kinds of the central inhibitions not only coordinate, but also protect (guard). Assume that the basic coordination role of presynaptic inhibition will consist in the oppression of afferent signals in the CNS. Due to direct postsynaptic inhibition, the activity of the antagonistic centres is coordinated.

Inverse inhibition, limited by the greatest frequency possibility of motoneuron discharge of the spinal cord, also carries out coordination role (coordinates the maximal frequency of motoneuron discharges with the reduction speed of muscular fibers which they innervate) and protective (prevents motoneuron excitation). For mammals, this kind of inhibition is widespread basically in spinal afferent systems. In the higher departments of the brain, namely in cortical substance of the brain, dominates postsynaptic inhibition.

Integrative function of the central neurons. Depending on the summation of the postsynaptic effect, which is caused by the impulses acting in the neuron at the present, excitation, which is distributed (AP), occurs or does not occur. Any neuron is an original "appraiser" of all signals convertible on it. As if it "decides" what to transfer to the following neural circuits and whether to transfer something at all. If it is examined on a simple example integration activity if the neuron "answers" by the principle "true-false", then in effect, at the level of the neuron (for example, the spinal cord) the problem is solved, whether the neuron will take a higher level of participation in the further "consideration" of the information which proceeds from the periphery.

With the characteristic of properties of the central neurons, it is important to take into account also the general condition of every neuron separately, which determines the character of its participation in the interaction of many neurons of a person's brain (excitation or inhibition, relief or depression, positive or negative after-action, etc.).

The condition of every neuron depends on the amount of synapses (excited or inhibited) on its surface and also on the amount of functional synapses at present and the intensity of their functioning. The total result of these influences the neuron and its change in directions of de- or hyperpolarization is determined, which eventually leads to the generation of AP with a certain frequency or the absence of this AP generation by the neuron.

Taking as the initial amount of neurons of the cortical substance of a person's brain 14 billion and taking into account that every neuron has on the average 5 thousand synaptic contacts to others neurons of the brain, it is hard even with having rich imagination to count up the amount of combinations of interaction neurons in the brain. Made in this direction, mathematicians' calculations gave a fantastic figure. It appeared that with the number of such combinations the condition of neurons (that is the degree of freedom) considering all those changes which were it is looked at earlier, can be expressed in a unit with such amount of zeros that they can be measured to be 9,500,000 km (!). It is enough only to imagine this "amount" to estimate grandiose reserves of brain activity of a person. Thus, every separate neuron potentially has a huge amount of degrees of freedom as the object of interaction with other neurons.

The universal factor of integrative activity of a neuron is the convergence of excitation on it. Thanks to this phenomenon in the same neuron, simultaneously, numerical streams of excitation of different sensorial and biological modality, which then are subject to complex processing and code conversion, form into uniform excitation which acts on axons.

Recently, rather frequently attempts have been made to draw an analogy between the activity of the CNS and the work of a computer. Neurons with a synaptic device are considered to be like a closet, in which logic operations, arithmetic actions, mathematical integration and differentiation can be carried out.

Processing the incoming information on neurons are connected both with synaptic processes and with a complex of chemical processes, which occur in the neuron (chemical integration), and also with the activization of nuclear genetic devices (genomes) of a neuron. In reply to the arrival of nervous impulses and the action of mediators, the synthesis of RNA and so-called specific brain fibers

increases in neurons, which modulates the sensitivity of neurons to chemical regulator (mediator, hormone, ion, prostaglandin, etc.), increases or decreases the nervous impulses. In reply to the arrival of nervous impulses and the action of mediators the synthèsis of RNA and so-called specific brain fibers increases in neurons which modulates the sensitivity of neurons to chemical regulators (mediator, hormone, ion, prostaglandin, etc.), increases or decreases the nervous impulses frequency

It is necessary to note, that in neurons genetic activity is highly expressed, than in the majority of other cells of the organism (the liver, spleen, kidneys, etc.). The organized activity of the genetic device of those departments of the brain which are greatly responsible for the integration function of the CNS (for example, volume of transcription DNA/RNA in frontal sites of brain) is more complex. In ontogenesis the measure of maturing of the brain and training genetic activity of the neurons gradually increases.

Different types of neurons are assigned different integrative abilities. In some the transmitting function prevails, that is the information is transferred by nervous channels without changes. It, in particular, concerns the so-called relay neurons with long axons and associative neurons of different departments of the CNS. At the same time mechanisms which adjust the excitability of neurons, can take part in the system integrative activity if the information is processed during its transfer from one link of the neuron grid to another.

All kinds of complex integrative activity of the CNS (the restriction of afferent stream with impulses, which goes from the receptors; inverse positive and negative connections which strengthen or limit the activity of a previous link of the neurons system; pre- and postsynaptic inhibition, which specifies afferent stream; the simplification or blocking of synaptic transfers, etc.) is carried out with the help of various mechanisms of excitation regulation and conduction opportunities of channels of nervous connection.

IV. PRINCIPLES OF REFLEXES COORDINATION

Interaction of neurons and nervous processes (excitation and inhibition), which provides the coordinated activity of the CNS, is why the organism carries the name of coordination of reflex activity.

Due to coordination, it is possible for the exact performance of different complex reflex acts, to which motor, secretory, vascular and other components belong. Coordination occurs in all departments of the CNS, in any nervous

centre. The main principles of reflex coordination are the following: reciprocal (incorporated) innervation; general final way; principles of feedback and dominant relations ('dominanta'), irradiation of excitation, convergence and divergence. Even the phenomena of simplification and occlusion take part in reflex coordination.

1. *Reciprocal innervation.* The basis of this phenomenon is the representation that excitation of the centre of one group of muscles (for example, motoneuron "pool" of muscles — flexor) is accompanied by reciprocal (incorporated) inhibition of the centres of muscles-antagonists (extenders). Simultaneously, on the opposite side (in symmetric endings) the centres of the extenders are raised and inhibited motoneurons muscles — flexors. Only for this condition is it possible for exact movement of flexing or extending. Reciprocal relations between groups of motoneurons and inserted neurons are the basis for so-called "rhythmic reflexes", which are characterized by correct watches of the opposite by functional value of muscular contraction, for example, flexing and extending. An example of such rhythmic reflexes in a mammal can be sneezing and step-by-step reflexes which occur in general during stimulation of receptor fields by long and weak stimulation.

The mechanism of reciprocal relations, consisting of afferent nervous fibers (for example, groups Aα of the muscular spindles) lets out branches in the spinal cord. One of the branches forms exciting synapse on the motoneurons of the flexor muscles and others with the help of inserted neurons inhibit synapses and motoneurons of the extensor muscles. Therefore, the stimulation of afferent nerves cause simultaneous excitation of the motoneurons of the flexor muscles and inhibition of the motoneurons of the extensor muscles. Reciprocal inhibition has the shortest reflectory arch known.

Nevertheless, it would be a mistake to count that between centres there are for ever fixed antagonistic relations. They have a dynamic character which can be determined by different conditions. There are no reasons to limit reciprocal relations at the site of innervations by anatomic anastomosis. In the same reciprocal relations others can enter, separated from the centres, if they regulate certain conditions of any organ or system. Similar reciprocal relations are observed between the centres of inhalation and exhalation, famine and saturation, breath and swallowing and etc. However reciprocal relations between the different centres are not constants, they can change (stop) under the influence of the cortical and subcortical structures of the brain.

2. *Principle of the general final way.* Ch. Sherrington (1906) entered the concepts about motoneurons as the general final way of the motor system, is a

way by which reprocessing and integration of all numerous and diverse excitation, which simultaneously enters it, in uniform (axon), regenerate excitation which goes to the effectors (muscles). The structural basis of the general final way is the significant excess of afferent "entrances" above efferent "exits" the presence of original "funnels", an expanded amount of afferent fibers, which converge on motoneurons, interneurons, and the narrowed part of the general final way motoneuron.

The neuron integrates like excitation and inhibition processes. AP occurs only in the case of the exciting phenomenon, that is EPSP at the axon hillock site as a result of summation reaches the critical level. In other words, these influences compete for mastering by the general final way. The result of this "struggle" for the general final way depends on the force of working stimulations, their biological value in the given situation, and the condition of the nervous centres, in particular dominant. Reflexes, which arches have the general final way, can mutually amplify (in case of so-called allied reflexes) or weaken (in case of antagonistic reflexes).

3. *Principle of dominant relations ('dominanta').* The known Russian physiologist A. A. Ukhtomsky (1875–1942) formulated the principle of dominant relations (dominanta), which is one of main principles of the CNS activity.

'Dominanta' — the prevailing in the CNS of excitation, which change and "subordinate" to itself, at present, the activity of other centres. The 'dominanta' is formed under the influence of various adequate stimuli of external and internal environments of an organism (an exhaustion nutritious substance, accumulation of sexual hormones in blood, etc.). All these stimuli raise the excitability of the corresponding nervous centres (digestive, sexual, etc.). Because of summation with extraneous (indifferent) stimulus of a different modality in the given centre, resistant and long excitation, incorporated with the inhibition of others is formed. So digestive, sexual, defensive and other 'dominants', which appreciably determine the realization of the given reflex, behavioural changes of the person and animals occurs.

Is it possible to consider any excitation which has arisen in the CNS as dominant? What are the features of a 'dominanta'? The dominant focus of excitation (any centre) could be characterized by such properties:

1) neurones hyperexcitability;
2) stability of excitation;
3) ability of excitation summation (to accumulate excitation due to extraneous stimulations);
4) ability to inhibit other nervous centres and reflectory systems;

5) ability to keep the excitation till the end of stimulation, which it has caused as 'dominant's' inertia.

Examples of 'dominanta' can be the following. Irritation of animals in some sites of the cortical substance of the brain, which is usually accompanied by bending of ends, during the act of defecation or urination such reaction does not occur. On the contrary, there is an amplification of these reflex acts. It is predetermined in that at present the centres of defecation or urination are in the condition of dominant excitation which "draws" to itself extraneous excitation and shows the incorporated inhibition of other nervous centres.

The dominant condition can occur under the influence of hormones which raise the excitability of the corresponding centres. For example, for a male frog during the spring period the sensitively of the sexual reflex amplifies under the influence of extraneous (for the sexual centre) irritations. At the same time, there is no protective bending reflex of this ending.

The phenomena, which is a reminder of the 'dominanta', has to be watched in clinical practice if the source of excitation which has arisen in the CNS, similar to a magnet, "draws" to itself various extraneous irritations, increasing at their expense. So, with an ulcer of the stomach and the duodenal in a stage of dominant excitation, a food centre is found, After a trauma of the nervous stem quite often in a condition of dominant excitation, there are centres of painful sensitivity in thalamo-cortical structures of the brain, increasing the pain in the wounded ending with different extraneous irritations (strong itching, etc.) dominants can be terminated in different ways: 1) due to the elimination of irritation or satisfaction of needs (biological, social) on the basis of which the 'dominanta' was generated; 2) with the occurrence in the CNS of a stronger 'dominanta', which inhibits the one there is; 3) due to the development in the dominant centre of secondary inhibition.

Due to the dominant the CNS is not stable, but dynamic coordination relations between the nervous centres and reflex systems are created. Dominants will play an important role in the mental activity of a person and animals, determining behavioural acts at a certain moment.

4. *Simplification of reflexes.* If the irritation of two receptor fields are so weak that each of them (although excites efferent fibers) will not cause the discharge of neurons and the reflex reply, but with their irritation simultaneously there can be a precise reflex reaction which will exceed beyond the intensity of the reply to any of the separated stimulations. This phenomenon received the name of simplification of reflexes. In its basis lays summation of EPSP with

spatial summation, which occurs in neurons of the subthreshold zones of the nervous centre and makes the occurrence of AP in them.

5. *Reflex occlusion.* With simultaneous strong irritations of two afferents the reflex reply can be significantly smaller than the sum of replies to any of the irritations, or identical to it. This phenomenon of reply reduction received the name occlusion (corking), and it does not have anything in common with inhibition. The basis of this phenomenon is the generality (blocking) of the discharge zones of two or several afferent entrances to the nervous centre. So, during irritation of a nervous fiber A if discharges 11 neurons, and during irritation of a nervous fiber — 9 neurons, 5 of which have already stretched in the impulse reaction. Therefore, the general number of neurons, which discharges, will be during simultaneous irritation of two afferent entrances less than the arithmetic sum of neurons which generate AP on any of the irritations separately. So, putting irritation on two nervous fibers (A and B) serves as the reason of impulse reaction only in 16 neurons, instead of in 20 as it should be with simple summation activity of the neurons, which are discharged on any irritations (A or B).

6. *Principle of feedback.* In studying the role of "feedback" or "inverse afferent" in mechanisms of coordination of reflex activity were engaged by P.K. Anokhin (1945), N.A. Bernstein (1966) and Ch. Cherrington (1906). For more information, read, please, Chapter # 4.

The essence of this principle of feedback is, that at the realization of any reflex act, the process is not limited to the working effect (for example, muscular movement), and accompanied by the excitation of receptors of the working organ (muscle, etc.) from which the information about the consequences action brings afferent conductors to the CNS, there it is compared to the programmed, expected result. Therefore, afferent signals which act in the CNS from receptors of the working organs, which belong to this or that reflectory arch, received the name of inverse, or secondary, afferent.

Unit III ROLE OF CNS IN REGULATION OF SOMATIC FUNCTIONS

Chapter 6. Organization and control of motor movements. Tone of skeletal muscles. The pyramidal and the extrapyramidal systems. Spinal cord, basal ganglia

Effective motor function requires that muscles move and that the mechanics of their movement be programmed in a manner that provides smooth and coordinated movement. In some cases, purposeless and disruptive movements can be almost as disabling as relative or complete absence of movement.

I. ORGANIZATION OF MOTOR MOVEMENTS

As with other parts of the nervous system, the motor systems are organized in functional hierarchy, with each concerned with levels of function (Fig. 6–1). The highest level of function, which occurs at the level of the frontal cortex, is concerned with the purpose and planning of the movement. The lowest level of the hierarchy occurs at the level of the spinal cord, which contains the basic reflex circuitry needed to coordinate the function of the motor units involved in the planned movement. Several anatomically distinct pathways project in parallel to the spinal cord from the higher motor centres. Above the spinal cord is the brain stem, and above the brain stem is the cerebellum and basal ganglia, structures that modulate the actions of the brain stem systems. Overseeing these supraspinal structures are the motor centres in the cerebral cortex.

The motor neurons in the spinal cord and cranial nerve nuclei, plus their axons and muscle fibers, constitute the *final common path*, the route by which

all central nervous activity influences the skeletal muscles. The motor neurons located in the ventral horns of the spinal grey matter and brainstem nuclei are influenced by both local reflex circuitry and by pathways that descend from the brainstem and cerebral cortex.

The brainstem-derived pathways include the rubrospinal, vestibulospinal, and reticulospinal tracts; the cortical pathways are the corticospinal and corticobulbar tracts. Although some of the cortically derived axons terminate directly on motor neurons, most of the axons of the cortical and the brainstem-derived tracts terminate on interneurons, which then influence motor neuron function. The outputs of the basal ganglia of the brain and cerebellum provide fine-tuning of cortical and brainstem influences on motor neuron functions.

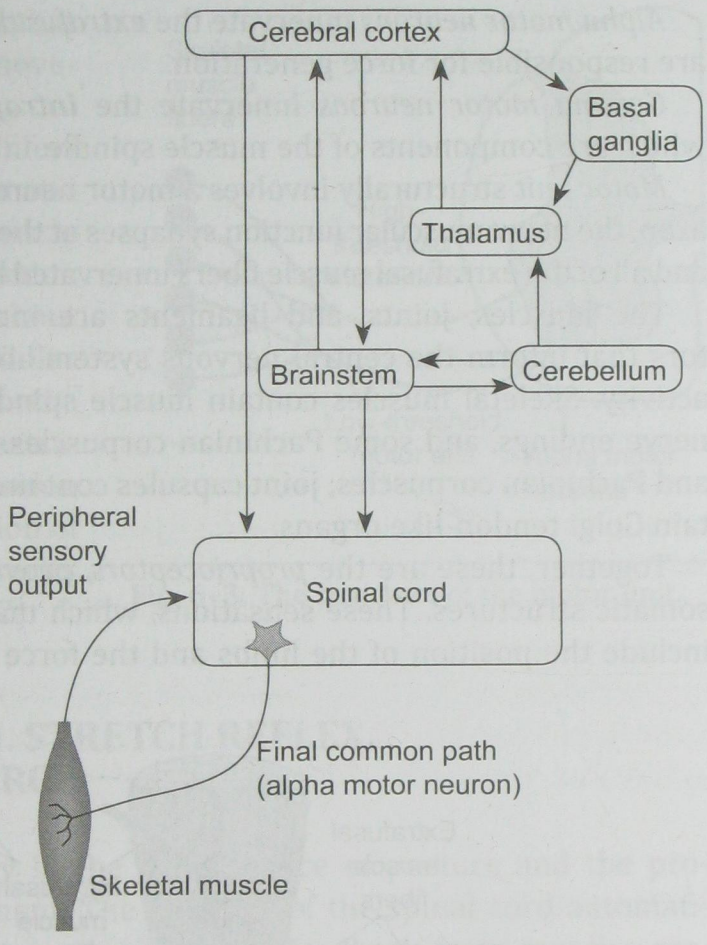


Fig. 6–1. Principal scheme of the motor system organization.

II. MOTOR UNIT

The major effects of the elaborate processing of movement information that takes place in the brain has to do with contraction of skeletal muscles. The neurons that control skeletal muscle contraction are referred to as *motoneurons* or sometimes as *alpha motoneurons.*

Motor neurons segregate into two major categories, alpha and gamma.

Alpha motor neurons innervate the ***extrafusal muscle fibers*** (Fig. 6–2), which are responsible for force generation.

Gamma motor neurons innervate the ***intrafusal muscle fibers*** (Fig. 6–2), which are components of the muscle spindle.

Motor unit structurally involves a motor neuron, its axon, the branches of the axon, the neuromuscular junction synapses at the distal end of each axon branch, and all of the extrafusal muscle fibers innervated by that motor neuron (Fig. 6–3).

The muscles, joints, and ligaments are innervated with sensory receptors that inform the central nervous system about body position and muscle activity. Skeletal muscles contain muscle spindles, Golgi tendon organs, free nerve endings, and some Pachinian corpuscles. Joints contain Ruffini endings and Pachinian corpuscles; joint capsules contain nerve endings; ligaments contain Golgi tendon-like organs.

Together, these are the *proprioceptors,* providing sensation from the deep somatic structures. These sensations, which may not reach a conscious level, include the position of the limbs and the force and speed of muscle contrac-

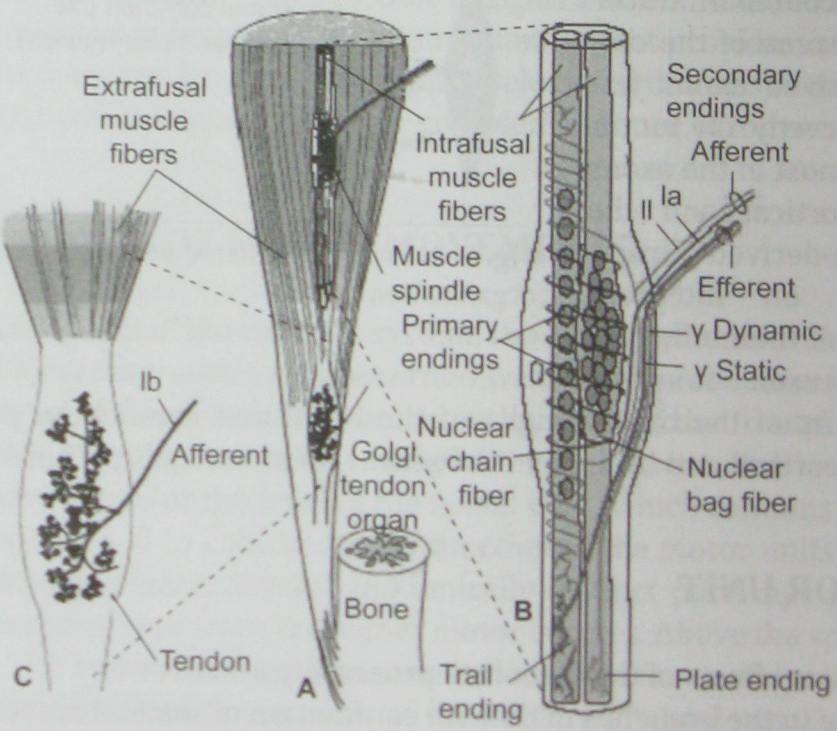


Fig. 6–2. The structure of the extrafusal (A) and intrafusal (B) muscle fibers together with Golgi tendon (C) within muscle spindles.

tion. They provide the feedback that is necessary for the control of movements.

Muscle spindles provide information about the muscle length and the velocity at which the muscle is being stretched. Golgi tendon organs provide information about the force being generated. Spindles are located in the mass of the muscle, in parallel with the extrafusal muscle fibers. Golgi tendon organs are located at the junction of the muscle and its tendons, in series with the muscle fibers (Fig. 6–2).

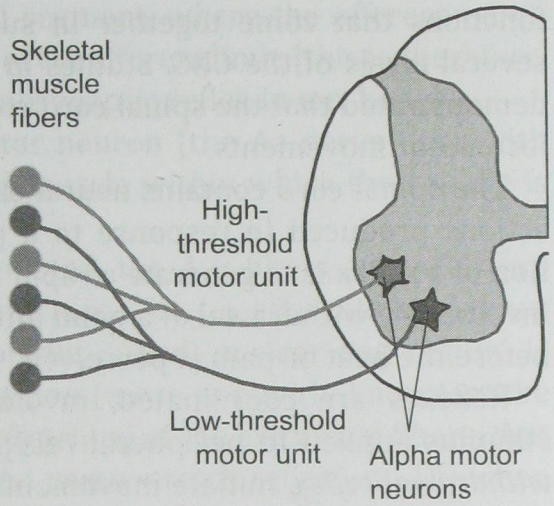


Fig. 6–3. The structure of the motor unit.

III. SPINAL MECHANISM. STRETCH REFLEX. THE REFLECTORY ARCH

Muscles interact extensively in the maintenance of posture and the production of coordinated movement. The circuitry of the spinal cord automatically controls much of this interaction. Sensory feedback from muscles reaches motor neurons of related muscles and, to a lesser degree, of more distant muscles. In addition to activating local circuits, muscles and joints transmit sensory information up the spinal cord to higher centers. This information is processed and can be relayed back to influence spinal cord circuits.

The cell bodies of the spinal cord motor neurons are grouped into pools in the ventral horns. The number of motor neurons that control a muscle varies in direct proportion to the delicacy of control required. The motor neurons are organized so that those innervating the axial muscles are grouped medially and those innervating the limbs are located laterally.

For locomotion, muscle action must occur in the limbs, but the posture of the trunk must also be controlled to provide a foundation from which the limb muscles can act. For example, when a human takes a step forward, not only must the advancing leg flex at the hip and knee, the opposite leg and bilateral truncal muscles must also be properly activated to prevent collapse of the body as weight is shifted from one leg to the other. Responsibility for the different

functions that come together in successful locomotion is divided between several levels of the CNS. Studies in experimental animals, mostly cats, have demonstrated that the spinal cord contains the capability for generating basic locomotor movements.

The spinal cord contains neural circuitry to generate *reflexes*, stereotypical actions produced in response to a peripherally applied stimulus. One function of a reflex is to generate a rapid response. A familiar example is the rapid, involuntary withdrawal of a hand after touching a dangerously hot object well before the heat or pain is perceived.

Reflexes are coordinated, involuntary motor responses initiated by a stimulus applied to peripheral receptors. Some reflexes, such as the *flexion-withdrawal reflex*, initiate movements to avoid hazardous situations; whereas others, such as the *stretch* or *crossed-extensor reflex*, serve to integrate motor movements so they function in a coordinated manner.

The anatomic basis of a reflex consists of (1) an afferent neuron, (2) the connection or synapse with CNS interneurons that communicate with the effector neuron, and (3) the effector neuron that innervates a muscle. Reflexes are essentially "wired into" the CNS so that they are always ready to function; with training, most reflexes can be modulated to become parts of more complicated movements.

1. Stretch Reflex

The *myotatic* (from the Greek "myo" means "muscle," "tatic" means "stretch") or stretch reflex controls muscle tone and helps maintain posture. Stretch reflexes can be evoked in many muscles throughout the body and are routinely tested (*e.g.*, knee-jerk reflex) during the clinical examination for the diagnosis of neurologic conditions. The period between the onset of a stimulus and the response, the *latency period*, is on the order of 30 msec for a knee-jerk reflex in a human. This response, called the *muscle stretch reflex*, is due to monosynaptic circuitry, where an afferent sensory neuron synapses directly on the efferent motor neuron (Fig 4–6).

The afferent limb of the stretch consists of:

- the receptor, called, the muscle spindle; from the central portion of the spindle arises;
- the afferent nerve belonging to I-α category of fibers (i.e., broad and fast conducting).

The center of reflex is the spinal cord segment, where the afferent nerve synapses directly with the nerve cell soma of the efferent limb; [this is, therefore, a monosynaptic reflex and is the only monosynaptic reflex in our body].

Efferent limb consist of (a) lower motor neuron (the Aa nerve fiber with its soma in the anterior horn) and the (ii) muscle within which the spindle is situated.

When the afferent (I-α) nerve is stimulated, the A-α fibers are reflectory stimulated and as a result the muscle contracts. The I-α afferents are stimulated when the central region of the spindle, called 'nuclear bag region' is stretched. The stretching of the central region (more popularly known as the 'equatorial zone') is caused by either (i) stretching of the extrafusal fibers, that is, when the muscle is stretched (hence the name stretch reflex), or, (ii) when the intrafusal fibers contract.

The myotatic reflex performs many functions. At the most general level, it produces rapid corrections of motor output in the moment-to-moment control of movement. It also forms the basis for postural reflexes, which maintain body position despite a varying range of loads and/or external forces on the body.

Disorders of muscle tone caused by dysregulated function of the stretch reflex are seen in persons with conditions such as spinal cord injury and stroke.

2. The Flexor Withdrawal Reflex

Cutaneous stimulation — such as touch, pressure, heat, cold, or tissue damage –can elicit a *flexor withdrawal reflex.* This reflex consists of a contraction of flexors and a relaxation of extensors in the stimulated limb. The action may be accompanied by a contraction of the extensors on the contralateral side. The axons of cutaneous sensory receptors synapse on interneurons in the dorsal horn. Those interneurons act ipsilaterally to excite the motor neurons of flexor muscles and inhibit those of extensor muscles. Collaterals of interneurons cross the midline to excite contralateral extensor motor neurons and inhibit flexors (Fig. 6–4).

There are two types of flexor withdrawal reflexes: those that result from innocuous stimuli and those that result from potentially injurious stimulation. The first type produces a localized flexor response accompanied by slight or no limb withdrawal; the second type produces widespread flexor contraction throughout the limb and abrupt withdrawal. The function of the first type of reflex is less obvious, but may be a general mechanism for adjusting the move-

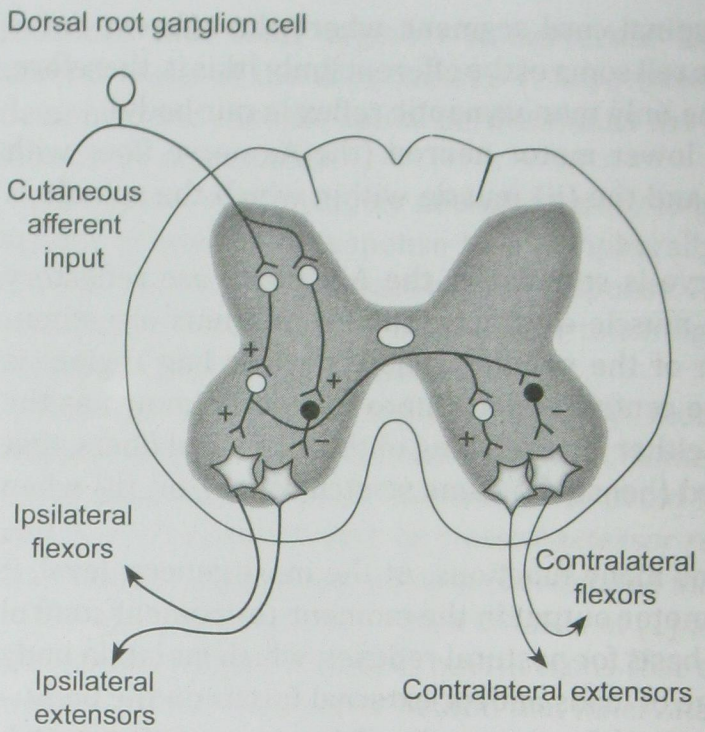


Fig. 6–4. Reflectory arch of the flexor withdrawal reflex.

ment of a body part when an obstacle is detected by cutaneous sensory input. The function of the second type is protection of the individual. Collectively, these reflexes provide for stability and postural support (the myotatic and inverse myotatic) and mobility (flexor withdrawal). The reflexes provide a foundation of automatic responses on which more complicated voluntary movements are built.

IV. SUPRASPINAL MECHANISM

From the works of Lindsay, Schreiner and Magoun, on the basis of their experiments on cat's brain in the 1940s, it may be decided that from cerebral cortex to the lowest level of medulla there are several areas, which give rise to facilitatory or inhibiting fibers. These fibers, following multisynaptic pathways, culminate into the extrapyramidal paths mentioned above.

In the diagram, + signs indicate facilitatory areas (=areas, whose discharge cause increase of tone) and — signs the inhibitory areas. Note, that, in the brain stem (midbrain + pons + medulla) there are two facilitatory areas (in addition to the vestibular nucleus) in the pons (nuclei pontis). In the lower part of the medulla there is an inhibitory center. The facilitatory areas are 'intrinsically active' that is the freed from all other connections, they will continue to discharge facilitatory impulses to the spinal center. They can be suppressed by inhibitory influences, reaching them from other areas of the brain.

On the other hand, the inhibitory center (of medulla) becomes active only when it receives impulses from the cerebral cortex or the cerebellum. When these connections are lost [as (partly) in decerebrate rigidity], its inhibitory discharges disappear.

Cerebral cortex is inhibitory and sends its inhibitory influences to the (i) medullary inhibitory center and to (ii) basal ganglia. In the cat, the cerebellum is also inhibitory and the cerebral influences are projected to the medullary inhibitory center.

Conclusion is that the resultant of the opposing effects of the areas determines the nature of supraspinal control and the degree of the tone. This net-result of the opposing influences, through the EP tracts, reach the α- and γ-motor neurons, influence them and thus influence the basic stretch reflex. Pyramidal tract fibers also have influences.

Descending motor fibers, (belonging to both pyramidal tracts as well as to the extrapyramid tract) terminate on the anterior horn cells (either directly or via inter neurons) and influence the γ- or α-motor activity and thus modify the tone. These descending tracts, terminate ultimately on either the α-motor neurons of the γ-motor neurons. About 30% of the anterior horn cells belong to the γ-motor neuron system.

1. Brainstem influences

Descending signals from the cervical spinal cord, brainstem, and cortex can influence the rate of motor neuron firing and the recruitment of additional motor neurons to increase the speed and force of muscle contraction. The brainstem contains the neural circuitry for initiating locomotion and for controlling posture. The maintenance of posture requires coordinated activity of both axial and limb muscles in response to input from proprioceptors and spatial position sensors, such as the inner ear. Cerebral cortex input through the corticospinal system is necessary for the control of fine individual movements of the distal limbs and digits. Each higher level of the nervous system acts on lower levels to produce appropriate, more refined movements.

Extrapyramidal system (EPS)

A short synopsis of the extra pyramidal system (EPS) is given above. EPS consists of the (i) subcortical gray matter (ie, basal ganglia), (ii) the motor nuclei or reticular formation of the brain stem and (iii) the vestibular nucleus

plus the (iv) descending fibers which convey the impulses from these nuclei to the spinal cord. Their principal function in man is to influence and maintain (i) tone, (ii) posture and (iii) equilibrium.

From cerebral cortex (e.g. from area 6, the premotor area) some fibers descend to merge with the EPS nuclei lower down; they are called 'cortically originating extrapyramidal fibers' (COEPS).

Three brainstem nuclear groups give rise to descending motor tracts that influence motor neurons and their associated interneurons. These consist of the *red nucleus*, the *vestibular nuclear complex*, and the *reticular formation* (Fig. 6–5).

The principal tracts of the extrapyramidal system are :
1) reticulospinal tract;
2) vestibulospinal tract;
3) rubrospinal tract;
4) tectospinal tract.

The other major descending influence on the motor neurons is the corticospinal tract, the only volitional control pathway in the motor system. In most cases, the descending pathways act through synaptic connections on interneurons. The connection is less commonly made directly with motor neurons.

Reticulospinal tract

Arises from two sources, (a) the pontine nuclei, particularly those at caudal pons, and (b) medullary nuclei (from nucleus reticularis gigantocellularis;

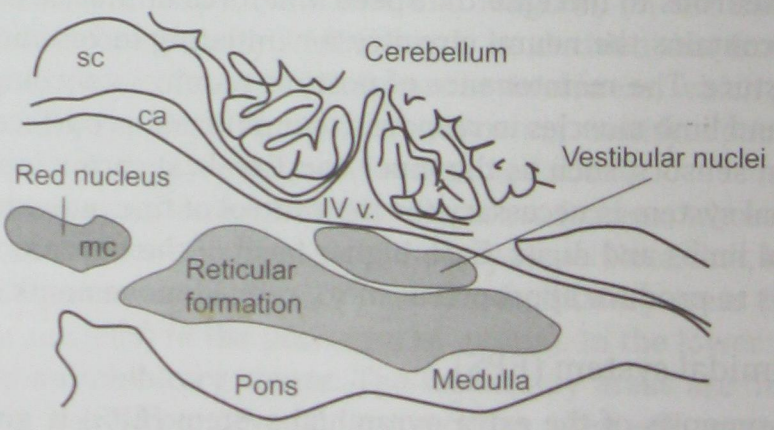


Fig. 6–5. Brainstem nuclei of the descending motor pathways.

Fig. 6–5). The tract remains uncrossed and terminate both on α as well as on γ-motor neurons of anterior horn. The tract contains both inhibitory and facilitatory fibers. Fibers from pons descend through anterior white column whereas those from medulla descend through lateral columns of the spinal cord. Facilitatory fibers facilitate while inhibitory fibers inhibit the tone. The fibers are called reticulospinal, because they arise from the reticular formation of brain stem terminate in the spinal cord. These axons have inhibitory influences on interneurons that modulate extensor motor neurons.

Vestibulospinal tract

The classical vestibulospinal tract arises from the lateral (Deiter's) nucleus of the vestibular nucleus, descends through the spinal cord, remaining uncrossed and terminates on the a motor neurons of the ventral horn. This tract, often called the Deiterospinal tract, is facilitatory, i.e., its discharge increases the muscle tone.

Another vestibulospinal tract, which is called the medial vestibulospinal tract, arising from the medial and inferior vestibular nuclei, runs through the medial longitudinal fasciculus (MLF). Some of these fibers terminate on the anterior horn cells.

Rubrospinal tract

The red nucleus of the mesencephalon receives major input from both the cerebellum and the cerebral cortical motor areas. Output via the rubrospinal tract is directed predominantly to contralateral spinal motor neurons that are involved with movements of the distal limbs.

Arising from the red nucleus (red nucleus is situated in the upper part of the mid brain), this tract decussates almost immediately (Forel's decussation) and eventually proceeds downwards through the lateral funiculus of the spinal cord, almost intermingling with the pyramidal tract, finally terminating on the anterior horn.

The red nucleus receives input from the cerebellum as well as from the motor cortex.

Rubrospinal action enhances the function of motor neurons innervating limb flexor muscles while inhibiting extensors. This tract may also influence γ-motor neuron function.

The vestibulospinal and reticulospinal tracts descend medially in the spinal cord and terminate in the ventromedial part of the intermediate zone, an area in the gray matter containing propriospinal interneurons (Fig. 6–6).

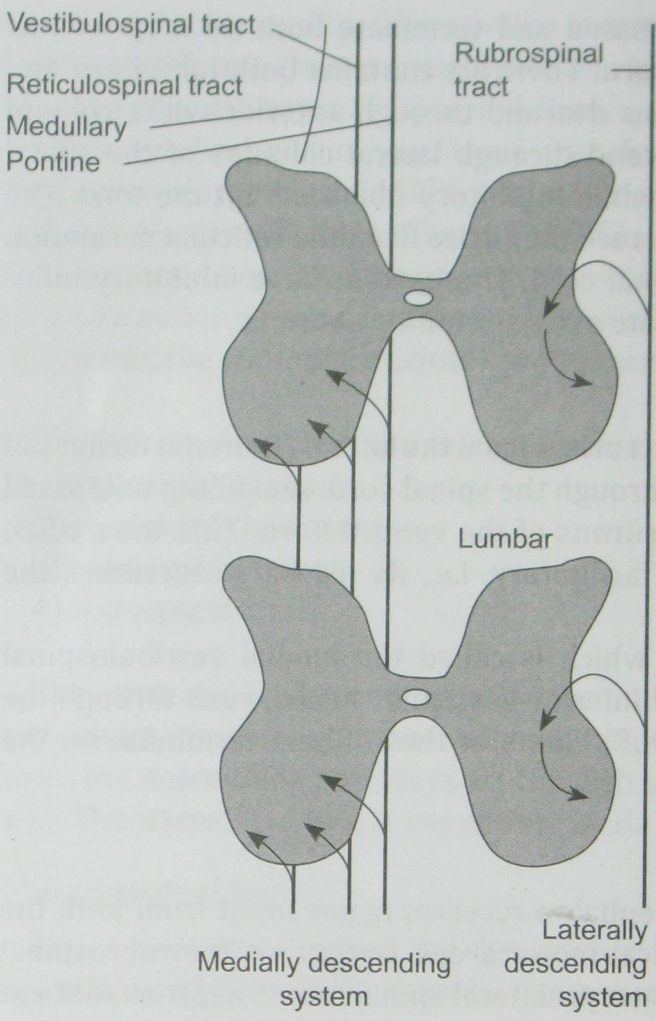


Fig. 6–6. Schematical draw of the spinal cord localization of reticulospinal, vestibulospinal and rubrospinal tracts second nuclei.

These tracts are the main CNS pathways for maintaining posture and head position during movement.

The rubrospinal tract descends laterally in the spinal cord and terminates mostly on interneurons in the lateral spinal intermediate zone, but it also has some monosynaptic connections directly on motor neurons to to muscles of the distal extremities. This tract supplements the medial descending pathways in postural control and the corticospinal tract for independent movements of the extremities.

Therefore, reticulospinal and vestibulospinal tracts are thought to be most important for the control of axial and proximal limb muscles, whereas the rubrospinal (and corticospinal) tracts are most important for the control of distal limb muscles, particularly the flexors.

2. Lesions of the descending motor tracts

If the *motor cortex or the internal capsule is damaged* (e.g., by bleeding or ischemia in the area supplied by the middle cerebral artery), impulse transmission in the immediately adjacent descending cortical tracts is interrupted (Fig. 6–7, 1). These make up the pyramidal tract and other connections of the motor cortex, such as those to the red nucleus and to the medullary reticular formation. The result is a reduced activity not only of the pyramidal tract but

also of the rubrospinal and medial reticulospinal tracts. The vestibulospinal and medial reticulospinal tracts are less affected, because they are under stronger noncortical influence, for example, from the cerebellum. An interruption of transmission in the area of the internal capsule thus ultimately results in an excessive activity of the extensors in the legs and the flexors in arms (Fig. 6–7, 2).

At first, however, *spinal shock* will set in due to cessation of supraspinal innervation of α-motoneurons (Fig. 6–7, 3a). The antigravity muscles are also affected, less so than the other muscles though, by the reduced supraspinal activation of the γ-motoneurons. In spinal shock the muscles are flaccid and no reflexes are elicited (areflexia).

However, partial "denervation" of the α- and γ-motoneurons as well as of interneurons leads to a gradual increase in sensitivity of these neurons. In addition, the endings of supraspinal neurons that are out of action are replaced by synapses with the spinal cord neurons (Fig. 6–7, 3b). As a consequence, the reflexes gradually gain a stronger influence on the activity of the α-motoneurons, and *hyperreflexia* occurs.

Another sequence is *spasticity*. After loss of function of the descending tracts, the activity of the α-motoneurons comes under the increasing influence of the muscle spindles and Golgi tendon organs (Fig. 6–7, 4).

Stretching the muscle spindles stimulates the α-motoneurons of the same muscle via a monosynaptic reflex; the increased influence of the muscle spindles results in massive contraction on stretching. Nevertheless, the response of the muscle spindles is mainly phasic, i.e., if they are stretched slowly or continuously their activity slowly decreases. As a result, the influence of the Golgi tendon organs becomes dominant: when the muscle is stretched they inhibit muscle contraction via an inhibiting interneuron. It is also under the influence of the Golgi tendon organs that on slow or continuous stretching the muscle will suddenly become flaccid after initial increase in tone (*clasp-knife effect*).

The predominance of the stretching muscles leads to extension of the big toe on stroking the sole of the foot (Fig. 6–7, 5), instead of its normal plantar flexion. This is called as *Babinski reflex*. It is taken as evidence for a lesion in the pyramidal tract. In fact the Babinski reflex is the result of a lesion of several descending cortical tracts, including the pyramidal tract. Isolated damage of the pyramidal tract (extremely rare) results in neither spasticity nor the Babinski reflex, but only minor disturbance of fine movement.

If the *red nucleus* has been destroyed (e.g., due to ischemia of the mid-brain or in Wilson's disease), coarse tremor will result. Neurons of the red nucleus are, among other functions, important for the dampening of oscillations that

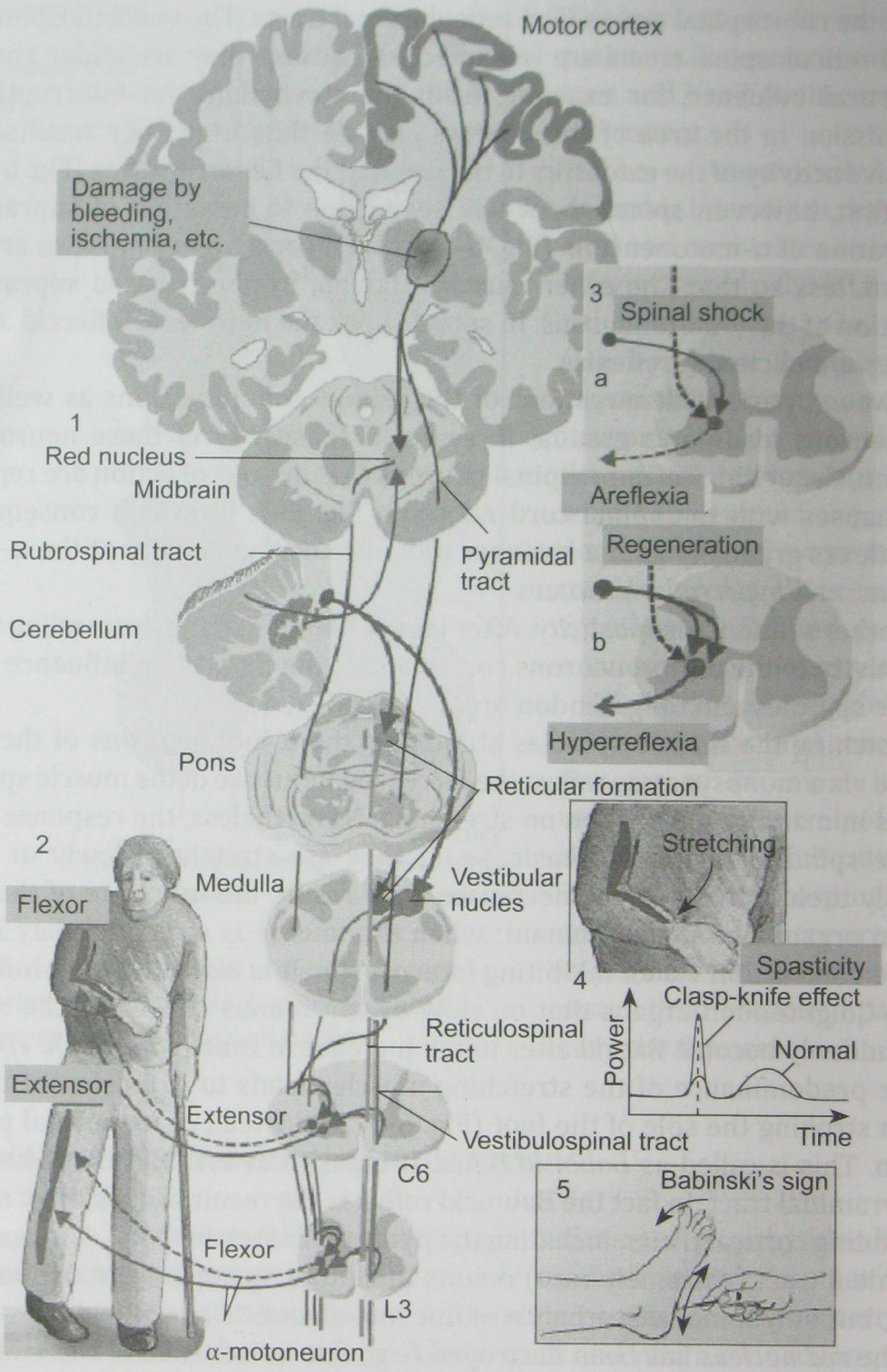


Fig. 6–7. Causes and sequences of the descending motor tracts lesions.

can occur as a result of a negative feedback in the control of γ-motoneurons activity.

In *lesions of the vestibular nucleus* abnormalities of balance with vertigo, nystagmus, and nausea predominate.

V. THE BASAL GANGLIA

The term basal ganglia (basal gangles) probably owes its origin to the fact that most of the structures in it originated, (in intrauterine life), from the basement membrane of the growing lateral ventricle.

The basal ganglia are a group of deep, interrelated subcortical nuclei that play an essential role in control of movement. They receive indirect input from the cerebellum and from all sensory ganglia systems, including vision, and the motor cortex. The structural components of the basal ganglia include the *caudate nucleus*, *putamen*, and the *globus pallidus* in the forebrain. The caudate and putamen are collectively referred to as the *neostriatum*, and the putamen and the globus pallidus form a wedge-shaped region called the *lentiform nucleus*.

Two other structures, the *subthalamic nucleus* of the diencephalon and the *substantia nigra* of the midbrain, are considered part of the basal ganglia.

The substantia nigra contains cells that use dopamine as a neurotransmitter and are rich in a black pigment called *melanin*. The high concentration of melanin gives the structure a black color, thus the name *substantia nigra*. The axons of the substantia nigra form the *nigrostriatal pathway*, which supplies dopamine to the striatum. The dopamine released from the substantia nigra regulates the overall excitability of the striatum and release of other neurotransmitters.

Their function is mainly to control movement in conjunction with the cerebellum, motor cortex, corticospinal tracts, and motor nuclei in the brainstem.

The basal ganglia have *input structures* that receive afferent information from outside structures, *internal circuits* that connect the various structures of the basal ganglia, and *output structures* that deliver information to other brain centers. The neostriatum represents the major input structure for the basal ganglia. Information coming from virtually all areas of the cortex and thalamus are projected to the neostriatum. The output areas of the basal ganglia have both ascending and descending components.

The major *ascending output* is transmitted to *thalamic nuclei*, which process all incoming sensory information that is transmitted to the cerebral cortex.

Descending output is directed to the *midbrain, brainstem* and *spinal cord.* The output functions of the basal ganglia are mainly inhibitory.

Although the circuitry of the basal ganglia appears complex at first glance, it can be simplified into input, output, and internal pathways (Fig. 6–8). Input is derived from all of the cerebral cortex except for the primary visual and auditory areas and is directed to the striatum and the subthalamic nucleus. The input is roughly somatotopic and is via neurons that use glutamate as the neurotransmitter.

Basal ganglia output is from the internal segment of the globus pallidus (GPi) and one segment of the substantia nigra. The GPi output is directed to ventrolateral and ventral anterior nuclei of the thalamus, which feed back to the cortical motor areas. The output of the GPi is also directed to a region in the upper brainstem termed the *midbrain extrapyramidal area.* This latter area then projects to the neurons of the reticulospinal tract.

The substantia nigra output arises from the pars reticulata. The output is directed to the superior colliculus of the mesencephalon, which is involved in eye movement control. The GPi and SNr output is inhibitory via neurons that use GABA as the neurotransmitter.

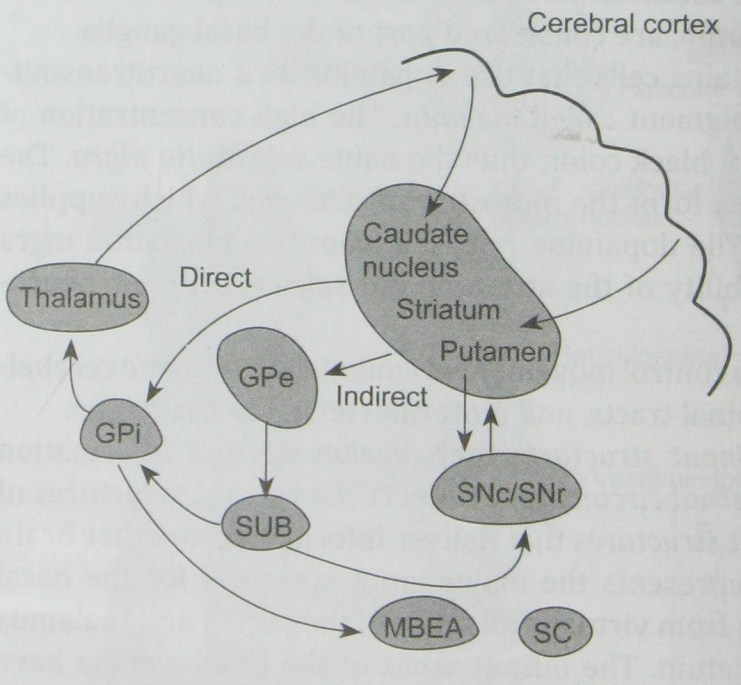


Fig. 6–8. Basal ganglia nuclei and circuitry (explanation see in the text.)

The internal pathway circuits link the various nuclei of the basal ganglia. The globus pallidus externa (GPe), the subthalamic nucleus, and the pars compacta region of the substantia nigra (SNc) are the nuclei in these pathways. The GPe receives inhibitory input from the striatum via GABA-releasing neurons. The output of the GPe is also inhibitory via GABA release and is directed to the GPi and the subthalamic

nucleus. The subthalamic nucleus output is excitatory and is directed to the GPi and the SNr.

This striatum- GPe-subthalamic nucleus-GPi circuit has been termed the *indirect pathway* in contrast to the *direct pathway* of striatum to Gpi (Fig. 6–8). The SNc receives inhibitory input from the striatum and produces output back to the striatum via dopamine-releasing neurons. The output can be either excitatory or inhibitory depending on the receptor type of the target neurons in the striatum. The action of the SNc may modulate cortical input to the striatum.

Therefore, summarizing, it should be stressed that striatal neurons are activated, via *glutamate,* by neurons of the cortex. The *internal interconnections* of the basal ganglia (Fig. 6–9) are mainly provided by the inhibitory transmitter GABA. Ultimately, the basal ganglia have an inhibitory effect on the thalamus via GABA-ergic neurons in the inner pallidum and the substantia nigra (pars reticulata). These neurons are activated via glutamate from the neurons of the subthalamic nucleus. Finally, the striatal neurons are partly activated and partly inhibited by dopamine from the substantia nigra (pars compacta), and also activated via cholinergic neurons. An imbalance between inhibitory and activating influences has a harmful effect on motor functions: too strong an inhibition of the thalamic nuclei has a hypokinetic, too little has a hyperkinetic effect.

The most is known about the inhibitory basal ganglia loop involved in modulating cortical motor control. This loop regulates release of stereotyped

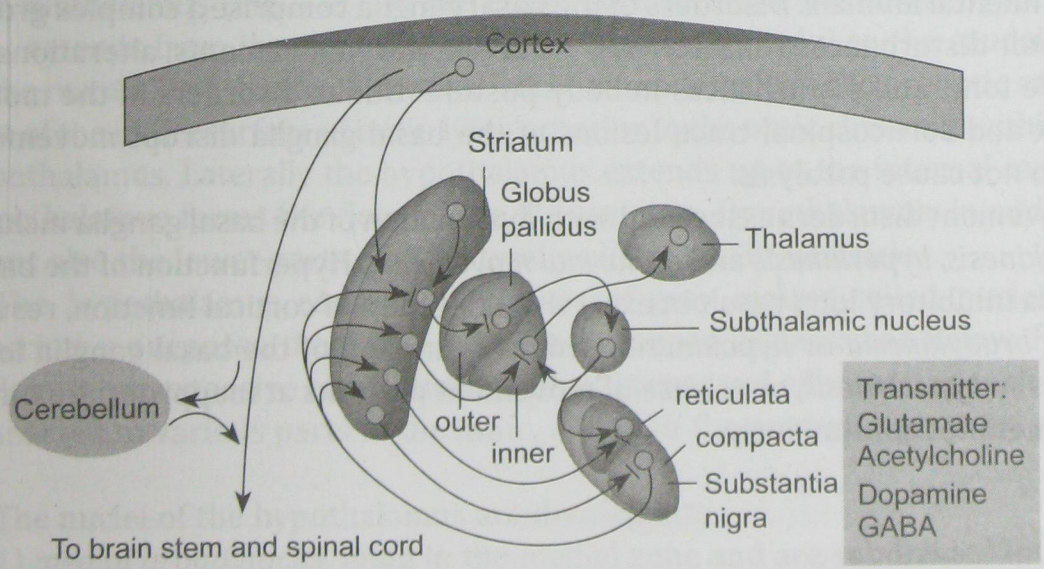


Fig. 6–9. Basal ganglia input, outer and internal neurotransmitters.

movement patterns that add efficiency and gracefulness to precise and delicate cortically controlled movements. These movements include inherited patterns that add efficiency, balance, and gracefulness to motion, such as the swinging of the arms during walking and running, and the highly learned automatic postural and follow through movements of throwing a ball or swinging a bat.

An additional modulating circuit involves a neostriatal inhibitory projection on the substantia nigra. The substantia nigra projects dopaminergic axons back on the neostriatum. A deficiency in the dopaminergic projection of this modulating circuit is implicated in Parkinson's syndrome. The function of the neostriatum also involves local cholinergic interneurons, and their destruction is thought to be related to the choreiform movements of Huntington's chorea, another basal ganglia-related syndrome.

Functions of basal ganglia are not clearly investigated still. The standard techniques of finding out the functions in a brain structure consists of (i) ablation of removal of the structure, (ii) stimulation of it, (iii) studying the electrical changes, and (iv) studying the human beings who are known to have disease of the structure concerned (clinical studies). The first three of the four methods can be performed in animals. But owing to the fact that the basal ganglia are very deep seated, such animal studies invariably result in injuries of other neighbouring structures leading to artifact formation in the results. The clinical studies therefore are very important.

Basal ganglia diseases produce profound motor dysfunction in humans and experimental animals. Disorders of the basal ganglia comprise a complex group of motor disturbances characterized by involuntary movements, alterations in muscle tone, and disturbances in body posture. Unlike disorders of the motor cortex and corticospinal tract, lesions of the basal ganglia disrupt movement but do not cause paralysis.

Movement disorders associated with dysfunction of the basal ganglia include *bradykinesis, hyperkinesis* and *abnormal movements*. Hyperfunction of the basal ganglia inhibitory loop result in excessive inhibition of cortical function, resulting in *bradykinesia* or *hypokinesis*. Reduced function of the basal ganglia loop results in *hyperkinesis*, or release of movement patterns at inappropriate times or sometimes continuously.

Chapter 7. Forebrain and cerebellar regulation of motor function and organism systemic activity. Physiology of thalamus and brain cortex. Electric activity of CNS and its clinical importance

I. INTRODUCTION. HYPOTHALAMUS

The hypothalamus integrates the body's autonomic, endocrine, and somatomotor functions. Neurons in the hypothalamus are responsible for regulating various homeostatic functions such as food intake, electrolyte and water metabolism, temperature regulation, and circadian rhythm. In addition, the functions are adapted in the hypothalamus to the required behavioral patterns, such as the fight and flight reaction, nutritive or sexual behavior. The programs required for the particular behavioral patterns are stored in the hypothalamus and are called up as needed, in particular by the neurons of the limbic system.

Hypothalamus (which weighs about only 4 g) together with thalamus, forms the *'diencephalon'*. The *'diencephalon'* includes the hypothalamus, thalamus, and subthalamus (Fig. 7–1). In the under surface of the brain, the hypothalamus extends from the region of optic chiasma (anteriorly) to the caudal part of the mammillary bodies. However, some authors include the preoptic area (a part of the undersurface of brain lying anterior to the optic chiasma) within the hypothalamus. Laterally the hypothalamus extends up to the internal capsule. Hypothalamus forms the floor of the 3rd ventricle. Hypothalamus, in addition, forms also the lower parts of the lateral walls of the 3rd ventricle.

The hypothalamus consists of masses of gray matter called nuclei. For descriptive purposes, the nuclei have been grouped as described below. Hypothalamus has extensive connections, both afferents and efferents, by which it is connected to various parts of the brain, and with the posterior pituitary gland.

The nuclei of the hypothalamus are divided into

1) *medial group nuclei*, lying in the medial zone and are subdivided into,

- preoptic
- paraventricular,

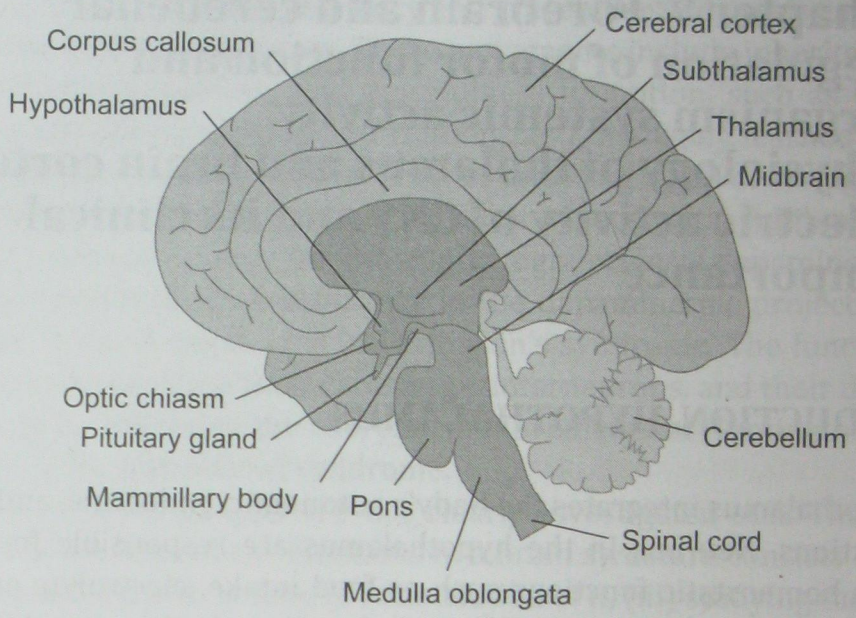


Fig. 7–1. A midsagittal section of the human brain showing the most prominent structures of the brainstem, diencephalon and forebrain.

- dorsomedial
- ventromedial
- infundibular and
- posterior nuclei (from rostral to caudal side), and

2) *lateral goup nuclei*, lying left or to the right side of the medial group nuclei and include,

- supraoptic
- large lateral,
- tuber mammillary and
- lateral tuberal nuclei (from rostral to caudal).

The standard landmarks on the undersurface of the hypothalamus are:

- optic chiasma, in the anterior
- tuber cinerium (tuber = swelling; cinerium = gray) and the infundibulum (where the stalk of the pituitary is attached), in the middle and
- mammillary bodies (a few swellings situated in the caudal part of the undersurface of the hypothalamus), in the posterior zone.

II. CONNECTIONS

A. Afferent

1. From limbic system:
- from hippocampus, and
- amygdala. Lymbic system is concerned with various aspects of emotion and behavior. Thus the hypothalamus is related to the physiology of emotion and behavior. Hippocampus and amygdala are also important sites for memory storage and memory consolidation.

2. From cortex, particularly from the frontal lobe (area 6 and prefrontal area). These fibers enter the hypothalamus directly. Many of other fibers enter the hypothalamus (from cortex) via the thalamus.
3. From olfactory lobe and paraolfactory area. The sense of olfaction, emotion and behavior and the hypothalamus are thus tied up.
4. Collaterals from the lemniscal system and fibers from the reticular activating system (RAS). The fibers which come from the RAS are probably concerned with maintenance of alertness. Collaterals from the lemniscal system keep the hypothalamus informed about the somatic sensation.
5. From the corpus striatum. (Recall, corpus striatum has strong control on the tone and some involuntary movements). Hypertonia and even shivering may occur in cold environment and for them to occur an intact hypothalamus is necessary. It is possible that hypertonia and shivering produced by hypothalamus are controlled by feed backs from basal ganglia.
6. From thalamus.
7. From brain stem: from the
- raphe nucleus (serotonergic and noradrenergic neurons);
- locus ceruleus (noradrenergic neurons);
- medulla (for the anatomical sites of raphe nucleus and locus ceruleus).

B. Efferent

1. To brain stem and the spinal cord. Probably via the pontine nuclei then reticulospinal fibers, descending impulses from hypothalamus terminate on the lateral horn cells, controlling the spinal sympathetic centers. Fibers from the hypothalamus also control the brain stem autonomic centers, particularly, the vasomotor center (VMC). Hypothalamus thus controls the blood pressure, heart rate, diameter of the blood vessels etc.

2. Fibers from supraoptic and paraventricular nuclei form the 'hypothalamo-hypophyseal path' to end in posterior pituitary. Via these fibers the hypothalamus sends the posterior pituitary hormones.
3. To thalamus.
4. To midbrain.

III. HYPOTHALAMIC NEUROTRANSMITTERS

Hypothalamic tracts, that is, groups of nerve fibers which end at the hypothalamus or emerge from it, secrete various monoamines as neurotransmitters. Examples of such monoamines are not adrenalin, dopamine, serotonin, adrenalin and histamine. This makes possible the control of hypothalamic functions by pharmacological agents.

IV. FUNCTIONS OF THE HYPOTHALAMUS

The nuclei of the hypothalamus contain groups of neurons that regulate several important physiological functions:

1. *Water and electrolyte balance* in magnocellular cells of the supraoptic and paraventricular nuclei.
2. *Secretion of hypothalamic releasing factors* in the arcuate and periventricular nuclei and in parvocellular cells of the paraventricular nucleus
3. *Temperature regulation* in the anterior and posterior hypothalamic nuclei
4. *Activation of the sympathetic nervous system and adrenal medullary hormone secretion* in the dorsal and posterior hypothalamus
5. *Thirst and drinking regulation* in the lateral hypothalamus
6. *Hunger, satiety and the regulation of eating behavior* in the arcuate nucleus, ventromedial nucleus, and lateral hypothalamic area
7. *Regulation of sexual behavior* in the anterior and preoptic areas
8. *Regulation of circadian rhythms* in the suprachiasmatic nucleus

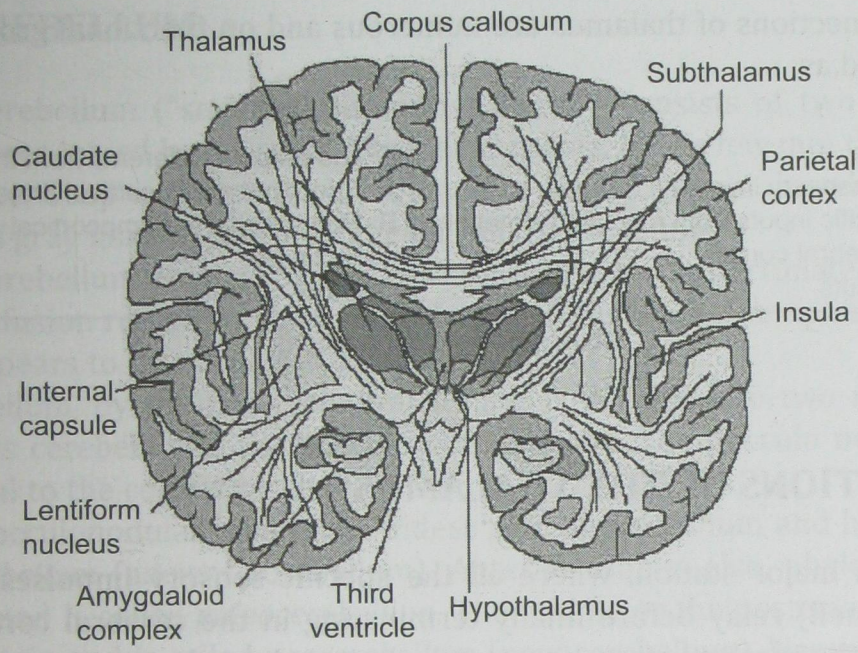


Fig. 7–3. Frontal section of the human brain showing the third ventricle, thalamus and the neighboring nuclei and structures.

V. THALAMUS

There are two thalami and in between the two thalami, the 3rd ventricle lies. Together with the hypothalamus, the thalami constitute the bulk of the diencephalon (Fig. 7–3).

The anterior end of each thalamus is narrow while the posterior end is broad and called the 'pulvinar'. Each thalamus rests on the superior surface of the midbrain. The medial surface of the thalamus forms the lateral wall of the 3rd ventricle. The internal capsule stands in between the lateral surface of the thalamus and the lentiform nucleus. A mass of gray matter, called, the 'massa intermedia' connects the two thalami at their anterior parts.

A sheet of white matter, called the internal medullary lamina resembling the English alphabet Y, divides each thalamus into three major parts, via, (i) anterior, (ii) medial, and the (iii) lateral parts, and each part contains a number of thalamic nuclei.

However description of these nuclei is unnecessary in a moderate sized text book of physiology lectures like this. Collectively, nuclei are called, anterior group, medial group and lateral group of nuclei, respectively.

The connections of thalamus are numerous and on functional ground, may be classified as:

I. The inputs	II. The outputs (efferents from thalamus)
A. Specific sensory inputs.	A. Specific thalamocortical projections.
B. Non specific inputs from reticular formation.	B. Non specific thalamocortical projections.
C. From cerebral cortex.	C. Other outputs.
D. Other inputs.	

III. Internuclear connections

VI. FUNCTIONS OF THE THALAMUS

1. It is a major station, where all the specific sensory impulses (excepting smell) relay before finally terminating in the cerebral cortex. Even when the cerebral cortex is severely damaged, crude sensory perceptions remain present, because the thalamus acts as a crude center for sense perception. Two neuronal circuits are significant in this regard. One is the pathway from the cerebral cortex to the pons and cerebellum and then, by way of the thalamus, back to the motor cortex. The second is the feedback circuit that travels from the cortex to the basal ganglia, then to the thalamus, and from the thalamus back to the cortex. The subthalamus also contains movement control systems related to the basal ganglia.
2. Majority of the nonspecific ascending impulses from the RAS relay in the thalamic nuclei, before proceeding further via non specific thalamocortical projections to the cerebral cortex. These nonspecific fibers, when active, cause wakefulness, alertness and consciousness. Through its connections with the ascending reticular activating system, the thalamus processes neural influences that are basic to cortical excitatory rhythms (i.e., those recorded on the electroencephalogram), to essential sleep-wakefulness cycles, and to the process of attending to stimuli.
3. Because of the intimate connections between the thalamus and the frontal cortex and the hypothalamus, the thalamus is involved in the subjective feeling of personality and various emotions. Thalamus thus may be regarded as a part of the limbic system.
4. Some information going to the cortex necessary for smooth operations of skeletal motor activities pass through the thalamus, relaying here. Examples are dentato-thalamicocortical path and fibers going to the corpus striatum.

VII. CEREBELLUM

The cerebellum ("small or diminutive brain") consists of two cerebellar hemispheres joined by a vermis. Cerebellar cortex, only a few mm thick, lies in the surface. Deep to the cortex lies the white matter. Within the white matter, masses of gray matter, called cerebellar nuclei, are found.

The cerebellum is divided into several lobes, but unfortunately, there is some confusion regarding the nomenclature of the lobes. A simplified version, which appears to be popular is given below:

Cerebellum, by the posterolateral fissure is divided into two main lobes, (i) corpus cerebelli and (ii) flocculo nodular lobe. The flocculo nodular lobe lies caudal to the corpus cerebelli.

The flocculonodular lobe is the oldest part of cerebellum and hence called *archicerebellum* (ancient cerebellum). Anterior lobe is also, phylogenetically very old and is called *paleocerebellum*. The middle or the posterior lobe is on recent origin and is called *neocerebellum* (new cerebellum). Neocerebellum is best developed in man.

Three pairs of peduncles, called superior, middle and inferior cerebellar peduncles connect the cerebellum with the brain system. The superior one (brachium conjuctivum) connects the cerebellum with the midbrain; the middle (brachius pontis) with the pons and the inferior (restiform body) with the medulla. These peduncles are made up of only nerve fibers, from the cerebellum to the brain stem (efferents) or to the cerebellum (afferents) from other parts of CNS.

The cerebellar surface is arranged in multiple, parallel, longitudinal folds termed *folia*. Several deep fissures divide the cerebellum into three main morphological components — the anterior, posterior, and flocculonodular lobes, which also correspond with the functional subdivisions of the cerebellum (Fig. 7–4).

The functional divisions are the *vestibulocerebellum*, the *spinocerebellum* and the *cerebrocerebellum*. These divisions appear in sequence during evolution. They have similar intrinsic circuitry; thus, the function of each depends on the nature of the output nucleus to which it projects.

The *vestibulocerebellum* is composed of the flocculonodular lobe. It receives input from the vestibular system and visual areas. Output goes to the vestibular nuclei, which can, in a sense, be considered as an additional pair of intrinsic cerebellar nuclei. The vestibulocerebellum functions to control equilibrium and eye movements.

The medially placed *spinocerebellum* consists of the midline vermis plus the medial portion of the lateral hemispheres, called the intermediate zones. Spino-

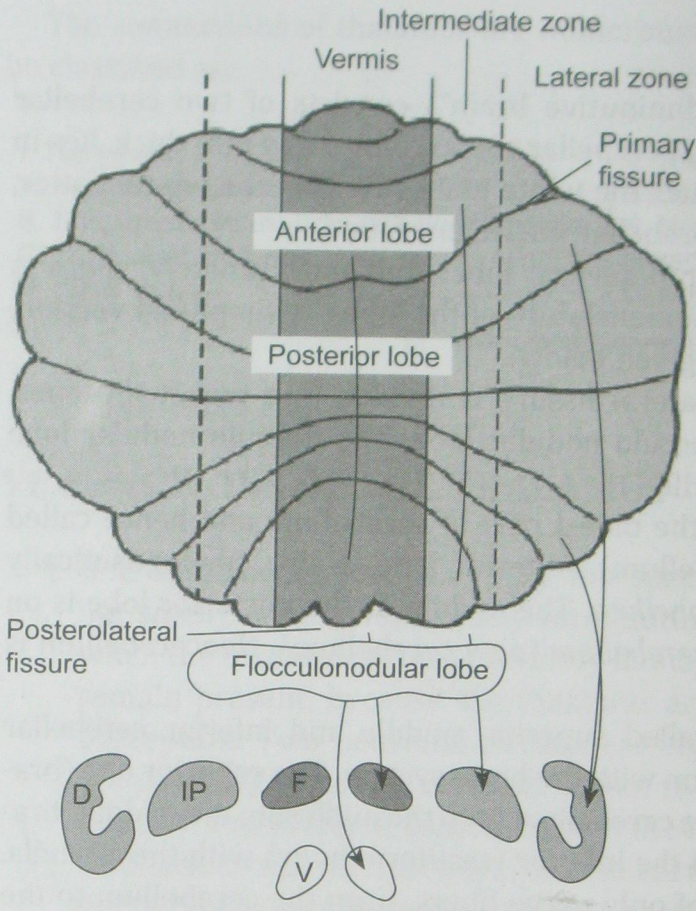


Fig. 7–4. The structure of cerebellum. Explanations see in the text.

cerebellar pathways carrying somatosensory information terminate in the vermis and intermediate zones in somatotopic arrangements. The auditory, visual, and vestibular systems and sensorimotor cortex also project to this portion of the cerebellum. Output from the vermis is directed to the fastigial nuclei, which project through the inferior cerebellar peduncle to the vestibular nuclei and reticular formation of the pons and medulla. Output from the intermediate zones goes to the interposed nuclei and from there to the red nucleus and, ultimately, to the motor cortex via the ventrolateral nucleus of the thalamus. It is believed that both the fastigial and interposed nuclei contain a complete representation of the muscles of the body. The fastigial output system controls antigravity muscles in posture and locomotion, while the interposed nuclei, perhaps, act on stretch reflexes and other somatosensory reflexes.

The *cerebrocerebellum* occupies the lateral aspects of the cerebellar hemispheres. Input comes exclusively from the cerebral cortex, relayed through the middle cerebellar peduncles of the pons. The cortical areas that are prominent in motor control are the sources for most of this input. Output is directed to the dentate nuclei and from there via the ventrolateral thalamus back to the motor and premotor cortices.

The cerebellar cortex is composed of five types of neurons arranged into three layers (Fig. 7–5). The *molecular layer* is the outermost and consists mostly of axons and dendrites plus two types of interneurons, *stellate cells*

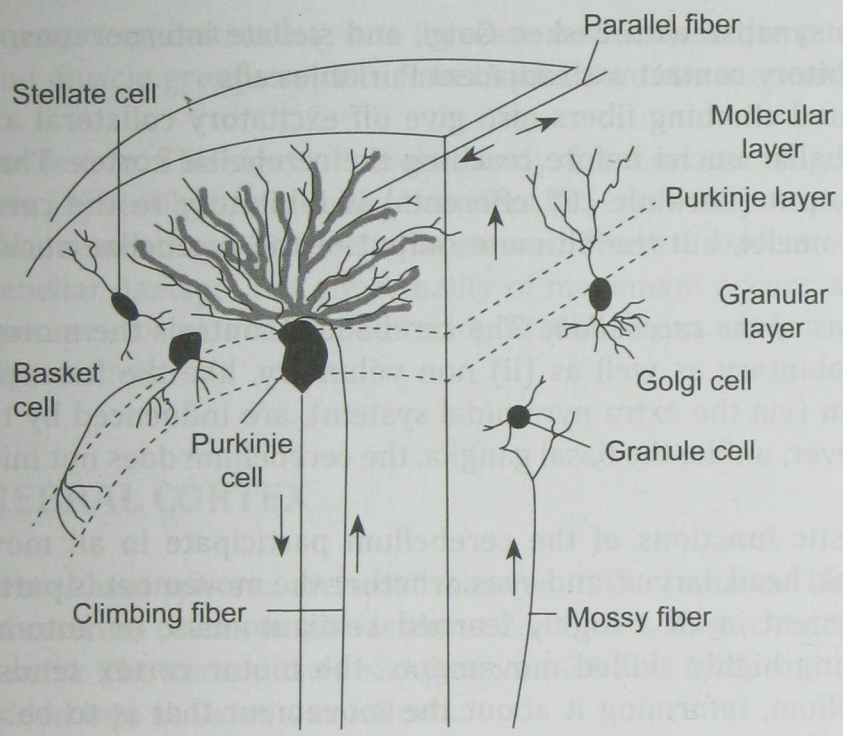


Fig. 7–5. The structure of the cerebellar cortex.

and *basket cells.* The next, *ganglionic layer,* contains the dramatic *Purkinje cells,* whose dendrites reach upward into the molecular layer in a fan-like array. The Purkinje cells are the only efferent neurons of the cere-bellar cortex. Their action is inhibitory via GABA as the neurotransmitter.

Deep to the Purkinje cells is the *granular layer,* containing *Golgi cells* and small local circuit neurons, the *granule cells.*

Afferent axons to the cerebellar cortex are of two types: *mossy fibers* and *climbing fibers. Mossy fibers* arise from the spinal cord and brainstem neurons, including those of the pons that receive input from the cerebral cortex.

Mossy fibers make complex multicontact synapses on granule cells. The granule cell axons then ascend to the molecular layer and bifurcate, forming the *parallel fibers.* These travel perpendicular to and synapse with the dendrites of Purkinje cells, providing excitatory input via glutamate..

Climbing fibers arise from the inferior olive, a nucleus in the medulla. Each climbing fiber synapses directly on the dendrites of a Purkinje cell and exerts a strong excitatory influence. One action potential in a climbing fiber produces a burst of action potentials in the Purkinje cell called a complex spike. Climbing

fibers also synapse with basket, Golgi, and stellate interneruons, which then make inhibitory contact with adjacent Purkinje cells.

Mossy and climbing fibers also give off excitatory collateral axons to the deep cerebellar nuclei before reaching the cerebellar cortex. The cerebellar cortical output (Purkinje cell efferents) is inhibitory to the cerebellar and vestibular nuclei, but the ultimate output of the cerebellar nuclei is mostly excitatory.

Functions of the cerebellum. The cerebellum controls the motor activities; both (i) voluntary as well as (ii) non voluntary, like the tone, posture and equilibrium (via the extra pyramidal system), are influenced by the cerebellum. However, unlike the basal ganglia, the cerebellum does not initiate movements.

Synergistic functions of the cerebellum participate in all movements of limbs, trunk, head, larynx, and eyes, whether the movement is part of a voluntary movement or of a highly learned semiautomatic or automatic movement. During highly skilled movements, the motor cortex sends signals to the cerebellum, informing it about the movement that is to be performed. The cerebellum makes continuous adjustments, resulting in smoothness of movement, particularly during delicate maneuvers. The readjustments are automatic, highly precise and smooth in operation. Highly skillful movement requires extensive motor training, and considerable evidence suggests many of these learned movement patterns involve cerebellar circuits. This means, if the cerebellum is damaged, there will be faults in reaching the goal of reaching in smoothly.

The cerebellum receives proprioceptor input from the vestibular system; feedback from the muscles, tendons, and joints; and indirect signals from the somesthetic, visual, and auditory systems that provide background information for ongoing movement. Sensory and motor information from a given area of the body is sent to the same area in the cerebellum. In this way, the cerebellum can assess continuously the status of each body part — position, rate of movement, and forces such as gravity that are opposing movement.

The cerebellum also gets a feed back from the which the extrapyramidal tracts arise and thus cerebellum remains aware of the extrapyramidal tract activities. The cerebellum, via its efferents, also control the above mentioned nuclei of the extra pyramidal tracts.

The cerebellum compares what is actually happening with what is intended to happen. It then transmits the appropriate corrective signals back

to the motor system, instructing it to increase or decrease the activity of the participating muscle groups so that smooth and accurate movements can be performed.

Cerebellar lesions. The signs of cerebellar dysfunction can be grouped into three classes: (1) vestibulocerebellar disorders, (2) cerebellar ataxia or decomposition of movement, and (3) cerebellar tremor. These disorders occur on the side of cerebellar damage. The abnormality of movement occurs whether the eyes are open or closed. Visual monitoring of movement cannot compensate for cerebellar defects.

VIII. CEREBRAL CORTEX

The cerebral hemispheres are most massive in the humans. The total surface area of cerebral cortex in man is about 2,200 cm^2 of which major parts lie within the sulci. The cortex itself is only about 24 mm thick and microscopically, is divided into six layers, as follows:

1. 'Molecular' (also called 'plexiform') layers.
2. External granular layer.
3. External pyramidal layer.
4. Internal granular layer.
5. Internal pyramidal (also called 'ganglionic') layer.
6. Multiform ('polymorphic cell') layer.

The cortex, a few-millimeters-thick outer shell of the cerebrum, has a rich, multilayered array of neurons and their processes forming columns perpendicular to the surface. The axons of cortical neurons give rise to descending fiber tracts and intrahemispheric and interheispheric fiber tracts, which, together with ascending axons coursing toward the cortex, make up the prominent white matter underlying the outer cortical gray matter.

A deep sagittal fissure divides the cortex into a right and left hemisphere, each of which receives sensory input from and sends its motor output to the opposite side of the body. A set of *commissures* containing axonal fibers interconnects the two hemispheres, so that processed neural information from one side of the forebrain is transmitted to the opposite hemisphere. The largest of these commisures is the *corpus callosum*, which interconnects the major portion of the hemispheric regions (Fig. 7–6).

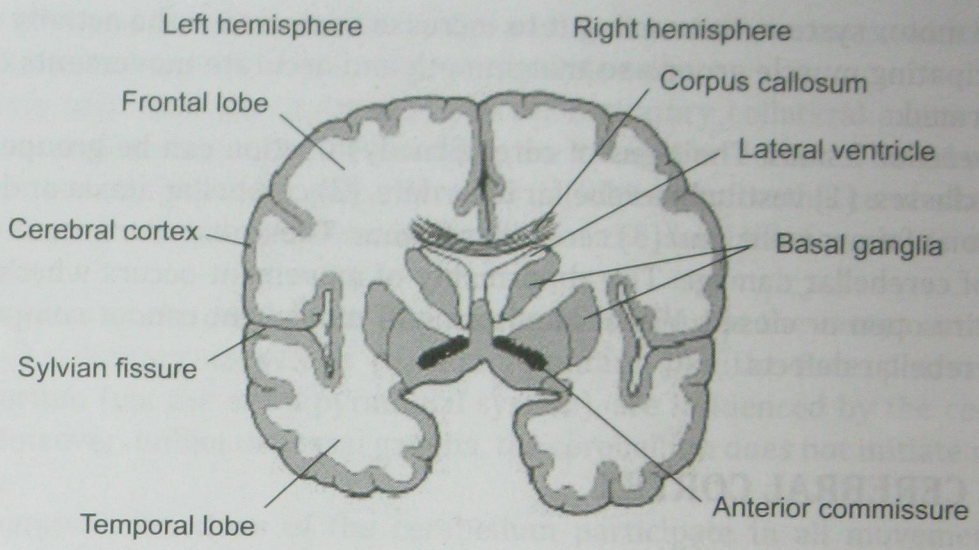


Fig. 7–6. The cerebral hemispheres and some deep structures in a coronal section through the rostral forebrain.

IX. THE MAJOR LOBES

The cerebral cortex has several areas and the areas can be classified in several ways:

1) functionally, e.g., (a) sensory cortex
2) anatomically e.g., (a) frontal, (b) parietal, (c) temporal, (d) occipital lobes.
3) cytoarchitonic classification of cortical areas is also known.

X. FUNCTIONS OF INDIVIDUAL AREAS

In the human brain, the surface of the cerebral cortex is highly convoluted, with **gyri** and **sulci** which are akin to hills and valleys, respectively. Deep sulci are also called **fissures**. Two deep fissures form prominent landmarks on the surface of the cortex; the **central sulcus** divides the **frontal lobe** from the **parietal lobe**, and the **sylvian fissure** divides the parietal lobe from the **temporal lobe** (Fig. 7–7).

The **occipital lobe** has less prominent sulci separating it from the parietal and temporal lobes. Topographically, the cerebral cortex is divided into areas of

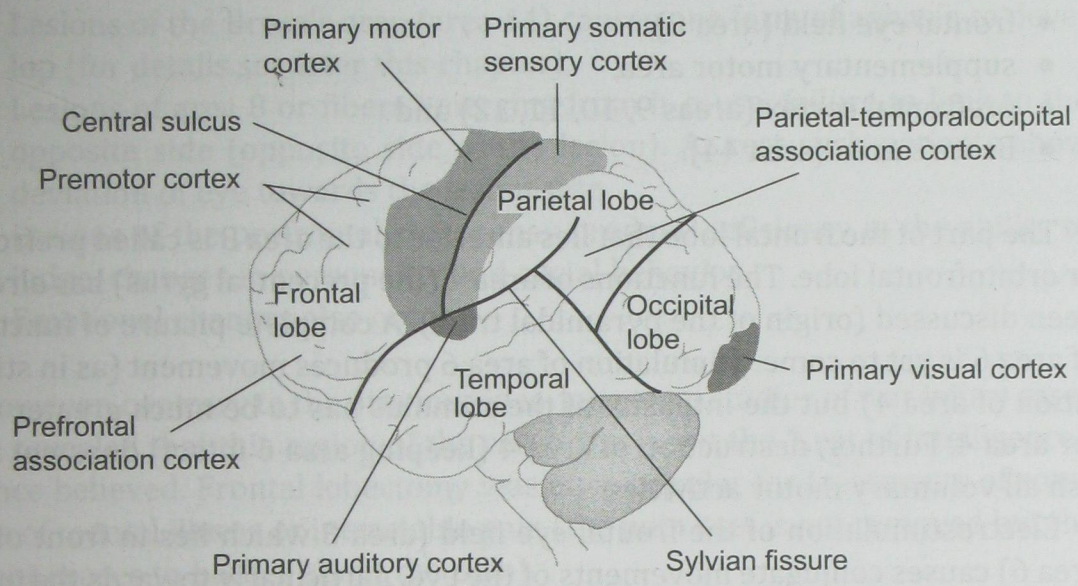


Fig. 7–7. The four cerebral cortex lobes containing primary sensory and motor areas and major association areas.

specialized functions, including the primary sensory areas for vision (occipital cortex), hearing (temporal cortex), somatic sensation (postcentral gyrus), and primary motor area (precentral gyrus).

As shown in Fig. 7–7, these well-defined areas comprise only a small fraction of the surface of the cerebral cortex. The majority of the remaining cortical area is known as *association cortex*, where the processing of neural information is performed at the highest levels of which the organism is capable; among verteb-rates, the human cortex contains the most exten-sive association areas.

The association areas are also sites of long-term memory, and they control such human functions as language acquisition, speech, musical ability, mathematical ability, complex motor skills, abstract thought, symbolic thought, and other cognitive functions.

1. Frontal lobe

The major areas in this lobe are:

- area 4 (motor area),
- area 6 (premotor area),

- frontal eye field (area 8),
- supplementary motor area,
- prefrontal cortex (areas 9, 10, 11, 12) and
- Broca's area (area 44).

The part of the frontal lobe that lies anterior to the area 8 is called prefrontal or orbitofrontal lobe. The functions of area 4 (the precentral gyrus) has already been discussed (origin of the pyramidal tract). A complete picture of functions of area 6 is yet to come. Stimulation of area 6 produces movement (as in stimulation of area 4) but the intensity of the stimulus has to be much greater than for area 4. Further, destruction of area 4 (keeping area 6 intact) does not abolish all voluntary motor activities.

Electrostimulation of the frontal eye field (area 8 which lies in front of the area 6) causes conjugate movements of the eye, particularly towards the opposite side. Therefore area 8 appears to be mainly responsible for the voluntary (but not the reflex) eye movement. It may also be responsible for the saccadic movements of the eye ball.

Area 44 in the dominant hemisphere is the motor area for speech (also called Broca's area). The dominant area for right handed persons is left hemisphere, whereas for left handed person, it is the right hemisphere.

On suitable electrostimulation of the prefrontal lobe **autonomic changes** (e.g. blood pressure, heart rate, gastrointestinal secretion and motility) can be elicited. The area is somehow related to **pain perception**. It is connected with limbic system and thus associated with **emotions**. The ability to **judge** with a situation and make small plans are also associated with this area.

Thus, this region influences autonomic fibers, emotions, pain perception and higher faculties.

Lesions in this lobe (due to vascular accidents/tumor/injury) cause signs and symptoms depending upon the exact site. Examples:

1. In lesions of area 4, signs and symptoms of pyramidal tract lesion will develop, although frank hypertonia of the muscles may not develop. When, however, there is concomitant injury of area 6, the muscles are frankly spastic (see also 'upper motor neuron lesion').
2. When there is an irritating focus in the area 4, a condition called 'Jaksonian epilepsy' develops in the area of the body which corresponds to the irritative site in the brain. The convulsion may remain local or it may spread.

3. Lesions of the Broca's area (area 44) causes one form of aphasia to develop (for details see later this chapter).
4. Lesions of area 8 or fibers emerging from it cause, failure to look to the opposite side (opposite side of the lesion). At rest such patients show deviation of eye towards the lesion side.
5. Lesions of the prefrontal cortex may produce deficiency in the ability to judge, changes in personality and social behavior.
6. Emotional changes also occur.

However, contrary to the previous idea, long term studies of war injury cases have revealed that this region of the cerebrum is not the "seat of intelligence", as once believed. Frontal lobectomy was once popular for treatments of some types of mental illness or intractable pain (the pain itsef is not removed but the patient ceases to bother about the pain).

2. Parietal lobe

Major areas in this lobe are:

- primary sensory area (also called somatic area I) or Brodmann's area 3, 1, 2, or the post central gyrus,
- secondary somatic area or somatic area II and
- somesthetic association area.

Primary somesthetic area (Brodmann's 3, 1, 2) is the highest area of general, conscious, sense perception. Fibers conveying impulses by spino-thalamic tract, lemniscal system, etc. (touch, pain, temperature senses etc), after relaying in the thalamus, end here, and thus we become aware of the specific senses. There is a somatotropic mapping, called sensory homunculus which describes pictorally, which area of somesthetic area I is responsible for sense perception from which area of the body.

However, numerous interconnections exist between area 4 and area 3, 2, 1 so that stimulation of primary sensory area can occasionally cause motor response; or stimulation of motor area (area 4) can elicit sense perception. Therefore, area 4 should be considered as an unit tied up with primary sensory area and these two areas, taken together, constitute the sensory motor cortex.

Somesthetic area II (SII) lies below the somesthetic area I, so that it is buried in the lateral (Sylvian) sulcus. Therefore, there are two (dual) representation

for each sensory perception from each field of the body although it appears somesthetic area I alone would have been sufficient.

Exact function of SII is not clear. It appears that for our ability to learn via various somesthetic senses (e.g., touch) , the presence of an intact SII is necessary. SII receives input from 2 sources, via, (i) from SI (post central gyrus) , and (ii) directly from thalamus.

Somatosensory association area (area 5 and 7 of Brodmann) lies posterior to the somesthetic area I, in the parietal lobe. Funnily, it does not receive large number of afferents nor does it give large number of efferent fibers.

It's major functions appears to be as follows:

1. *Various sensory input* (from SI, olfactory, auditory, visual areas of the brain etc.) reach the posterior part of parietal lobe, also called posterior parietal lobe (areas 5 & 7 of Brodmann = somatosensory association area) and are integrated there. This integration produces a motive and the motive in turn can lead to an action.

Example. A man touches an item of food. At the same time he is seeing and smelling it. Synthesis occurs in posterior parietal lobe (automatic synthesis, without effort, i.e., reflexy) and perception occurs that the item is food and the food is very expensive and very delicious. Motive to eat, develops. Action, i.e., eating now follows. In damages of posterior parietal lobe, tactile agnosia (inability to recognize by touching) can develop.

2. Another function of this (i.e., posterior parietal lobe) area is that its presence is necessary for the *maintenance of the body image*. What is body image? Examples will help to understand: we are aware of the:

- interrelationship between the different parts of the body (e.g. head is above the trunk),
- presence of the different parts of our body (e.g. right hand/left hand etc).
- presence of objects in the immediate surrounding space in relation to sides of the body (e.g. I am walking by the side of a wall and my left side is close to the wall; I am aware of both my left side as well as of the wall).

All these awareness are, however, really in a subconscious state and are called body image. Loss of one posterior parietal lobe can thus produce (parietal) neglect. A person suffering from such (unilateral) neglect shaves only one of his cheeks neglecting the other.

3. The parietal lobe also contains angular gyrus. Destruction of angular gyrus causes agraphia (inability to write) which is a form of aphasia (described later).

Lesion of area 3, 1 & 2, result in severe disturbances of sensations. Crude sense perceptions (e. g. perceptions of crude touch/temperature) however persist, but this is due to the presence of thalamus (thalamus can cause crude sense perceptions but not the finer ones). As a result, tactile discrimination, weight discrimination, localization of sensory stimulus, temperature discrimination (all of which are rather finer forms of sense perception) disappear. In addition there is also motor disturbances, like loss of muscle tone. Pain perception is however not much affected by damage of post central gyrus.

Lesions of area 5 & 7 cause a loss of stereognosis (tactile agnosia). Most spectacular effect of loss of area 5 and 7 might be explained as 'neglect'.

3. Occipital lobe

Major areas of this lobe are:

- primary visual area (area 17),
- visual association area (area 18), and
- occipital eye field (area 19).

The primary visual area, area 17 (also called the visuosensory area of area striata) is the highest area for perception of visual senses. Area 18 is the area where visual senses are interpreted and integrated in the light of past experience. For example, there is a light and area 17 'sees' the light. It is area 18 which interprets that the light is from a candle stick. Thus, this can be viewed as a homologue of stereognosis.

In the anterior part of the lateral surface of occipital lobe, lies the occipital eye field. Stimulation of this area causes, (a) movement of the eyes, and (b) visual hallucination (cf. frontal eye field stimulation, which causes only movement of the eyes but no hallucination).

Occipital lobe pathology

Extensive lesions of area 17 produces homonymous hemianopsia, whereas discrete lesions can produce scotoma.

Lesions of area 18 produce loss of the ability to 'recognize', in the corresponding field of vision.

4. Temporal lobe

Major areas in this lobe are:

- Primary auditory area, in the gyrus of Heschl (area 41) which is situated in the floor of the lateral sulcus plus a small part of the superior temporal gyrus.
- In area 41 sound is "heard" while in the Wernicke's area the sound is 'interpreted'.

Temporal lobe pathology

1. Because of the bilateral representation, lesion of area 41 of one temporal lobe does not cause deafness.

XI. EEG

Electrodes may be applied over the scalp of an individual and the electrical activity of the underlying brain may be recorded thereby. Such a record is called 'electroencephalogram', and the machine, electroencephalograph. The devise was introduced by the German neurologist Hans Berger in the 1930s and hence the records of EGG are also called 'Berger rhythm'. However, the great expectations of Berger (who thought many diseases of the brain could be diagnosed by EEG) were not fully realized and Berger ended as a much frustrated mentally broken down man.

Therefore, the neurons of the cerebral cortex, when their membrane potential is changed, generate varying electrical fields on the surface of the skull that can be recorded with suitable leads. The EEG can provide valuable clues to neuronal functions and as a result has gained great clinical importance.

1. Description of EEG waves

1. In adult awake persons, with the eyes closed and the mind relaxed and wandering, when a pair of electrodes are placed in the occipital region of the scalp, α-waves or α-rhythm (also called 'synchronised' waves) are seen. The waves have a 8 to 14 Hz frequency (1 Hz = 1 cycle/sec) and are

of about 50 mV amplitude. Further, the waves are sinusoidal. This is the fundamental EEG wave and may be obtained from temporal and parietal regions as well. However, in the frontal region the frequency is rather high and the amplitude low.

Now, if the person opens his eyes or engages his mind into a problem, the α-wave disappears and in its place comes a new type of wave, called 'a block' or 'desynchronization' wave or 'alert response', characterized by low voltage, fast, irregular waves. Desynchronization, therefore, equals to faster rhythm + low amplitude.

2. Besides, the α-wave and β-wave, the δ-waves may appear in EEG (Fig. 7–8). Alpha waves, a rhythm ranging from 8 to 13 Hz, are observed when the person is awake but relaxed with the eyes closed. When the eyes are open, the added visual input to the cortex imparts a faster rhythm to the EEG, ranging from 13 to 30 Hz and designated beta waves. The slowest waves recorded occur during sleep: theta waves at 4 to 7 Hz and delta waves at 0.5 to 4 Hz, in deepest sleep.

2. Variation of EEG

1. As stated already, *a-rhythm* is obtainable, mostly from the occipital region, when eyes are closed and the mind is wandering (relaxed). But on opening the eyes or attempting to work the mind, block rhythm (desynchronization rhythm) appears. *b-rhythm* means that the cerebral cortex is not very active and the mind is at rest whereas desynchronization means that the cortex is active and mind is not fully resting.
2. EEG waves vary according to the sleep stages (see Chapter #41).
3. The amplitude of the waves changes remarkably with several drugs, hypoglycemia and hypocapnia.
4. Abnormal wave patterns are seen *in epilepsy*, a neurological disorder of the brain characterized by spontaneous discharges of electrical activity, resulting in abnormalities ranging from momentary lapses of attention, to seizures of varying severity, to loss of consciousness if both brain hemispheres participate in the electrical abnormality. The characteristic waveform signifying seizure activity is the appearance of spikes or sharp peaks, as abnormally large numbers of units fire simultaneously. Examples of spike activity occurring singly and in a spike-and-wave pattern are shown in Fig. 7–8.

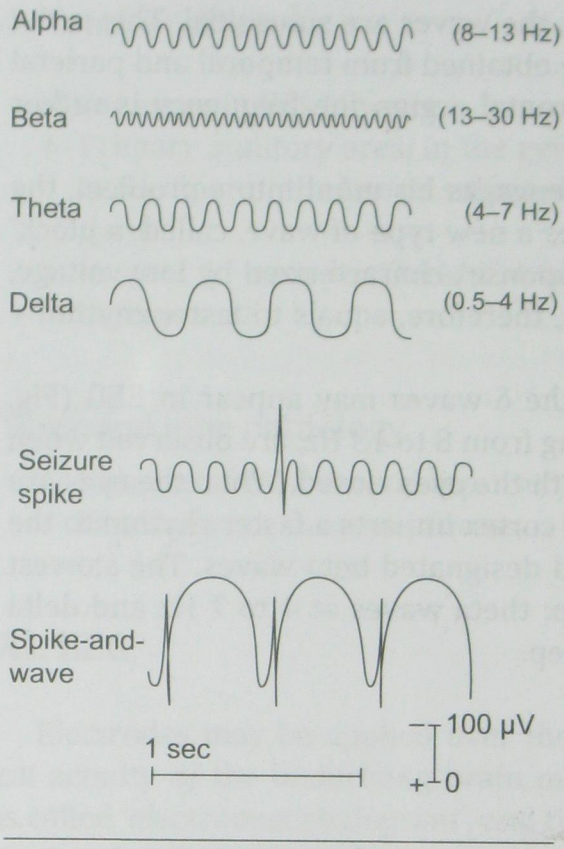


Fig. 7–8. EEG waves.

3. Mechanism of EEG development

The exact mechanism is still not very clear. However, from various evidences it appears that the EEG waves are the result of summated (= algebraically added up) potentials of the dendrites of the gray matter of the superficial layer of the cerebral cortex. These dendrite potentials are not action potential (AP). Therefore, the potential changes on the cortical surface largely depend on the ***postsynaptic potentials at dendrites of the pyramidal cells*** (Fig. 7–9).

Although the postsynaptic potentials have a lower amplitude than the AP, they last significantly longer. Because the pyramidal cells are positioned at right angles to the cortical surface, their local activity generates dipoles in the direction of the surface much more easily than other cells in the cortex. They thus have a much greater impact on the surface potential than other neurons. Furthermore, they are all orientated in parallel to one another, so that equidirectional potential changes of neighboring pyramidal cells are summated. EEG deflections are to be expected only if (around the lead electrode) several pyramidal cells are simultaneously depolarized, i.e., there is a synchronized event.

During an EPSP, Na^+ enters the cell and thus leaves behind a local negative extracellular potential (Fig. 7–9, 1). The depolarization promotes an efflux of K^+ ions along the remaining cell membrane, this efflux in turn generating a local positive extracellular potential. If an excitatory synapse at the apical end of a dendrite is activated, the extracellular space in the area is relatively negative, but relatively positive at the base of the dendrite. As a result a dipole is generated that creates a negative potential at the surface.

Commissural fibers from the other cortical hemisphere and nonspecific parts of the thalamus form excitatory synapses mainly at the surface; excitation

via these fibers thus leads to a negative potential at the surface electrode (Fig. 7–9, 1).

Conversely, activation of specific *thalamocortical fibers* are more likely to lead to positive potentials at the surface (Fig. 7–9, 2) because they act near the cell body, i.e., deep in the cerebral cortex. Inhibition in the area of the cell body theoretically results in a negative potential at the surface, but it is not strong enough to be registered at the surface of the scalp (Fig. 7–9, 3).

The neurons in the thalamus that excite the cortical pyramidal cells undergo a rhythmical activity due to negative feedback (Fig. 7–9, 4). This rhythm is

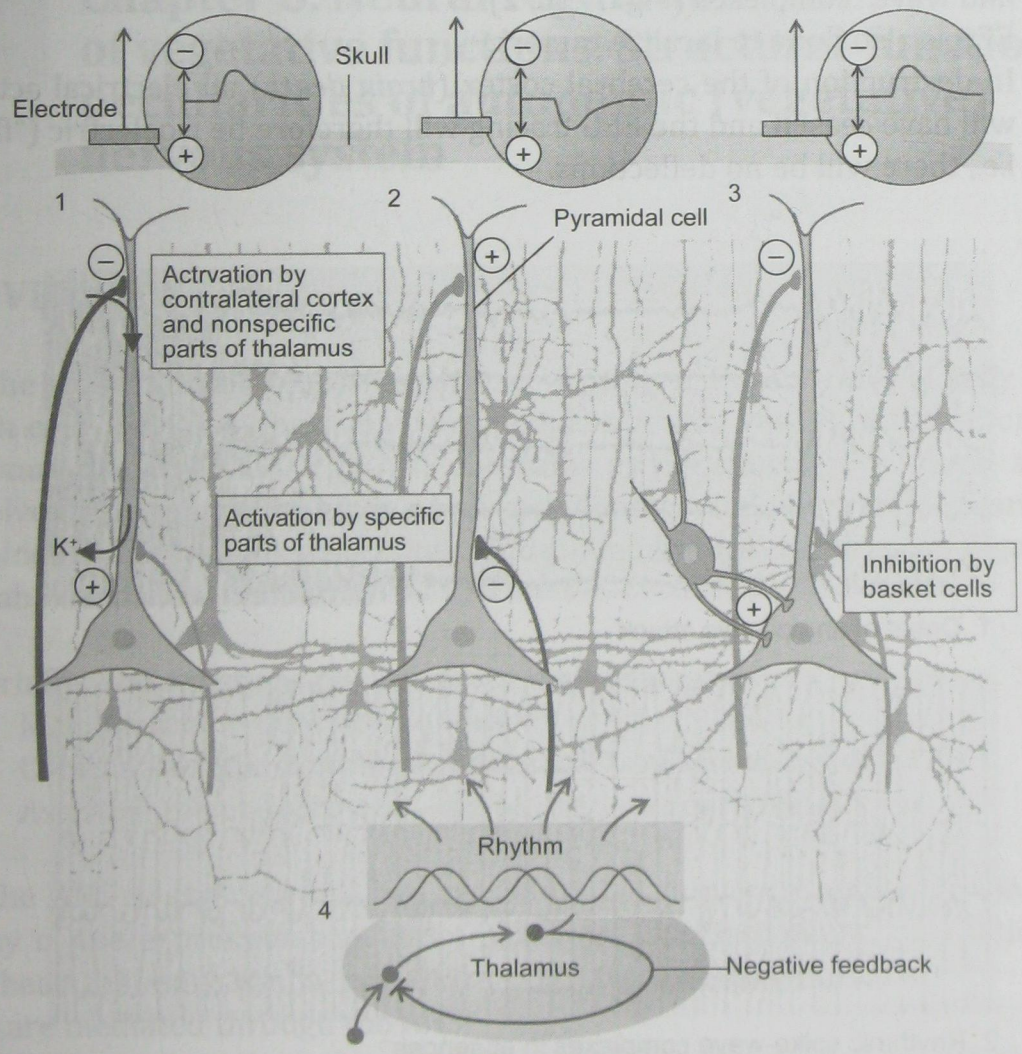
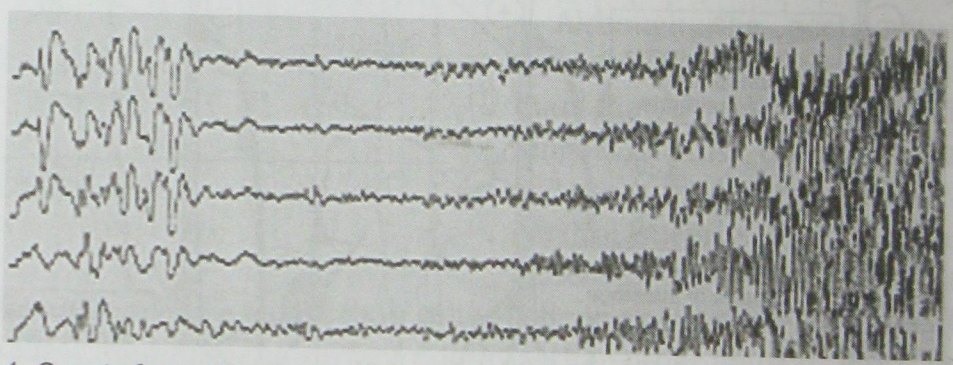


Fig. 7–9. Schematical illustrations of the EEG genesis.

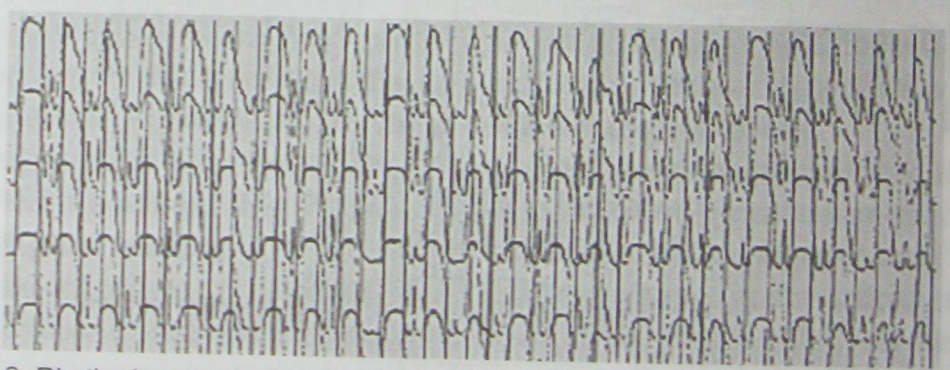
transmitted by the thalamocortical tracts to the pyramidal cells, with one thalamic neuron simultaneously exciting several pyramidal cells. Because of this, subcortical lesions are better registered in the EEG than small cortical ones.

4. EEG clinical importance

1. EEG recording is very commonly done in patients suffering from epilepsy, which is characterized by massive synchronized excitation of cortical neurons. It causes "spike" activity ("seizure spikes"; Fig. 7–10, 1) or "spike and wave" complexes (Fig. 7–10, 2).
2. EEG is also done to localize tumors.
3. In destruction of the cerebral cortex (*brain death*) all electrical activity will have ceased and the EEG tracing will therefore be isoelectric ("flat"), i.e., there will be no deflections.



1. Onset of an epileptic attack



2. Rhythmic spike-wave complexes in absences

Fig. 7–10. Types of EEG in different forms of the epileptic seizures

Unit IV PHYSIOLOGY OF AUTONOMIC NERVOUS SYSTEM

Chapter 8. Neural regulation of vegetative functions. Structural-functional peculiarities of autonomic (vegetative) nervous system

I. OVERVIEW

The ability to maintain homeostasis and perform the activities of daily living in an ever-changing physical environment is largely vested in the autonomic nervous system (ANS). The ANS functions at the subconscious level and is involved in regulating, adjusting, and coordinating vital visceral functions such as blood pressure and blood flow, body temperature, respiration, digestion, metabolism, and elimination.

Briefly, ANS functions fall into three major categories:

- Maintaining homeostatic conditions within the body;
- Coordinating the body's responses to exercise and stress;
- Assisting the endocrine system to regulate reproduction.

The ANS is strongly affected by emotional influences and is involved in many of the expressive aspects of behavior. Blushing, pallor, palpitations of the heart, clammy hands, and dry mouth are several emotional expressions that are mediated through the ANS.

As with the somatic nervous system, the ANS is represented in both the CNS and the peripheral nervous system (PNS). Traditionally, the ANS has been defined as a general efferent system innervating visceral organs. The efferent outflow

from the ANS has two divisions: the sympathetic nervous system and the parasympathetic nervous system. The afferent input to the ANS is provided by visceral afferent neurons, usually not considered to be part of the ANS. Besides, on the basis of anatomic, functional, and neurochemical differences, the ANS is usually subdivided into three divisions: sympathetic, parasympathetic, and enteric.

II. ANS SYMPATHETIC AND PARASYMPATHETIC DIVISIONS INTERACTIONS

The functions of the sympathetic and parasympathetic divisions are often simplified into a two-part scheme. The sympathetic division is said to preside over the utilization of metabolic resources and emergency responses of the body. The parasympathetic division presides over the restoration and buildup of the body's reserves and the elimination of waste products. In reality, most of the organs supplied by the ANS receive both sympathetic and parasympathetic innervation. In many instances, the two divisions are activated in a reciprocal fashion, so that if the firing rate in one division is increased, the rate is decreased in the other. An example is controlling the heart rate: increased firing in the sympathetic nerves and simultaneous decreased firing in the parasympathetic nerves result in increased heart rate.

The functions of the sympathetic nervous system include maintaining body temperature and adjusting blood flow and blood pressure to meet the changing needs of the body that occur with activities of daily living, such as moving from the supine to the standing position. The sympathoadrenal system also can discharge as a unit when there is a critical threat to the integrity of the individual — the so-called fight-or-flight response.

During a stress situation, the heart rate accelerates; the blood pressure rises; blood flow shifts from the skin and gastrointestinal tract to the skeletal muscles and brain; blood sugar increases; the bronchioles and pupils dilate; the sphincters of the stomach and intestine and the internal sphincter of the urethra constrict; and the rate of secretion of exocrine glands that are involved in digestion diminishes. Emergency situations often require vasoconstriction and shunting of blood away from the skin and into the muscles and brain, a mechanism that, should a wound occur, provides for a reduction in blood flow and preservation of vital functions needed for survival.

Sympathetic function often is summarized as catabolic in that its actions predominate during periods of pronounced energy expenditure, such as when

survival is threatened. In contrast to the sympathetic nervous system, the functions of the parasympathetic nervous system are concerned with conservation of energy, resource replenishment and storage (*i.e.*, anabolism), and maintenance of organ function during periods of minimal activity. The parasympathetic nervous system slows heart rate, stimulates gastrointestinal function and related glandular secretion, promotes bowel and bladder elimination, and contracts the pupil, protecting the retina from excessive light during periods when visual function is not vital to survival.

The two divisions of the ANS usually are viewed as having opposite and antagonistic actions (*i.e.*, if one activates, the other inhibits a function). Exceptions are functions, such as sweating and regulation of arteriolar blood vessel diameter, that are controlled by a single division of the ANS, in this case the sympathetic nervous system.

The sympathetic and parasympathetic nervous systems are continually active. The effect of this continual or basal (baseline) activity is referred to as *tone*. The tone of an effector organ or system can be increased or decreased and usually is regulated by a single division of the ANS. For example, vascular smooth muscle tone is controlled by the sympathetic nervous system. Increased sympathetic activity produces local vasoconstriction from increased vascular smooth muscle tone, and decreased activity results in vasodilatation caused by decreased tone. In structures such as the sinoatrial node and atrioventricular node of the heart, which are innervated by both divisions of the ANS, one division predominates in controlling tone. In this case, the tonically active parasympathetic nervous system exerts a constraining or braking effect on heart rate, and when parasympathetic outflow is withdrawn, similar to releasing a brake, the heart rate increases. The increase in heart rate that occurs with vagal withdrawal can be further augmented by sympathetic stimulation.

III. ANS EFFERENT REFLECTORY PATHWAYS

In the ANS, the efferent path consists of a two-neuron chain with a synapse interposed between the CNS and the effector cells (Fig. 8–1). The cell bodies of the autonomic motor neurons are located in the spinal cord or specific brainstem nuclei. The first motoneuron, called the *preganglionic neuron*, lies in the intermediolateral cell column in the ventral horn of the spinal cord or its equivalent location in the brainstem.

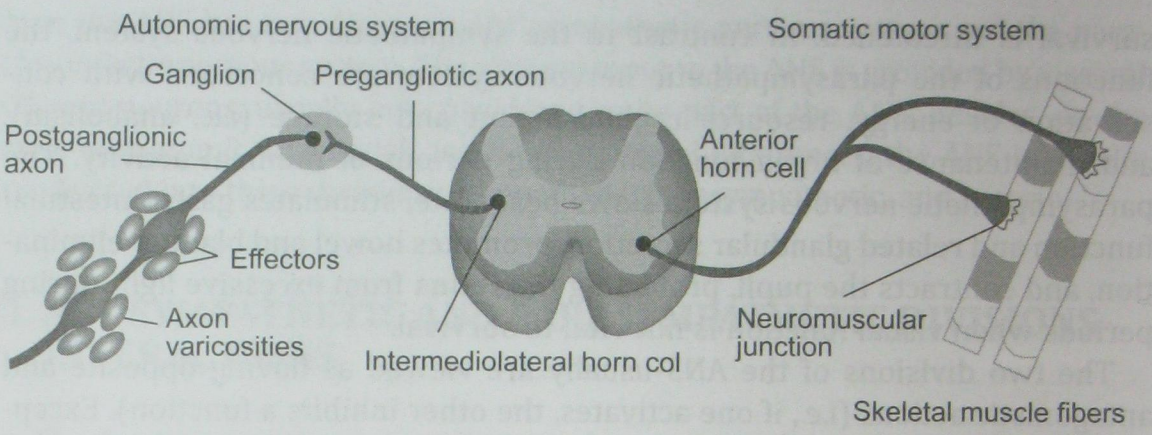


Fig. 8–1. Schematic illustration of the ANS efferent pathway compared with the somatic motor system.

The second motoneuron, called the *postganglionic neuron*, synapses with a preganglionic neuron in an autonomic ganglion located in the PNS. An efferent fiber emerges as the *preganglionic axon* and then synapses with neurons located in a peripheral ganglion. The neuron in the ganglion then projects a *postganglionic axon* to the autonomic effector cells.

The two divisions of the ANS differ in terms of location of preganglionic cell bodies, relative length of preganglionic fibers, general function, nature of peripheral responses, and preganglionic and postganglionic neuromediators (Table 8–1). This two-neuron outflow pathway and the interneurons in the autonomic ganglia that add further modulation to ANS function are features distinctly different from the arrangement in somatic motor innervation.

The fibers of the sympathetic nervous system are distributed to effectors throughout the body, and as a result, sympathetic actions tend to be more diffuse than those of the parasympathetic nervous system, in which there is a more localized distribution of fibers. The preganglionic fibers of the sympathetic nervous system may traverse a considerable distance and pass through several ganglia before synapsing with postganglionic neurons, and their terminals make contact with a large number of postganglionic fibers. In some ganglia, the ratio of preganglionic to postganglionic cells may be 1 : 20; because of this, the effects of sympathetic stimulation are diffuse. There is considerable overlap, and one ganglion cell may be supplied by several preganglionic fibers.

In contrast to the sympathetic nervous system, the parasympathetic nervous system has its postganglionic neurons located very near or in the organ of

Table 8–1. Properties of the sympathetic and parasympathetic compartment of the ANS

Characteristic	Sympathetic compartment	Parasympathetic compartment
Location of preganglionic cell bodies	T_1-T_{12}, L_1 and L_2	Cranial nerves: III, VII (n. intermedius), IX, X; S_2, S_3, and S_4
Relative length of preganglionic fibers	Short -to paravertebral chain of ganglia or to aortic prevertebral of ganglia	Long — to ganglion cells near or in the innervated organ
General function	Catabolic — mobilizes resources in anticipation of challenge for survival (preparation for "fight-or-flight" response)	Anabolic — concerned with conservation, renewal, and storage of resources
Nature of peripheral response	Generalized	Localized
Transmitter between preganglionic terminals and postganglionic neurons	ACh	ACh
Transmitter of postganglionic neuron	ACh (sweat glands and skeletal muscle vasodilator fibers); NE (most synapses); NE and E (secreted by adrenal gland)	ACh

innervation. Because the ratio of preganglionic to postganglionic communication often is 1 : 1, the effects of the parasympathetic nervous system are much more circumscribed.

IV. SYMPATHETIC NERVOUS SYSTEM

The preganglionic neurons of the sympathetic nervous system are located primarily in the thoracic and upper lumbar segments (Th_1 to L_2) of the spinal cord; thus, the sympathetic nervous system often is referred to as the *thoracolumbar division* of the ANS. These preganglionic neurons, which are located

primarily in the ventral horn intermediolateral cell column, have axons that are largely myelinated and relatively short. The postganglionic neurons of the sympathetic nervous system are located in the paravertebral ganglia of the sympathetic chain that lie on either side of the vertebral column, or in prevertebral sympathetic ganglia such as the celiac ganglia (Fig. 8–2).

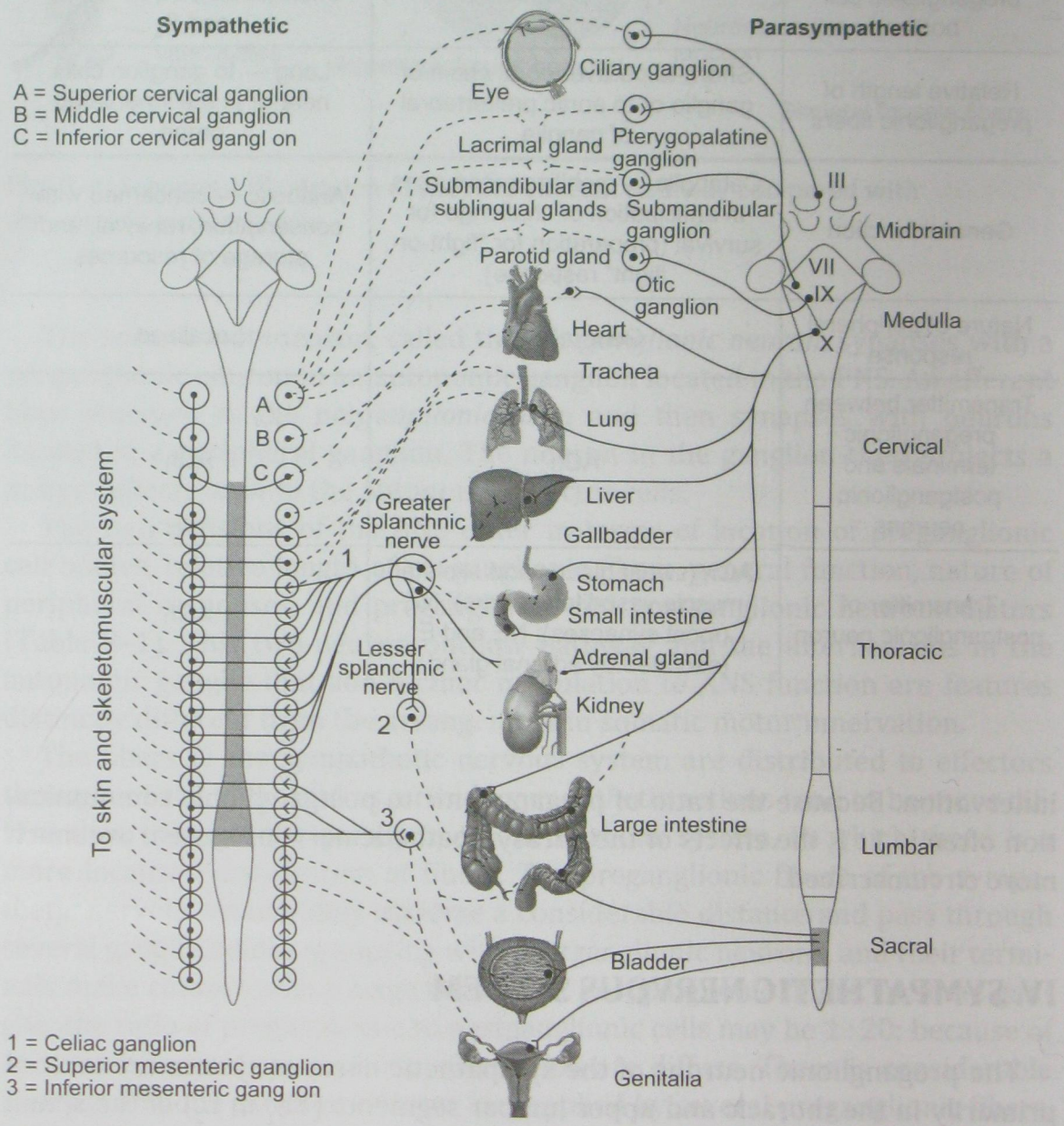


Fig. 8–2. The general scheme of the ANS organization.

In addition to postganglionic efferent neurons, the sympathetic ganglia contain neurons of the internuncial, short-axon type, similar to those associated with complex circuitry in the brain and spinal cord. Many of these inhibit and others modulate preganglionic-to-postganglionic transmission. The axons of the preganglionic neurons leave the spinal cord through the ventral root of the spinal nerves (Th_1 to L_2), enter the ventral primary rami, and leave the spinal nerve through white rami of the rami communicantes to reach the paravertebral ganglionic chain (Fig. 8–3). In the sympathetic chain of ganglia, preganglionic fibers may synapse with neurons of the ganglion it enters, pass up or

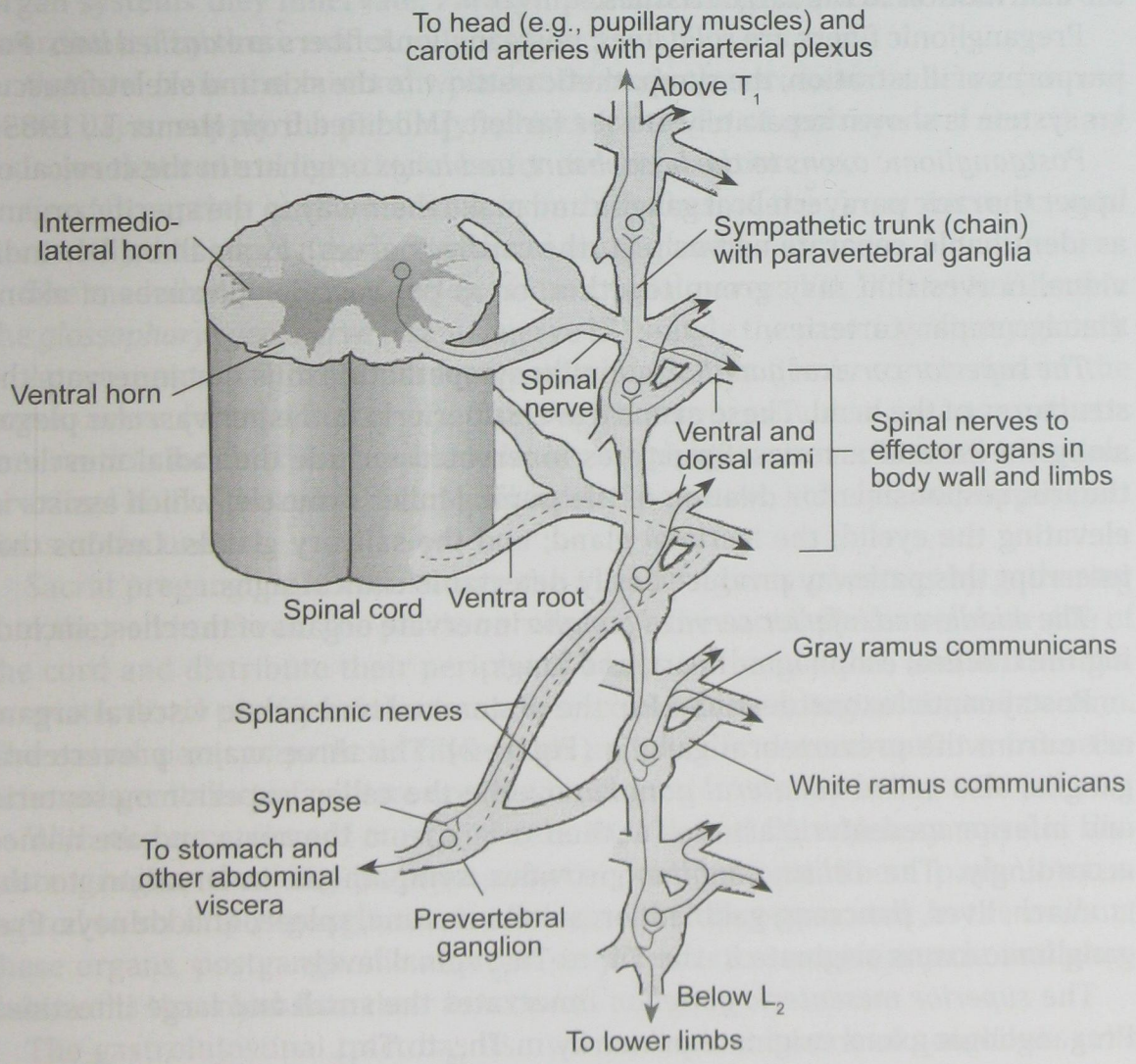


Fig. 8–3. Sympathetic pathways.

down the chain and synapse with one or more ganglia, or pass through the chain and move outward through a splanchnic nerve to terminate in one of the prevertebral ganglia (*i.e.*, celiac, superior mesenteric, or inferior mesenteric) that are scattered along the dorsal aorta and its branches. The adrenal medulla, which is part of the sympathetic nervous system, contains postganglionic sympathetic neurons that secrete sympathetic neurotransmitters directly into the bloodstream.

Postganglionic axons that are destined for somatic structures — such as sweat glands, pilomotor muscles, or blood vessels of the skin and skeletal muscles — leave the paravertebral ganglion in the gray ramus and rejoin the spinal nerve for distribution to the target tissues.

Preganglionic fibers are *solid lines;* postganglionic fibers are *dashed lines.* For purposes of illustration, the sympathetic outflow to the skin and skeletomuscular system is shown separately –to the far left. [Modified from Hemer L., 1983].

Postganglionic axons to the head, heart, and lungs originate in the cervical or upper thoracic paravertebral ganglia and make their way to the specific organs as identifiable, separate nerves (e.g., the cardiac nerves), as small-caliber individual nerves that may group together, or as perivascular plexuses of axons that accompany arteries.

The superior cervical ganglion supplies sympathetic axons that innervate the structures of the head. These axons travel superiorly in the perivascular plexus along the carotid arteries. Structures innervated include the radial muscle of the iris, responsible for dilation of the pupil; Muller's muscle, which assists in elevating the eyelid; the lacrimal gland; and the salivary glands. Lesions that interrupt this pathway produce easily detectable clinical signs.

The middle and inferior cervical ganglia innervate organs of the chest, including the trachea, esophagus, heart, and lungs.

Postsynaptic axons destined for the abdominal and pelvic visceral organs arise from the prevertebral ganglia (Fig. 8–2). The three major prevertebral ganglia, also called *collateral ganglia*, overlie the celiac, superior mesenteric, and inferior mesenteric arteries at their origin from the aorta and are named accordingly. The *celiac ganglion* provides sympathetic innervation to the stomach, liver, pancreas, gallbladder, small intestine, spleen, and kidneys. Preganglionic axons originate in the Th_5 to Th_{12} spinal levels.

The *superior mesenteric ganglion* innervates the small and large intestines. Preganglionic axons originate primarily in Th_{10} to Th_{12}.

The *inferior mesenteric ganglion* innervates the lower colon and rectum, urinary bladder, and reproductive organs. Preganglionic axons originate in L_1 to L_3.

V. PARASYMPATHETIC NERVOUS SYSTEM

The preganglionic fibers of the parasympathetic nervous system, also referred to as the *craniosacral division* of the ANS, originate in some segments of the brain stem and sacral segments of the spinal cord (Fig. 8–2).

The central regions of origin are the midbrain, pons, medulla oblongata, and the sacral part of the spinal cord. Four of the twelve cranial nerves — numbers III, VII, IX, and X — contain parasympathetic axons. The nuclei of these nerves, which occupy areas of the tectum in the midbrain, pons, and medulla, are the centers for the initiation and integration of autonomic reflexes for the organ systems they innervate. Parasympathetic and sympathetic activities are coordinated by these nuclei.

Outflow from the midbrain passes through the *oculomotor nerve* (cranial nerve III) to supply the papillary sphincter muscle of each eye and the ciliary muscles that control lens thickness for accommodation.

Caudal pontine outflow comes from branches of the *facial nerve* (cranial nerve VII) that supply the lacrimal and nasal glands.

The medullary outflow develops from cranial nerves VII, IX, and X. Fibers in the *glossopharyngeal nerve* (cranial nerve IX) supply the parotid salivary glands.

Approximately 75% of parasympathetic efferent fibers are carried in the *vagus nerve* (cranial nerve X). The vagus nerve provides parasympathetic innervation for the heart, trachea, lungs, esophagus, stomach, small intestine, proximal half of the colon, liver, gallbladder, pancreas, kidneys, and upper portions of the ureters.

Sacral preganglionic axons leave the S_2 *to* S_4 *segmental nerves* by gathering into the pelvic nerves. The pelvic nerves leave the sacral plexus on each side of the cord and distribute their peripheral fibers to the bladder, uterus, urethra, prostate, distal portion of the transverse colon, descending colon, and rectum. The sacral parasympathetic fibers also supply the venous outflow from the external genitalia to facilitate erectile function.

With the exception of cranial nerves III, VII, and IX, which synapse in discrete ganglia, the long parasympathetic preganglionic fibers pass uninterrupted to short postganglionic fibers located in the organ wall. In the walls of these organs, postganglionic neurons send axons to smooth muscle and glandular cells that modulate their functions.

The gastrointestinal tract has its own intrinsic network of ganglionic cells located between the smooth muscle layers, called the *enteric* (or *intramural*) *plexus*, which controls local peristaltic movements and secretory functions.

This network of parasympathetic postganglionic neurons and interneurons runs from the upper portion of the esophagus to the internal anal sphincter. Local afferent sensory neurons respond to mechanical and chemical stimuli and communicate these influences to motor fibers in the enteric plexus. The number of neurons in the enteric neural network (108) is so large that it approximates that of the spinal cord. It is thought that this enteric nervous system is capable of independent function without control from CNS fibers. The CNS has a modulating role, by way of preganglionic innervation of the plexus, converting local peristalsis to longer-distance movements, thereby speeding the transit of intestinal contents.

VI. AUTONOMIC NEUROTRANSMISSION

The generation and transmission of impulses in the ANS occur in the same manner as in the CNS. There are self-propagating action potentials with transmission of impulses across synapses and other tissue junctions by way of neurohumoral transmitters.

The postganglionic fibers of the ANS form a diffuse neural plexus at the site of innervation. The membranes of the cells of many smooth muscle fibers are connected by conductive protoplasmic bridges, called *gap junctions*, that permit rapid conduction of impulses through whole sheets of smooth muscle, often in repeating waves of contraction. Autonomic neurotransmitters released near a limited portion of these fibers provide a modulating function extending to a large number of effector cells. The muscle layers of the gut and of the bladder wall are examples.

The main neurotransmitters of the ANS are *acetylcholine* and the catecholamines — *epinephrine* and *norepinephrine*. In the somatic nervous system, neurotransmitter is released from specialized nerve endings that make intimate contact with the target structure. The mammalian motor endplate, with one nerve terminal to one skeletal muscle fiber, illustrates this principle. This arrangement contrasts with the ANS, where postganglionic axons terminate in varicosities, swellings enriched in synaptic vesicles, which release the transmitter into the extracellular space surrounding the effector cells. The response to the ANS output originates in some of the effector cells and then propagates to the remainder via gap junctions.

Acetylcholine (ACh) is the transmitter released by the preganglionic nerve terminals of both the sympathetic and the parasympathetic divisions (Fig. 8–4).

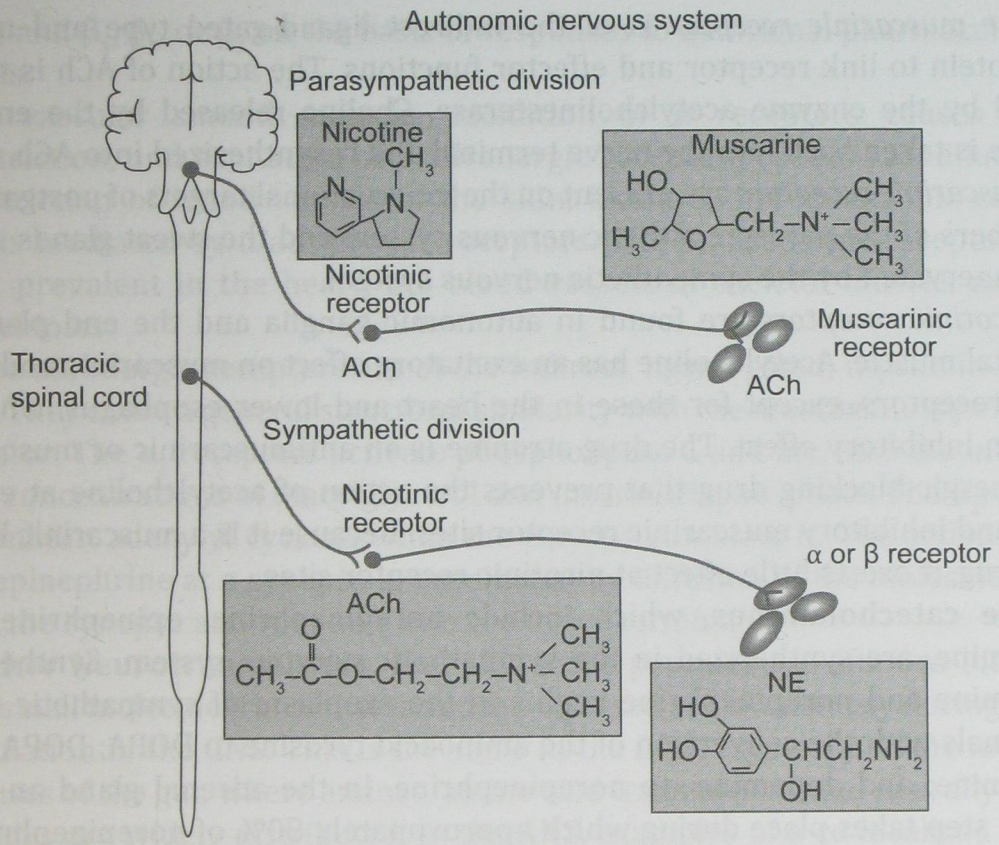


Fig. 8–4. The neurochemistry of the autonomous pathways.

The synapse at those sites utilizes a nicotinic receptor similar in structure to the receptor at the neuromuscular junction.

Parasympathetic postganglionic neurons release ACh at the synapse with the effectors. The postganglionic sympathetic neurons to the sweat glands and to some blood vessels in skeletal muscle also use ACh as the neurotransmitter. The synapse between the postganglionic neuron and the target tissues utilizes a *muscarinic receptor*. This receptor classification scheme is based on the response of the synapses to the alkaloids nicotine and muscarine, which act as agonists at their respective type of synapse. The nicotinic receptor of the ANS is blocked by the antagonist hexametho-nium, in contrast to the neuromuscular junction receptor, which is blocked by curare. The muscarinic receptor is blocked by atropine.

The nicotinic receptor is of the direct ligand-gated type, meaning that the receptor and the ion channel are contained in the same structure.

The muscarinic receptor is of the indirect ligand-gated type and uses a G-protein to link receptor and effector functions. The action of ACh is terminated by the enzyme acetylcholinesterase. Choline released by the enzyme action is taken back into the nerve terminal and resynthesized into ACh.

Muscarinic receptors are present on the innervational targets of postganglionic fibers of the parasympathetic nervous system and the sweat glands which are innervated by the sympathetic nervous system.

Nicotinic receptors are found in autonomic ganglia and the end plates of skeletal muscle. Acetylcholine has an excitatory effect on muscarinic and nicotinic receptors, except for those in the heart and lower esophagus, where it has an inhibitory effect. The drug atropine is an antimuscarinic or muscarinic cholinergic-blocking drug that prevents the action of acetylcholine at excitatory and inhibitory muscarinic receptor sites. Because it is a muscarinic blocking drug, it exerts little effect at nicotinic receptor sites.

The catecholamines, which include norepinephrine, epinephrine, and dopamine, are synthesized in the sympathetic nervous system. Synthesis of dopamine and norepinephrine begins in the axoplasm of sympathetic nerve terminals with the conversion of the aminoacid tyrosine to DOPA; DOPA — to dopamine; and dopamine -to norepinephrine. In the adrenal gland an additional step takes place during which approximately 80% of norepinephrine is transformed into epinephrine.

The catecholamine *norepinephrine (NE)* is the neurotransmitter for postganglionic synapses of the sympathetic division (Fig.8-4). Catecholamines can cause excitation or inhibition of smooth muscle contraction, depending on the site, dose, and type of receptor present. The excitatory or inhibitory responses of organs to sympathetic neurotransmitters are mediated by interaction with special structures in the cell membrane called *receptors*.

The synapses that utilize NE receptors can also be activated by the closely related compound epinephrine (adrenaline), which is released into the general circulation by the adrenal medulla — hence, the original designation of these type receptors as adrenergic.

Adrenergic receptors are classified as either alpha (α) or beta (β), based on their responses to pharmacological agents that mimic or block the actions of NE and related compounds. *Alpha receptors* respond best to epinephrine, less well to NE, and least well to the synthetic compound isoproterenol. *Beta receptors* respond best to isoproterenol, less well to epinephrine, and least well to NE. Propranolol is a drug that acts as an antagonist at α receptors but has no action on β receptors. Each class of receptors is further classified as α_1

or α_2, and β_1, β_2, or β_3 on the basis of responses to additional pharmacological agents.

In vascular smooth muscle, excitation of α receptors causes vasoconstriction, and excitation of β-adrenergic receptors causes vasodilatation. Constriction of blood vessels in the skin, kidneys, and splanchnic circulation is mediated by α-adrenergic receptors. The β-adrenergic receptors are most prevalent in the heart, the blood vessels of skeletal muscle, and the bronchioles.

The adrenergic receptors are of the indirect, ligand gated, G-protein-linked type. They share a general structural similarity with the muscarinic type of ACh receptor. The α_1 receptors activate phospholipase C and increase the intracellular concentrations of diacylglycerol and inositoltrisphosphate. The α_2, receptors inhibit adenylyl cyclase, while the β types stimulate it. The action of NE and epinephrine at a synapse is terminated by diffusion of the molecule away from the synapse and reuptake into the nerve terminal.

Other Neurotransmitters. Neurally active peptides are often co-localized with small molecule transmitters and are released simultaneously during nerve stimulation in the CNS. This is the same in the ANS, especially in the intrinsic plexuses of the gut, where amines, amino acid transmitters, and neurally active peptides are widely distributed. In the ANS, examples of a co-localized amine and peptide are seen in the sympathetic division, where NE and neuropeptides Y are co-released by vasoconstrictor nerves. Vasoactive intestinal polypeptide (VIP) and calcitonin-gene-related peptide (CGRP) are released along with ACh from nerve terminals innervating the sweat glands.

Nitric oxide is another type of neurotransmitter produced by some autonomic nerve endings. The definition *nonadrenergic noncholinergic* has been applied to such nerves. Nitric oxide is a highly diffusible substance important in the regulation of smooth muscle contraction.

VII. CENTRAL REGULATIVE INFLUENCES

General visceral afferent fibers accompany the sympathetic and parasympathetic outflow into the spinal and cranial nerves, bringing chemoreceptor, pressure, and nociceptive information from organs of the viscera to the brain stem, thoracolumbar cord, and sacral cord. Local reflex circuits relating visceral afferent and autonomic efferent activity are integrated into a hierarchic control system in the spinal cord and brain stem. Progressively greater complexity in

the responses and greater precision in their control occur at each higher level of the nervous system.

For most autonomic-mediated functions, *the hypothalamus* serves as the major control center (Fig. 8–5). The hypothalamus, which has connections with the cerebral cortex, the limbic system, and the pituitary gland, is in a prime position to receive, integrate, and transmit information to other areas of the nervous system. The neurons concerned with thermoregulation, thirst, and feeding behaviors are found in the hypothalamus. The hypothalamus also is the site for integrating neuroendocrine function. Hypothalamic releasing and inhibiting hormones control the secretion of anterior pituitary hormones.

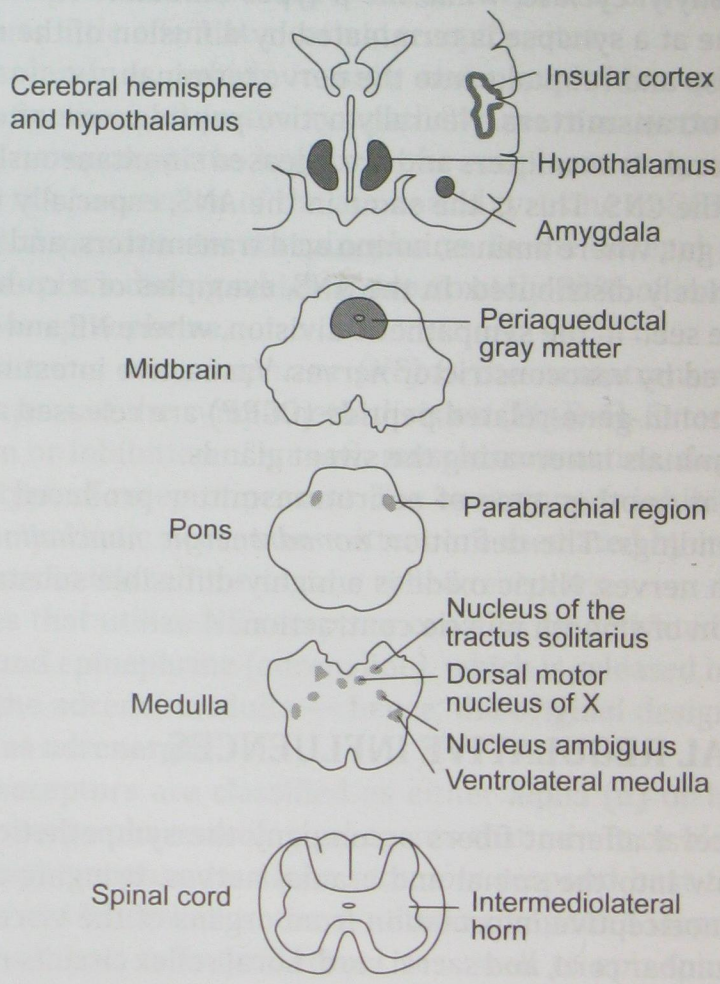


Fig. 8–5. The central autonomic network

The organization of many life-support reflexes occurs in the *reticular formation of the medulla* and *pons* (Fig. 8–5). These areas of reflex circuitry, often called centers, produce complex combinations of autonomic and somatic efferent functions required for the cough, sneeze, swallow, and vomit reflexes, as well as for the more purely autonomic control of the cardiovascular system. At the hypothalamic level, these reflexes are integrated into more general response patterns, such as rage, defensive behavior, eating, drinking, voiding, and sexual function. Forebrain and especially limbic system control of these behaviors involves inhibiting or facilitating release of the response patterns according to social pressures during learned emotion provoking situations.

Reflex adjustments of cardiovascular and respiratory function occur at the level of the *brainstem*. A prominent example is the carotid sinus baroreflex. One of the striking features of ANS function is the rapidity and intensity with which it can change visceral function. Within 3 to 5 seconds, it can increase heart rate to approximately twice its resting level. Bronchial smooth muscle tone is largely controlled by parasympathetic fibers carried in the vagus nerve. These nerves produce mild to moderate constriction of the bronchioles.

Other important ANS reflexes are located at the level of the *spinal cord* (Fig. 8–5). As with other spinal reflexes, these reflexes are modulated by input from higher centers. When there is loss of communication between the higher centers and the spinal reflexes, as occurs in spinal cord injury, these reflexes function in an unregulated manner.

Unit V ENDOCRINE PHYSIOLOGY

Chapter 9. General endocrinology. The hypothalamic-pituitary system

Endocrinology is the branch of physiology concerned with the description and characterization of processes involved in the regulation and integration of cells and organ systems by a group of specialized chemical substances called hormones. The diagnosis and treatment of a large number of endocrine disorders is an important aspect of any general medical practice. Certain endocrine disease states, such as diabetes mellitus, thyroid disorders, and reproductive disorders, are fairly common in the general population; therefore, it is likely that they will be encountered repeatedly in the practice of medicine.

The endocrine system is involved in all of the integrative aspects of life, including growth, sex differentiation, metabolism, and adaptation to an ever-changing environment.

I. HORMONES

The endocrine system uses chemical substances called *hormones* as a means of regulating and integrating body functions.

1. *Definition. Hormones* (from Greece 'hormaino' — to do something, to move something) — *the group of the biologically active substances synthesized by*:

- endocrinal glands (i.e., thyroid gland);
- endocrinal tissue of the organs that perform non-endocrinal functions (i.e., the pancreatic gland);
- endocrinal cells diffusely localized in an organism (i.e., APUD system of the digestive system, cells of the kidneys and atrium).

The English physiologists Beylis and Starling firstly proposed the term 'hormone'.

Hormones are secreted into the interstitial space and diffuse from there into the bloodstream. As a result, hormones can exert their effects over relatively long distances. Because hormones are greatly diluted in the bloodstream and interstitial fluid, they must be able to act at very low concentrations (typically 10^{-8} M). The following processes are involved into hormonal signal transduction:

- synthesis
- release
- transport to target cell
- detection of hormonal signal by a specific receptor
- transduction of a signal into the cell, resulting in a change in cellular metabolism
- removal of signal (regulation)

Some, but not all hormones operate by causing cells to change expression of individual genes or gene sets; other hormones affect the action of already existing enzymes or cell structures.

A characteristic of hormones is that a single hormone can exert various effects in different tissues or, conversely, a single function can be regulated by several hormones. For example, estradiol, which is produced by the ovary, can act on the ovarian follicles to promote their maturation, on the uterus to stimulate its growth and maintain the cyclic changes in the uterine mucosa, on the mammary gland to stimulate ductal growth, on the hypothalamic-pituitary system to regulate the secretion of gonadotropins and prolactin, and on general metabolic processes to affect adipose tissue distribution. Lipolysis, which is the release of free fatty acids from adipose tissue, is an example of a single function that is regulated by several hormones, including the catecholamines, glucagon, and secretin. Table 9-1 lists the major functions and sources of body hormones.

The events involved in responding to a hormone include detection of the hormonal signal by a specific receptor; transduction of a signal into the cell, resulting in a change in the cellular metabolism; and removal of the signal (regulation). Numerous amounts of hormones resulted in their metabolic, morphogenetic, kinetic and correcting (reactogenic) influences on the organism.

Although the endocrine system once was thought to consist solely of discrete endocrine glands, it is now known that a number of other tissues release chemical messengers that modulate body processes.

Table 9–1. Main Action and the Sources of the Selected Hormones

Place of Origin	Hormone	Main Action
Hypothalamus	Releasing and inhibiting hormones: Corticotropin-releasing hormone (CRH Thyrotropin-releasing hormone (TRH) Growth hormone-releasing hormone (GHRH Gonadotropin-releasing hormone (GnRH)	Controls the release of pituitary hormones
Anterior pituitary	Growth hormone (GH)	Stimulates growth of bone and muscle, promotes protein synthesis and fat metabolism, decreases carbohydrate metabolism
	Adrenocorticotropic hormone (ACTH)	Stimulates synthesis and secretion of adrenal cortical hormones
	Thyroid-stimulating hormone (TSH)	Stimulates synthesis and secretion of thyroid hormone
	Follicle-stimulating hormone (FSH)	*Female:* stimulates growth of ovarian follicle, ovulation *Male:* stimulates sperm production
	Luteinizing hormone (LH)	*Female:* stimulates development of corpus luteum, release of oocyte, production of estrogen and progesterone *Male:* stimulates secretion of testosterone, development of interstitial tissue of testes
	Rolactin	Essential for milk production by lactating mammary glands
	Melanocite-stimulating hormone	Essential for pigment turnover
Posterior pituitary	Vasopressin, or Antidiuretic hormone (ADH) Oxytocin	Increases water reabsorption by the kidney Stimulates contraction of pregnant uterus, milk ejection from breasts after childbirth

Adrenal cortex	Mineralocorticosteroids, mainly aldosterone	Increases sodium absorption, potassium loss by the kidney
	Glucocorticoids, mainly cortisol	Affects metabolism of all nutrients; regulates blood glucose levels, affects growth, has anti-inflammatory action, and decreases effects of stress
	Adrenal androgens, mainly dehydroepiandrosterone (DHEA) and androstenedione	Have minimal intrinsic androgenic activity; they are converted to testosterone and dihydrotestosterone in the periphery
Adrenal medulla	Epinephrine Norepinephrine	Serve as neurotransmitters for the sympathetic nervous system
Thyroid gland: Follicular cells	Thyroid hormones: triiodothyronine (T_3), thyroxine (T_4)	Increase the metabolic rate; increase protein and bone turnover; increase responsiveness to catecholamines; necessary for fetal and infant growth and development
Thyroid C-cells	Calcitonin	Lowers blood Ca^{2+} and phosphate levels
Parathyroid glands	Parathyroid hormone	Regulates serum calcium
Pancreatic islet cells	Insulin	Lowers blood glucose by facilitating glucose transport across cell membranes of muscle, liver, and adipose tissue
	Glucagon	Increases blood glucose concentration by stimulation of glycogenolysis and glyconeogenesis
	Somatostatin	Delays intestinal absorption of glucose
Kidney	1,25-Dihydroxyvitamin D	Stimulates calcium absorption from the intestine
	Renin	Increases arterial pressure
	Erythropoietin	Essential for erythropoiesis
Ovaries	Estrogsen	Affects development of female sex organs and secondary sex characteristics
	Progesterone	Influences menstrual cycle; stimulates growth of uterine wall; maintains pregnancy
Testes	Androgens, mainly testosterone	Affect development of male sex organs and secondary sex characteristics; aid in sperm production

The functions of the endocrine system are closely linked with those of the nervous system and the immune system. For example, neurotransmitters such as epinephrine can act as neurotransmitters or as hormones. The mechanisms by which hormone binding results in a metabolic response will be discussed below.

2. *Types of hormonal secretion.* In the past, hormones were described as chemical substances that were released into the bloodstream and transported to distant target sites, where they exerted their action (Fig. 9–1). Although many hormones travel by this mechanism, some hormones and hormone-like substances never enter the bloodstream but instead act locally in the vicinity in which they are released. When they act locally on cells other than those that produced the hormone, the action is called

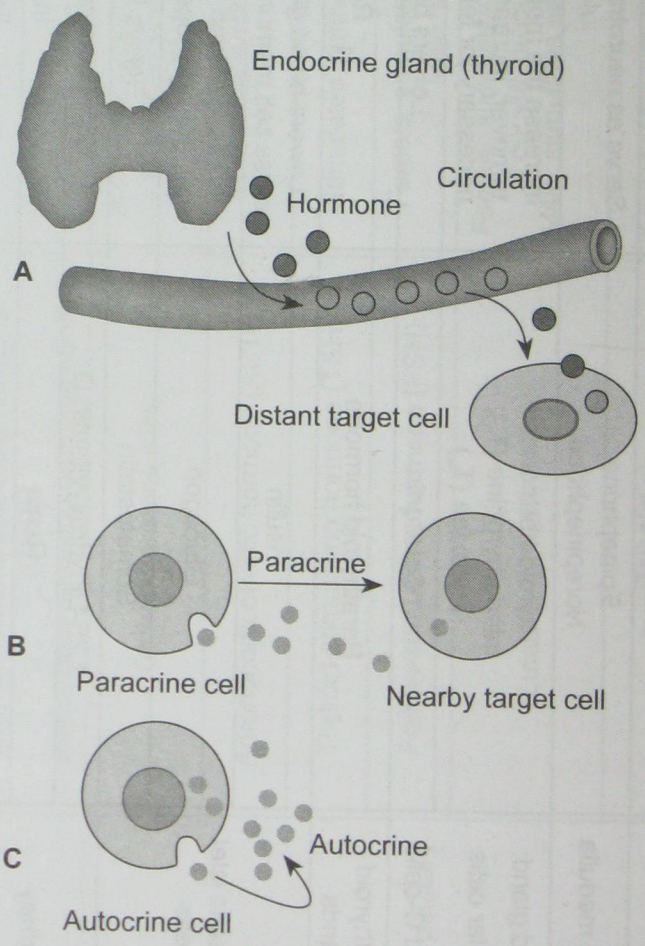


Fig. 9–1. Schematical illustration of endocrine (A), paracrine (B) and autocrine (C) types of secretion.

paracrine. The action of sex steroids on the ovary is a paracrine action. Hormones also can exert an *autocrine* action on the cells from which they were produced. For example, the release of insulin from pancreatic beta cells can inhibit its release from the same cells. *Juxtacrine* refers to a mechanism whereby a cytokine that is embedded in, bound to, or associated with the plasma membrane of one cell interacts with a specific receptor in a juxtaposed cell.

3. *Hormones' classification.*

Hormones can be categorized by a number of criteria. Grouping them by chemical structure is convenient, since in many cases, hormones with similar structures also use similar mechanisms to produce their biological effects. In addition, hormones with similar chemical structures are usually produced by tissues with similar embryonic origins. All hormones according to their biochemical structure divided into 3 groups:

- hormones of the protein nature;
- hormones of the steroid nature;
- derivatives of amino acids.

Table 9–2. Represents the main chemical groups of the hormones.
Chemical classification of hormones

Group	Hormones	Place of synthesis
Derivatives of amino acids:	norepinephrine;	CNS, medullar part of the suprarenal glands.
Amines	epinephrine;	Medullar part of the suprarenal glands.
	dopamine;	The peripheral chromaffine tissue of the suprarenal gland
	melatonin.	Epiphysis
Iodtyronines:	thyroxin (T_4); triiodtyronine (T_3)	Thyroid gland
Hormones of protein-peptide nature:	Arginine-vasopressine (antidiuretic hormone)	Hypothalamus
	Oxytocine	Hypothalamus
Small peptides	Melanocite-stimulating hormone (MSH)	Hypothalamus
	Prolactin	Adenohypoplysis

Group	Hormones	Place of synthesis
Glycoproteins	Parathyrine (parat-hormone)	Parathyroid glands
	b-lypotropin and encephalin	Hypophysis, CNS
	Calcitonine	C-cells of the thyroid gland
	ACTH	Adenohypophysis
	Secretin	Digestive tract, CNS
	Cholecystokinin	Digestive tract, CNS
	Gastrine	Digestive tract, CNS
	Follicle-stimulating hormone	Adenohypophysis
	Luteinizing hormone (LH)	Adenohypophysis
	Thyrotrophin-releasing hormone	Hypothalamus
	Gonadotrophin-releasing hormone	Hypothalamus
	Corticotrophin- releasing hormone	Hypothalamus
Proteins	Angiotensin	Liver, blood, CNS
	Insulin	Pancreatic b-cells, CNS
	Glucagon	Pancreatic a-cells
	Somatotrophin (growth hormone)	Adenohypophysis
	Chorionic gonadotropin	Placenta
	Thyrotropin	Adenohypophysis
	Placental lactogen	Placenta
Steroids	Estrogens	Ovaries, placenta
	Progesterone	Ovaries, placenta
	Testosterone	Testis
	Glucocorticoids	Cortical layer of the suprarenal glands
	Aldosteron	Cortical layer of the suprarenal glands
	Vitamin D hormonal form	Kidneys

The 1st category, the amines, includes norepinephrine and epinephrine, which are derived from a single amino acid (*i.e.*, tyrosine), and the thyroid

hormones, which are derived from two iodinated tyrosine aminoacid residues. The 2nd category, the peptides, polypeptides, glycoproteins, and proteins, can be as small as thyrotropin-releasing hormone (TRH), which contains three aminoacids, and as large and complex as growth hormone (GH) and follicle-stimulating hormone (FSH), which have approximately 200 amino acids. Glycoproteins are large peptide hormones associated with a carbohydrate (*e.g.*, FSH). The 3rd category comprises the steroid hormones, which are derivatives of cholesterol.

4. Categories of hormones.

1) *Lipophilic*, e.g steroids, thyroxin, retinoic acid:

- small; diffuse through the membrane ;
- bind to cytosol or nuclear receptors;
- stability-hours to days;
- poorly soluble in water so they are transported in the blood by tightly bound carrier proteins.

For example, most of the thyroid and steroid hormones circulate bound to plasma binding proteins. The complex of hormone and carrier protein is restricted in its ability to leave blood vessels due to the large size of the complex, and thus represents an inactive reserve of hormone. Under certain conditions, the complex dissociates and the free hormone readily diffuses out of the blood to the target tissue (Fig. 9–2). The binding protein stabilizes the

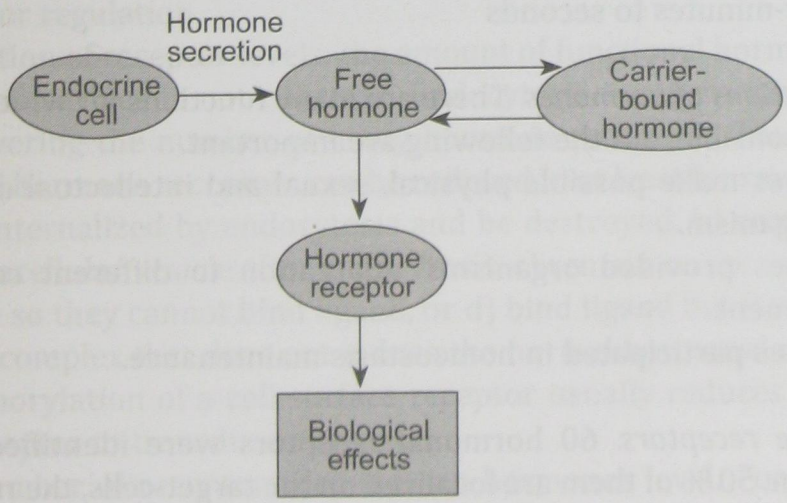


Fig. 9–2. Schematical representation of free and carrier-bound hormones.

hormone, since the bound hormone cannot diffuse out of the circulation, it is prevented from entering tissues (e.g., the liver) which degrade the hormone to inactive forms. Thyroid hormones are degraded during the time of several days and exert their effects over a similar period of time. In contrast, a protein hormone such as insulin, which does not associate with binding proteins, has the duration of approximately 5 min and exerts its maximal duration of approximately 5 minutes and exerts its maximal effect on blood glucose in approximately 30 minutes. A hormone that is relatively loosely bound will have a rapid onset and short duration of action. A hormone that is bound with high affinity will have a slow onset and long duration of action. In part, the difference in plasma protein binding between T_3 (Good binding) and T_4 (tight binding) accounts for the more rapid onset and shorter duration of T_3 actions than those of T_4.

The presence of some hormones in a bound state has major practical implications for the diagnosis of hormone deficiency or excess. Only the measurement of the free hormone concentration will give a reliable indication of function

2) *Hydrophilic*, most hormones and neurotransmitters, e.g., large polypeptides like insulin, growth hormone, glucagon; small charged compounds like epinephrine:

- can't diffuse across the plasma membrane to interact with intracellular receptors;
- bind to cell surface receptors;
- stability-minutes to seconds

5. *The functions of hormones.* There are many functions for which hormones are responsible, but the following are important.

- Hormones make possible physical, sexual and intellectual development of an organism.
- Hormones provided organisms' adaptation to different conditions of environment.
- Hormones participated in homeostasis maintenance.

6. *Hormone receptors.* 60 hormonal receptors were identified in human organism. 50 % of them are localized on the target-cells, the rest — inside the cells.

1) Specificity and hormone bidding.

Hormone receptors bind ligands reversibly but with great specificity and high affinity. This accounts for the ability of the target tissue to sense and respond to very small amounts of circulating hormones. This is important, because otherwise the secreting organ would need to produce very large quantities for rapidly metabolized hormones. Binding a hormone to a receptor involves the same types of weak interactions that characterize the specific binding of a substrate to an enzyme. The specificity of a receptor refers to its ability to distinguish closely related substances. KD, the dissociation constant of the receptor-ligand complex, measures the affinity of the receptor for the ligand.

A particular cell can express more than one type of receptor. A particular ligand may bind to more than one receptor. A particular receptor may occur on different cell types and binding may have different consequences.

Hormone analogs are important in the study of receptor action and in the manipulation of hormonal responses.

2) Detection of receptors

It is difficult to identify and purify hormone receptors, mainly because they are present in such minute amounts. Usually receptors are detected and measured by a functional assay: their ability to bind radioactive hormones to a cell or to cell fragments. Many cell-surface receptors can be solubilized from preparations of plasma membrane and fully retain their hormone-binding activity. Other receptors, that are expressed at such low levels that they cannot be purified by conventional approaches, have been cloned using recombinant DNA technology.

3) Receptor regulation

- Regulation of receptor levels: the amount of functional hormone receptor on a cell surface is not constant. Endocytosis is a principal mechanism for lowering the number of many hormone receptors. The level of cell-surface hormone receptors can be reduced in at least four ways. They can a) be internalized by endocytosis and be destroyed, b) remain stored in an intracellular vesicle after endocytosis, c) remain on the cell surface but change so they cannot bind ligand, or d) bind ligand but form a receptor-ligand complex that does not induce the normal hormonal response.
- Phosphorylation of a cell-surface receptor usually reduces the ability of the receptor to transduce a signal.
- The number of receptors depends on hormones level. For instance, the number of receptors decreases as the result of plasma hormones' level increasing and vice versa.

II. MECHANISMS OF ACTION

Hormones produce their effects through interaction with high-affinity receptors, which in turn are linked to one or more effector systems within the cell. These mechanisms involve many of the cell's metabolic activities, ranging from ion transport at the cell surface to stimulation of nuclear transcription of complex molecules. The rate at which hormones react depends on their mechanism of action. The neurotransmitters, which control the opening of ion channels, have a reaction time of milliseconds. Thyroid hormone, which functions in the control of cell metabolism and synthesis of intracellular signaling molecules, requires days for its full effect to occur.

Hormones exert their action by binding to high-affinity receptors located either on the surface or inside the target cells. The function of these receptors is to recognize a specific hormone and translate the hormonal signal into a cellular response. The structure of these receptors varies in a manner that allows target cells to respond to one hormone and not to others.

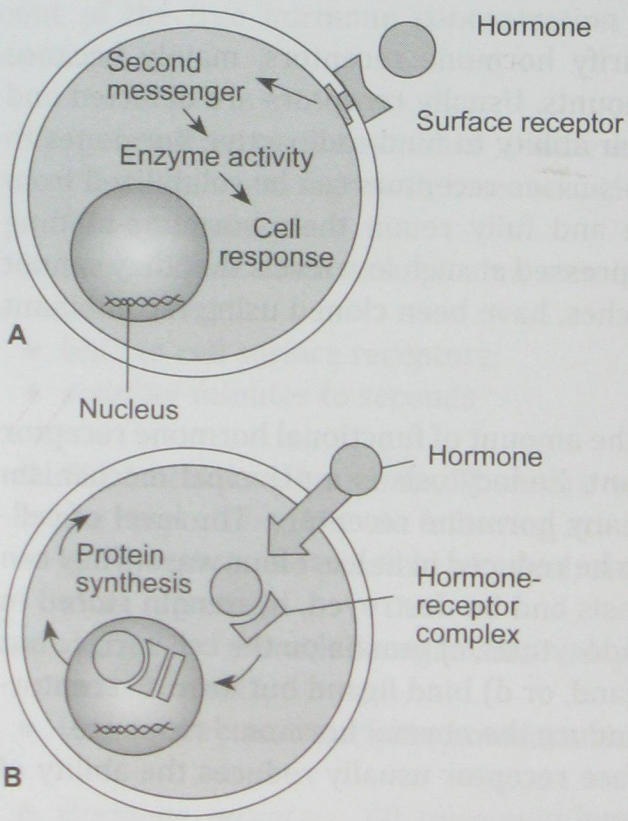


Fig. 9–3. Surface (A) and intracellular (B) receptors

1. *Cell surface receptors.* Cell surface receptors are generally bound to the plasma membrane so that part of the protein is on each side of the lipid bilayer. The hormone binds to the external portion of the receptor and does not penetrate the membrane. The resulting conformational change in the receptor results in the transduction of a signal across the membrane and activation of a "second messenger" system, and the hormone is considered to be the "first messenger" (Fig. 9–3).

Antibodies that bind to the receptor can mimic the effect of ligand binding. This phenomenon underlies some disease states. For example, the usual cause of hyperthyroid disease is the abnormal production of antibodies that bind to thyroid-stimulating-hormone receptors, activating the receptor and causing too much thyroid hormone to be produced

There are three major types of second messenger systems used by surface hormone receptors:

A. cAMP and G proteins;

One group of receptors is linked to GTP-binding regulatory proteins (G- proteins). The G-protein linked receptors usually activate a chain of events that alters the concentration of one or more small intracellular signaling molecules, often referred to as intracellular messengers. Two of the most important messengers are cyclic AMP (cAMP) and Ca^{2+}. cAMP acts as the second messenger in the following hormones effects: ACTH, TTH, FSH, glucagons, parathormone, gonadotrophins etc.

These intracellular messengers alter the activity of other target proteins in the cell.

For cAMP to function as an intracellular mediator, its intracellular concentration (normally £10^{-6}M) must be able to change rapidly up or down in response to extracellular signals: upon hormonal stimulation, cyclic AMP levels can change by fivefold in seconds. Such responsiveness requires that rapid synthesis of the molecule will be balanced by rapid breakdown or removal. Cyclic AMP is synthesized from ATP by the plasma-membrane-bound enzyme adenylatecyclase, and it is rapidly and continuously destroyed by one or more cyclic AMP phosphodiesterases. Theophyline and caffeine are inhibitors of the phosphodiesterase; how would you expect them to act with hormones that use cAMP as a second messenger? Receptors and adenylate cyclase molecules do not interact directly; GTP-binding regulatory proteins ("G proteins") are involved.

G_S (stimulatory G protein) transduces the signal from the receptor to adenylate cyclase,. When the receptor binds hormone, the G_s protein is activated and it binds a molecule of GTP. Carrying the GTP, it now activates the adenylate cyclase. Individuals who are genetically deficient in G_s have decreased responses to many hormones and, consequently, fail to grow or mature sexually, are mentally retarded, and have many metabolic abnormalities. Eventually, hydrolysis of the GTP to GDP by the G_s protein terminates the activation of the cyclase. Cholera toxin is an enzyme that catalyzes the alteration of a subunit of G_S so that it can no longer hydrolyze its bound GTP. The adenylate

cyclase will thus remain in the active state indefinitely. The resulting prolonged elevation in cAMP levels within intestinal epithelial cells causes a large efflux of Na^+ and water into the gut, which is responsible for the severe diarrhea that is characteristic of cholera.

Some receptors are coupled to adenylate cyclase by an inhibitory G protein (G_i). Pertussis toxin inactivates the GTPase activity of G_i, thus the G_i complex fails to inhibit adenylatecyclase in response to receptor activation.

B. Inositol phosphates, Ca^{2+}, and diacylglycerol;

Ca^{2+} acts as the second messenger for the following hormones action: oxytocin, vasopressin, gastrin, angiotensin etc.

Ca^{2+} functions as an intracellular messenger. The concentration of Ca^{2+} in the cytosol of any cell is extremely low ($\approx 10^{-7}$M), whereas its concentration in the extracellular fluid and in a specialized intracellular compartment is high ($>10^{-3}$M). Thus, there is a large gradient tending to drive Ca^{2+} into the cytosol across the plasma membrane and the membrane of the intracellular compartment when a signal transiently opens Ca^{2+} channels in either of these membranes. In nerve cells, depolarization of the plasma membrane causes an influx of Ca^{2+} through the voltage-gated Ca^{2+} channels. In almost all eukaryotic cells, the engagement of cell-surface receptors causes the release of Ca^{2+} from the intracellular calcium-sequestering compartment through Ca^{2+} channels in the internal membrane. This release of calcium involves another intracellular messenger molecule, inositoltrisphosphate. More than 25 different cell-surface receptors have been shown to utilize the inositol-phospholipid transduction pathway.

An activated receptor transduces a signal, via a G protein, to the plasma-membrane enzyme phospholipase C. In less than a second, this enzyme hydrolyzes *phosphatidylinositol-bisphosphate* and cleaves it into two products: *inositol trisphosphate* and *diacylglycerol.*

Inositol-trisphosphate is a small water-soluble molecule that diffuses to the ER where it binds to a specific receptor and causes the release of Ca^{2+} from intracellular compartments into the cytosol. *Calmodulin* is a small protein, ubiquitous in eukaryotic cells, that forms a complex with Ca^{2+} and mediates most Ca^{2+}-regulated processes. Each calmodulin binds four Ca^{2+} ions, and binding each Ca^{2+} ion facilitates binding additional Ca^{2+}, thus a small change in the level of cytosolic Ca^{2+} causes a large change in the level of active calmodulin. A typical animal cell contains more than 10^7 molecules of calmodulin, which can constitute as much as 1% of the protein mass of the cell.

Diacylglycerol (DAG) remains in the plasma membrane and activates a Ca^{2+}-dependent enzyme called protein kinase C. The activation of PKC is transient

because within seconds DAG is inactivated. PKC is a key regulatory enzyme; it appears to be involved in synaptic transmission, neuronal development, axonal regeneration, growth, differentiation, learning, and tumor promotion.

C. Tyrosine kinases.

Some receptors, termed catalytic receptors, act like enzymes when they are activated by their ligand. Almost all of the known catalytic receptors are transmembrane proteins that function as tyrosine-specific protein kinases, as opposed to other kinases that act on serine and threonine. This class of receptors includes the insulin receptor and many growth factor receptors. Abnormal function or regulation of insulin receptors has been demonstrated in certain cases of diabetes. Ligand binding induces a conformational change in the receptor. This activates the catalytic domain on the cytoplasmatic side of the plasma membrane, in this way stimulating the phosphorylation of selected proteins in the target cell in addition to "autophosphorylation", i.e. phosphorylation of the receptor itself. Following ligand binding, many catalytic receptors are endocytosed.

Some oncogenes encode abnormal receptors that express constitutive kinase activity. Cells carrying such faulty receptors behave as though they are constantly being signaled to proliferate.

2. *Steroid hormone receptors.*

Steroid hormones (i.e., corticosterone, cortisol, aldosterone, testosterone, progestin's, estriol, estradiol, estrone, steroids, vitamin D_3, etc.) are synthesized from cholesterol. They are small molecules and cross the membrane by simple diffusion. Once inside the target cell, each type of steroid hormone binds tightly but reversibly to a specific receptor. These receptors typically comprise less than 0.01% of the total mass of protein in the cell. Experiments with antibody to the glucocorticoid receptor show that before hormone treatment the receptors are found mainly in the nucleus. The binding of the hormone causes the receptor to undergo a conformational change ('receptor activation') that increases its affinity for DNA and enables it to bind to specific nucleotide sequences in the genes that regulates and influence their transcription.

The important functional domains of these receptors are:

- carboxyl-terminal domain-hormone-binding region;
- middle domain — DNA-binding region;
- amino-terminal domain-activates (or suppresses) gene transcription.

How does the binding of the receptor-hormone complexes to a gene activate its transcription?

The characteristic DNA-binding sites for all the major steroid hormone receptors are known. These sites, or hormone response elements (HRE), are distinct for different hormones. DNA recognition sites associated with steroid-hormone-responsive genes function as receptor-dependent transcriptional enhancers, which can stimulate transcription even when they are located thousands of nucleotides away from the promoter where RNA synthesis begins. Thyroid hormones work in the same way. The responses to steroid hormones are different in different target cells. More than one type of gene-regulatory protein generally must bind to a eukaryotic gene in order to activate its transcription. A steroid hormone receptor can activate a gene only if the right combination of other regulatory proteins is also present. And some of these are cell-type specific.

Role of steroids in brain function: using DNA technology, steroid receptors have been found to be expressed in regions of the brain that exhibit no steroid responsiveness. This is thought to indicate a possible role for steroid hormones as neurotransmitters and is an area of active investigation.

Testicular feminization syndrome — the gene for the testosterone receptor is defective. Understanding the molecular basis of this syndrome is of practical importance. Would the administration of exogenous testosterone be expected to have a therapeutic effect?

Estrogen receptors and breast and prostate tumors; the best prognostic indicator is whether the cells are estrogen receptor (ER) positive or estrogen receptor negative. ER+ cells are dependent on estrogen for hypophysectomy or adrenalectomy. The current treatment of choice is the use of estrogen antagonists, such as tamoxifen, which binds to the receptor and blocks uptake of estrogen.

In summary, responses follow the binding of a hormone to a specific receptor. Changes in protein synthesis may result from relatively direct effects on gene transcription by steroid or thyroid hormones. Some hormones have less direct effects on protein synthesis, presumably mediated by protein kinases (e.g., cAMP-dependent protein kinase, protein kinase C) acting on transcription factors.

III. REGULATION OF HORMONES' SYNTHESIS

Hormone secretion varies widely during a 24-hour period. Some hormones, such as GH and ACTH, have diurnal fluctuations that vary with the sleep-wake

cycle. Others, such as the female sex hormones, are secreted in a complicated cyclic manner. The levels of hormones such as insulin and ADH are regulated by feedback mechanisms that monitor substances such as glucose (insulin) and water (ADH) in the body. The levels of many of the hormones are regulated by feedback mechanisms that involve the hypothalamic-pituitary-target cell system.

There are at least three ways of hormone synthesis and secretion regulation: neural, hormonal and intracellular.

1. *Neural regulation* occurs by four main directions:

Hypothalamic-pituitary system (Fig. 9–4). 6 liberins and 3 statins are produced within the hypothalamus which entering the adenohypophysis enhances (liberins) or suppresses (statins) hormones production. The hypothalamic

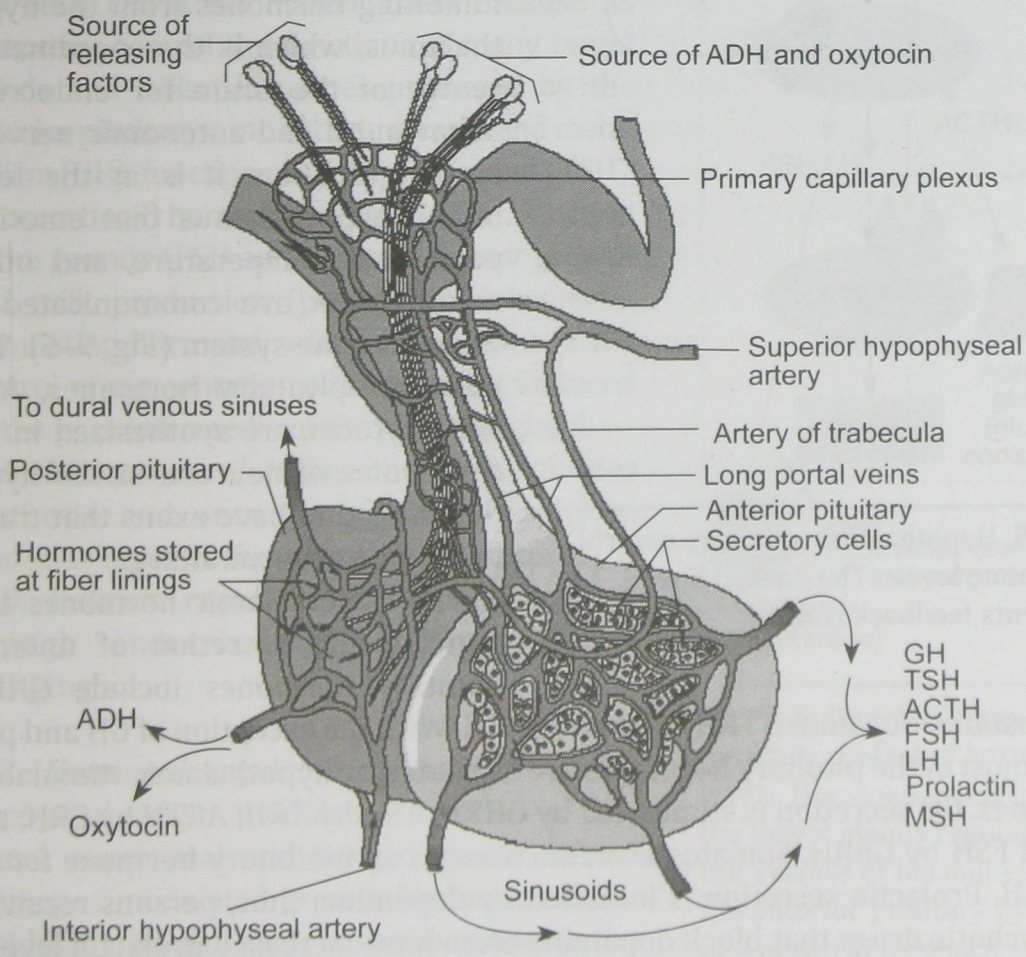


Fig. 9–4. Hypothalamus and the anterior and posterior parts of the pituitary gland.

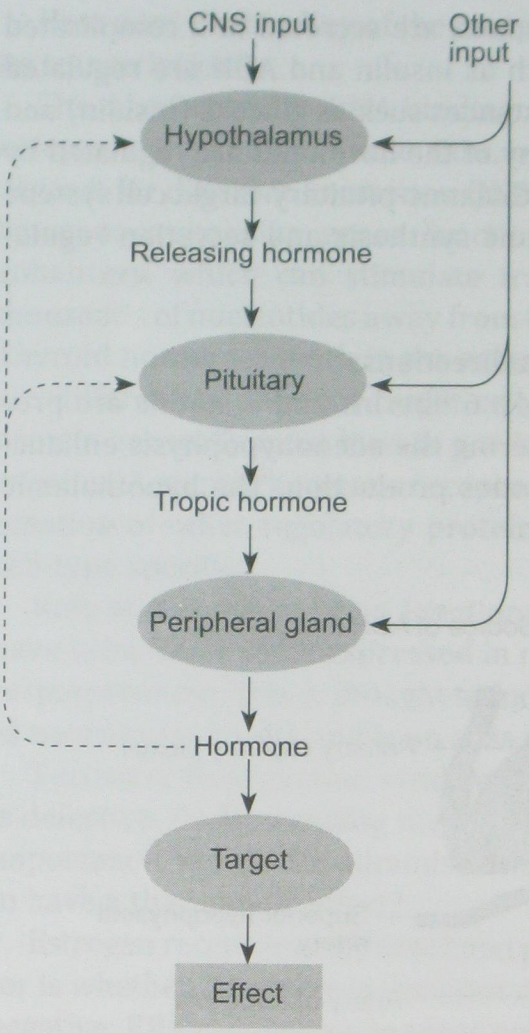


Fig. 9–5. Hypothalamic-pituitary control of hormone levels. The dashed line represents feedback control.

releasing or inhibiting hormones are transported to the anterior pituitary by way of the portal vessels. ADH and oxytocin are produced by nerve cells in the supraoptic and paraventricular nuclei of the hypothalamus and then transported through the nerve axon to the posterior pituitary, where they are released into the circulation.

A.Hypothalamic hormones. The synthesis and release of anterior pituitary hormones are largely regulated by the action of releasing or inhibiting hormones from the hypothalamus, which is the coordinating centre of the brain for endocrine, behavioural and autonomic nervous system function. It is at the level of the hypothalamus that emotion, pain, body temperature, and other neural input are communicated to the endocrine system (Fig. 9–5). The posterior pituitary hormones, ADH and oxytocin, are synthesized in the cell bodies of neurons in the hypothalamus that have axons that travel to the posterior pituitary.

The hypothalamic hormones that regulate the secretion of anterior pituitary hormones include GHRH, somatostatin, dopamine, TRH, CRH and GnRH. With the exception of GH and prolactin, most of the pituitary hormones are regulated by hypothalamic stimulatory hormones. GH secretion is stimulated by GHRH; TSH by TRH; ACTH by CRH; and LH and FSH by GnRH. Somatostatin functions as an inhibitory hormone for GH and TSH. Prolactin secretion is inhibited by dopamine; thus, persons receiving antipsychotic drugs that block dopamine often have increased prolactin levels.

The activity of the hypothalamus is regulated by both hormonally mediated signals (*e.g.*, negative feedback signals) and by neuronal input from a num-

ber of sources. Neuronal signals are mediated by neurotransmitters such as acetylcholine, dopamine, norepinephrine, serotonin, GABA and opioids. Cytokines that are involved in immune and inflammatory responses, such as the interleukins, also are involved in the regulation of hypothalamic function. This is particularly true of the hormones involved in the hypothalamic-pituitary-adrenal axis. Thus, the hypothalamus can be viewed as a bridge by which signals from multiple systems are relayed to the pituitary gland.

B. Pituitary Hormones. The pituitary gland has been called the master gland because its hormones control the functions of many target glands and cells. Hormones produced by the anterior pituitary control body growth and metabolism (GH), function of the thyroid gland (TSH), glucocorticoid hormone levels (ACTH), function of the gonads (FSH and LH), and breast growth and milk production (prolactin). Melanocyte-stimulating hormone, which is involved in the control of pigmentation of the skin, is produced by the pars intermedia of the pituitary gland.

C. Direct neural influences on hormones synthesis and secretion (neurohypophysis, medullar layer of the suprarenal glands).

D. Indirect regulation by neural system through endocrinal gland blood supply changing.

2. *Hormonal regulation.* Melatonin of the epiphyseal origin modulates function of the sexual, suprarenal and thyroid glands. Sexual hormones modulated growth hormone effects. Besides, the level of many of the hormones in the body is regulated by negative feedback mechanisms (Fig. 9–6) that will be described below.

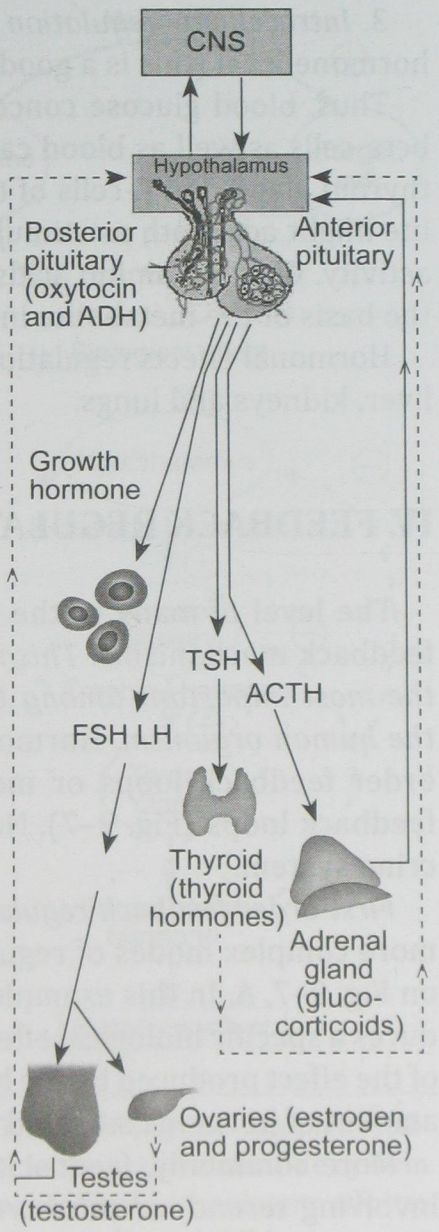


Fig. 9–6. Hypothalamic-pituitary control of hormone levels. Hormone levels from the target glands regulate the release of hormones from the anterior pituitary by means of a negative feedback system. The dashed line represents feedback control.

3. *Intracellular regulation* occurs by the compound direct influence on the hormone level (this is a good example of the negative feedback mechanism).

Thus, blood glucose concentration regulates insulin exit out of pancreatic beta-cells as well as blood calcium level defines secretory activity of the parathyroid gland and C-cells of the thyroid glands. Therefore, regulative metabolite might acts both as stimulator and inhibitor of endocrinal gland functional activity. Glucose, amino acids, Ca^{2+}, Na^{+}, K^{+} etc levels self regulation occurs on the basis of the metabolite interaction with endocrinal gland cells.

Hormonal effects regulation occurs also due to hormones degradation in the liver, kidneys and lungs.

IV. FEEDBACK REGULATION

The level of many of the hormones in the body is regulated by negative feedback mechanisms. *This regulative mechanisms is considered to be one of the most important among the principles of the functions regulation within the human organism.* Hormone secretion may be regulated via simple first-order feedback loops or more complex multilevel second- or third-order feedback loops (Fig. 9–7). Negative feedback is most prevalent in the endocrine system.

First-order feedback regulation is the simplest type and forms the basis for more complex modes of regulation. A simple first-order feedback loop shown on Fig. 9–7, A. In this example, an endocrine cell secretes a hormone that produces a specific biological effect in its target tissue. It also senses the magnitude of the effect produced by the hormone. As the biological response increases, the amount of hormone secreted by the endocrine cell is appropriately decreased.

More commonly, feedback regulation in the endocrine system is complex, involving *second- or third-order feedback loops.* For example, multiple levels of feedback regulation may be involved in regulating hormone production by various endocrine glands under the control of the anterior pituitary (Fig. 9–7, B). The regulation of target gland hormone secretion, such as adrenal steroids or thyroid hormones, begins with production of a releasing hormone by the hypothalamus. The releasing hormone stimulates production of a trophic hormone by the anterior pituitary, which, in turn, stimulates the production of the target gland hormone by the target gland. As indicated by the dashed lines in Fig. 9–7, B, the target gland hormone may have negative-feedback effects to inhibit secretion of both the trophic hormone from the anterior pituitary and

the releasing hormone from the hypothalamus. In addition, the trophic hormone may inhibit releasing hormone secretion from the hypothalamus, and in some cases, the releasing hormone may inhibit its own secretion by the hypothalamus.

The more complex multilevel form of regulation appears to provide certain advantages compared with the simpler system. Theoretically, it permits a greater degree of fine-tuning of hormone secretion, and the multiplicity of regulatory steps minimizes changes in hormone secretion in the event that one component of the system is not functioning normally.

Both types of feedback (negative and positive) control occur because the endocrine cell, in addition to synthesizing and secreting its own hormone product, has the ability to sense the biological consequences of secretion of that hormone. This enables the endocrine cell to adjust its rate of hormone secretion to produce the desired level of effect, ensuring the maintenance of homeostasis.

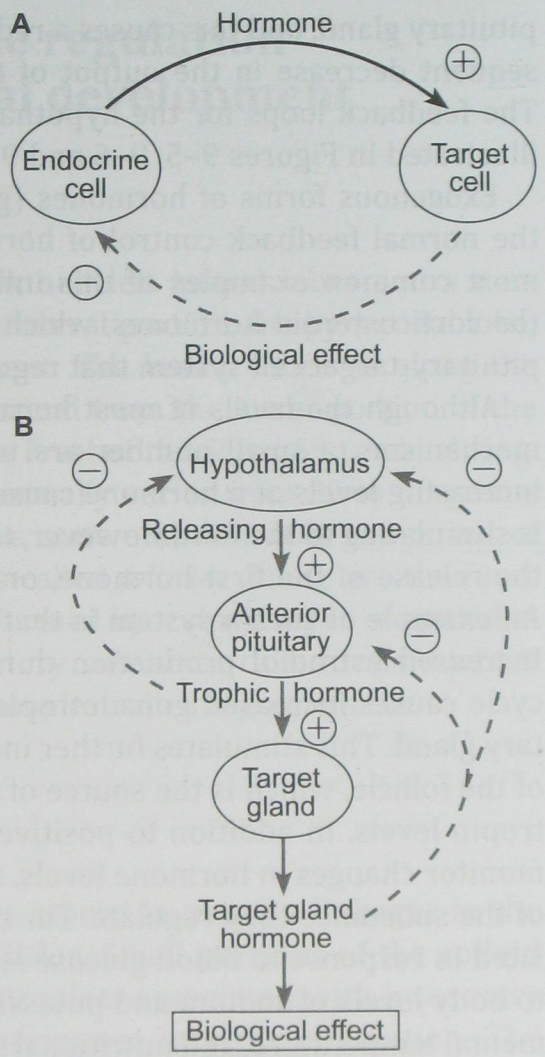


Fig. 9–7. Simple (A) and complex feedback (B) loops in the endocrine system. Solid lines indicate stimulatory effects; dashed lines indicate inhibitory, negative-feedback effects.

The function of this type of system is similar to that of the thermostat in the heating system. In the endocrine system, sensors detect a change in the hormone level and adjust hormone secretion so that body levels are maintained within an appropriate range. When the sensors detect a decrease in hormone levels, they initiate changes that cause an increase in hormone production; when hormone levels rise above the set point of the system, the sensors cause hormone production and release to decrease. For example, an increase in thyroid hormone is detected by sensors in the hypothalamus or anterior

pituitary gland, and this causes a reduction in the secretion of TSH, with a subsequent decrease in the output of thyroid hormone from the thyroid gland. The feedback loops for the hypothalamic-pituitary feedback mechanisms are illustrated in Figures 9–5, 9–6 and 9–7.

Exogenous forms of hormones (given as drug preparations) can influence the normal feedback control of hormone production and release. One of the most common examples of this influence occurs with the administration of the corticosteroid hormones, which causes suppression of the hypothalamic-pituitary-target cell system that regulates the production of these hormones.

Although the levels of most hormones are regulated by negative feedback mechanisms, a small number are under positive feedback control, in which increasing levels of a hormone cause another gland to release a hormone that is stimulating to the first. However, there must be a mechanism for shutting off the release of the first hormone, or its production would continue unabated. An example of such a system is that of the female ovarian hormone estradiol. Increased estradiol production during the follicular stage of the menstrual cycle causes increased gonadotropin (FSH) production by the anterior pituitary gland. This stimulates further increases in estradiol levels until the demise of the follicle, which is the source of estradiol, results in a decrease in gonadotropin levels. In addition to positive and negative feedback mechanisms that monitor changes in hormone levels, some hormones are regulated by the level of the substance they regulate. For example, insulin levels normally are regulated in response to blood glucose levels and those of aldosterone in response to body levels of sodium and potassium. Other factors such as stress, environmental temperature, and nutritional status can alter feedback regulation of hormone levels.

Chapter 10. Endocrine regulation of physical and mental development

The thyroid gland is composed of right and left lateral lobes and a narrow connecting portion, the isthmus. A unilateral portion, the pyramidal lobe, extending from the isthmus may be present. The lateral lobes lie along the anterolateral surface of the trachea and extend from the lower portion of the thyroid cartilage of the larynx down to the level of the 5th or 6th tracheal ring. The thyroid gland weight in the human normally varies from 20 to 40 grams but is influenced by age, sex, reproductive state and diet. It is composed of a large number of tiny, sac-like structures called *follicles*. These are the functional units of the thyroid. Each follicle is formed by a single layer of epithelial (follicular) cells and is filled with a secretory substance called *colloid* which consists largely of a glycoprotein-iodine complex called *thyroglobulin*.

I. THE THYROID HORMONES

The synthesized thyroid hormones are stored in peptide linkages in the thyroglobulin (colloid) of the thyroid follicles. Small portions of the colloid are taken into the thyroid follicle cell as droplets associated with lysosomes and hydrolyzed to release the thyroid hormones into the circulation. The thyroglobulin is degraded, and its component parts are reused. Thyroglobulin is not released normally into the circulation, if it gains access to the circulation, it behaves as an antigenic foreign protein with pathological consequences. There are two principal thyroid hormones which can be isolated from the plasma, are of physiologic importance, and are the topic of this section.

1.1. Anterior Pituitary — Thyroid Axis

TSH (thyroid stimulating hormone, thyrotropin)

The secretion of thyroid hormone is regulated by the hypothalamic-pituitary-thyroid feedback system (Fig. 10–3). In this system, thyrotropin-relea-sing hor-

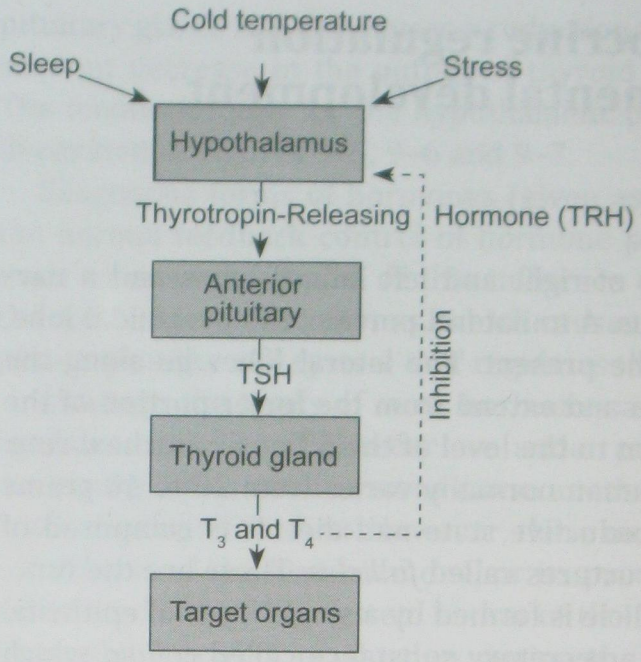


Fig. 10–3. Scheme of hypothalamic-pituitary-thyroid feedback system that regulates the body levels of thyroid hormone secretion.

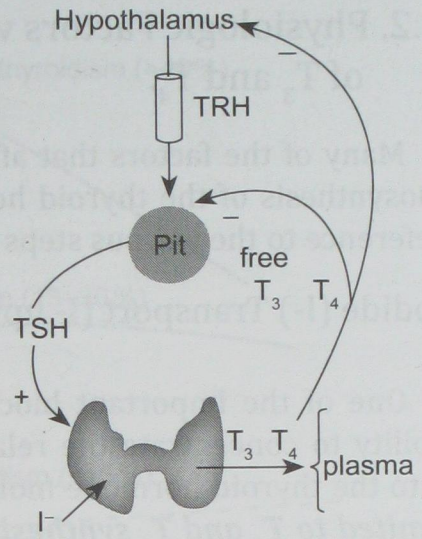
mone (TRH), which is produced by the hypothalamus, controls the release of thyroidstimulating hormone (TSH) from the anterior pituitary gland. TSH is a glycoprotein composed of an α and a β chain. The α chain is closely similar in sequence to the a chains of FSH, LH and hCG. The β chain aminoacid sequence is different for each of the aforementioned glycoprotein hormones and confers the biologic activity and specificity to each of these hormones.

a) **Actions of TSH**

Thyrotropin stimulates the thyroid follicular cells *to increase the synthesis and release of the thyroid hormones;* these are two separate and distinct actions of TSH. TSH has the following specific actions:

- *mode of action.* The mechanism of TSH activation of increased thyroid gland activity is via combination with the TSH receptor and rapid subsequent generation of intracellular (second) messengers. The intracellular concentrations of the following messengers are increased in the follicular cells: cAMP, DAG, IP_3 and Ca^{2+}-Calmodulin.
- *rapid increase in follicle cell metabolism.* There are increases in O_2 used, CO_2 production, glucose uptake and utilization, etc. There is also a marked increase in blood flow that occurs hours to days after TSH administration.
- *rapid increase in the secretion of stored thyroid hormones.* This effect can be detected within minutes. The mechanism is via activation of the cyclase-cAMP system which in turn activates a proteolytic enzyme which hydrolyzes the peptide bonds of the colloid to release free T_3 and T_4. This process occurs in the colloid droplets which associate with lysosomes and migrate from the follicular border to the basal border where the free T_3 and T_4 diffuse into the capillaries. As a consequence of the increased se-

cretion there is less colloid, and the *total thyroidal iodine* content is decreased.



b) **Control of TSH Secretion**

- *Stimulation:*

Cold environmental temperature, acting via peripheral neural input or central core temperature will cause an increased TSH release, presumably acting through the hypothalamus. This response (increased TRH, TSH, T_3 and T_4) is readily demonstrated in laboratory animals as a chronic adjustment to cold. Similarly, the neonatal human will respond to cold exposure with an increase in TSH. However, with age, there is a decreased sensitivity, and this reflex is not easily demonstrated in the adult human and requires total exposure (nakedness) for several weeks.

- Inhibition:

The plasma levels of *free* T_3 and T_4 exert an inhibitory effect on TSH release from the pituitary; the primary site of action is directly on the thyrotrophs of the pituitary. There is evidence that T_3 is the intracellular hormone responsible for the inhibition; free T_4 is converted by the pituitary cells to T_3 which then inhibits TSH release, even when TRH is present. The plasma levels of free T_3 and T_4 also inhibit the hypothalamic synthesis of TRH. It has been demonstrated that thyroidectomy will increase, whereas administered T_4 will decrease the mRNA for ProTRH in hypothalamic cells. Thus, the negative feedback control occurs at hypothalamic and pituitary sites.

c) **Summary Diagram:**

The right side diagram summarizes and depicts the controls which have been discussed above. It emphasize; that it is the plasma free T_3 and T_4 which exert the *negative feedback control directly on the anterior pituitary cells and hypothalamus.* This inhibition of TSH is accomplished by first increasing protein synthesis in the anterior pituitary, a finding consistent with the view that many T_3 and T_4 actions are mediated via an effect on protein synthesis. There is evidence to indicate that the intracellular inhibition of TSH release is due to T_3 only and that T_4 is effective because the pituitary, like many other tissues, can convert T_4 to T_3; this is consistent with the view that T_3 is the active intracellular hormone. The thyroid hormones influence the hypothalamus by decreasing the mRNA for proTRH in hypothalamic cells.

1.2. Physiologic Factors which Influence the Biosynthesis of T_3 and T_4.

Many of the factors that affect thyroidal function do so by influencing the biosynthesis of the thyroid hormones. These factors are discussed best with reference to the various steps of the biosynthetic process.

Iodide (I-) Transport (I- uptake)

One of the important biochemical characteristics of thyroid tissue is its ability to concentrate the relatively rare element, iodine, and incorporate it into the thyroid hormone molecule. ***The initial step, I^- transport, is the rate limited to T_3 and T_4 synthesis.*** Thus, a measure of I^- transport is an index of the T_3 and T_4 synthesis. It is an active process and is a carrier mediated one. The I^- is concentrated against a chemical and electrical gradients, and the process displays all the characteristics of active transport. It is located at the basal region of the cell which is well supplied with mitochondria and the required ATPase.

The major controls of the I^- transport mechanism are: (a) *TSH* and (b) *autoregulation*. TSH stimulates the mechanism to increase I^- transport after a 30 minute latency. TSH is the principal controler. TSH increases the T/S [I^-] ratio, and by accelerating the rate-limiting step, it increases T_3 and T_4 synthesis. The auto-regulatory control is determined by the thyroidal iodine storage. I^- transport is increased when the iodine store is decreased; I^- transport is decreased when the iodine store is increased. Autoregulation is independent of TSH, but it will increase or decrease the response to TSH. It provides the moment to moment corrections for an increase or decrease in I^- ingestion. Thus, in a person with a low I^- intake, the activity of the transport mechanism is greatly increased, and with a high I^- intake, its activity is greatly decreased. The mechanism provides a degree of protection against a deficiency or excess. It is independent of the slower, chronic adjustments made by altered TSH release.

Clinically, one can not use the T/S [I^-] ratio evaluation, but one can use the non-invasive technique of measuring the total radioactive I^- uptake by the thyroid gland for the assessment of T_3 and T_4 synthesis. A tracer dose (2μC) of $Na^{131}I$ is given orally, and at timed intervals the radioactivity at the neck region over the thyroid gland is measured. What are the two compelling reasons for the use of a tracer dose? The highly idealized graph below depicts the thyroidal uptake of ^{131}I in hypo-, hyper- and euthyroid subjects. Note that better discrim-

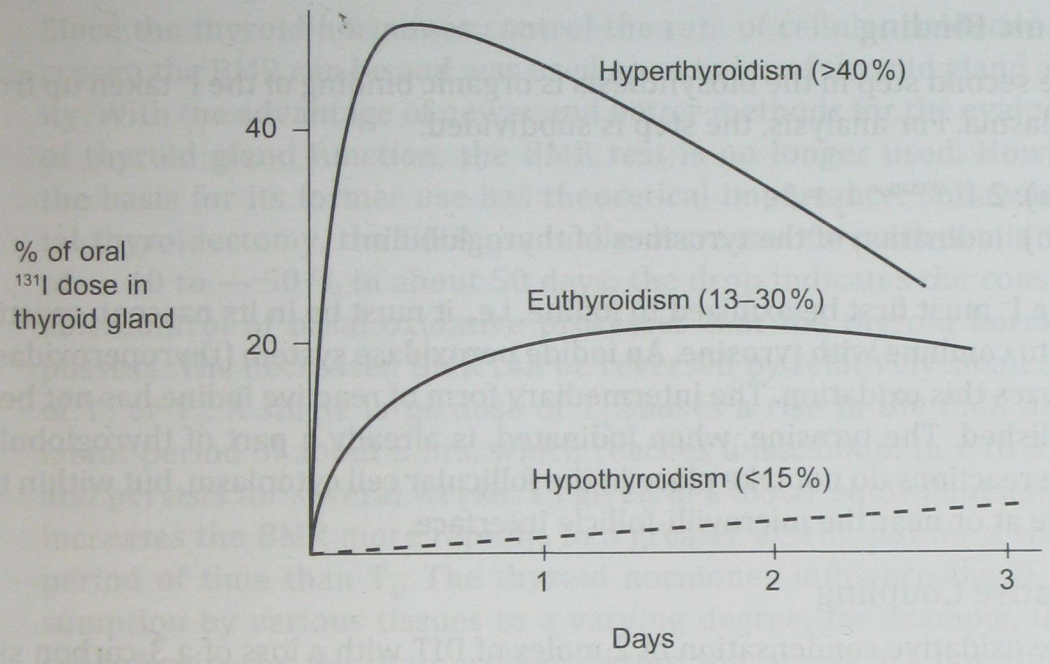


Fig. 10–4. Highly idealized graphs which depict thyroidal uptake Of ^{131}I in hypothyroidism, euthyroidism and hyperthyroidism (the 24-hr uptakes are shown in parentheses).

ination is obtained at earlier times (2, 6 or 12 hrs), although the 24 hr test was commonly used for the patient's convenience (return to lab the next day rather than wait for hours).

The radioiodide uptake test is no longer commonly used clinically because of the development of reliable and safe methods for the measurement of circulating thyroid hormone concentrations and indices of those plasma concentrations (Fig. 10–4).

Knowledge of I^- handling by the thyroid is of fundamental importance because it reveals the kinetics of thyroid function and aids our understanding of thyroidal physiology and pathophysiology. The graph of the euthyroid subject reflects the rapid uptake of plasma radioiodide by thyroid follicular cells following ingestion of radioiodide, the rapid incorporation of the radioiodide into the thyroid hormone molecule, storage of the radioactive hormone, and the subsequent slow release rate (secretion) of the radiohormone. Note that the hyperthyroid gland takes the I^- up at a greater rate, incorporates a greater total quantity into T_3 and T_4 and secretes the T_3 and T_4 at a greater rate than those of the normal gland.

Organic Binding

Tie second step in the biosynthesis is organic binding of the I^- taken up from the plasma. For analysis, the step is subdivided:

(2a): $2\ I^- \xrightarrow{\text{oxidation}} I_2 + 2e$
(2b): iodination of the tyrosines of thyroglobulin

The I^- must first be oxidized to iodine, i.e., it must be in its nascent, reactive form to combine with tyrosine. An iodide peroxidase system (thyroperoxidase) catalyzes this oxidation. The intermediary form of reactive iodine has not been established. The tyrosine, when iodinated, is already a part of thyroglobulin. These reactions do not take place in the follicular cell cytoplasm, but within the follicle at or near the microvilli-follicle interface.

Ovidative Coupling

The oxidative condensation of 2 moles of DIT with a loss of a 3-carbon side chain under the influence of a coupling enzyme results in the formation of T_4. The reaction does not proceed under anaerobic conditions. T_3 may arise from the coupling of a mole of DIT with a mole of MIT.

1.3. Actions of T_3 and T_4

1) *increase basal 0_2 consumption, therefore increase basal heat production (calorigenic effect).* Basal metabolism is the heat production of an individual who is in a post-absorptive state, rested, awake, calm and at rested (no motion). The basal metabolic rate (BMR) can be expressed as the basal head production in kcal/hr/M^2, the normal value for the young adult female is 36 kcal/hr/M^2, and for the young adult male it is ~ 40 kcal/hr/M^2 of body surface area. Clinically, it is expressed as a ± % deviation from the predicted normal value for that subject. See equation given below. A ± 15 % deviation from the predicted value covers more than 90 % of the population and is taken as the normal range. Values below this range indicate hypothyroidism, and values above this range indicate hyperthyroidism. The BMR is no longer used clinically.

$$\frac{\text{measured BMR-predicted BMR}}{\text{predicted BMR}} \times 100 = \pm\ \%\ \text{deviation} = \text{BMR}$$

Since the thyroid hormones control the rate of cellular oxidative processes, the BMR can be and was used as an index of thyroid gland activity. With the advantage of newer and better methods for the evaluation of thyroid gland function, the BMR test is no longer used. However, the basis for its former use has theoretical importance. Following total thyroidectomy, the BMR gradually decreases to reach a minimum of — 40 to — 50 % in about 50 days; the drop indicates the considerable control of basal oxidative processes that the thyroid hormones possess. The decreased BMR can be reversed by relatively small doses of T_4 or T_3. A single large dose of T_4 causes a rise in the BMR after a latent period of about 8 hrs, which reaches a maximum in 2 to 3 days and persists for several weeks. T_3, given in a molar equivalent amount, increases the BMR more rapidly, to a greater extent, but for a shorter period of time than T_4. The thyroid hormones influence the O_2 consumption by various tissues to a varying degree; for example, the O_2 consumption by the brain, spleen, testes and others are not affected, though its consumption by the heart and gastric mucosa is greatly increased, while the O_2 use by most other tissues is increased to an intermediate extent.

If protein synthesis is blocked, the T_3 and T_4 effect on O_2 use and calorigenesis is blocked. This indicates that the mode of action of T_3 and T_4 is via a stimulation of protein synthesis. T_3, after diffusing through the cytoplasm region, binds with a nuclear receptor; this is followed shortly by an increase in RNA polymerase, increased RNA and finally an increase in protein synthesis. These events occur rapidly enough to account for all other sequelae. The thyroid hormones may influence O_2 use by their effects on mitochondria: increased number of cristae, slight swelling to influence permeability and steric relationships and an increase in mitochondrial protein synthesis. A large fraction of the T_3 and T_4 effect on O_2 consumption can be attributed to their increase of Na^+ pumping. If the Na^+ pump is blocked, basal thermogenesis is decreased by about 50 to 65 %. T_4 increases the activity of the Na^+/K^- ATPase, decreases the Na^+/K^+ intracellular ratio and decreases the ATP/ADP ratio. Thus, a large fraction of basal O_2 consumption is used to maintain the Na^+ pump activity, and the latter is controlled by thyroid hormones. Though the problem is unsettled, an effect of T_3 and T_4 on protein synthesis to synthesize specific proteins which then initiate the specific actions may be the principal mode of action.

2) *are required for developmental growth.* The thyroid hormones are anabolic hormones, and the above discussed stimulatory effects on protein synthesis provide one means by which this effect could take place. The need for T_3 and T_4 for developmental growth is dramatically exemplified by the cretinism, a dwarf whose thyroid hypofunction had its inception *in utero* and in the early neonatal period. The cretinized person is characterized by a small stature, small bones, delay in development of centers of ossification, delay or failure of epiphyseal closure, poorly formed teeth, mental retardation and a "pot belly" as a result of weak muscles. The reproductive system may remain infantile. These multiple developmental growth defects can be reversed if thyroid hormone replacement is begun soon enough. There is apparently a critical period, extending from *in utero* through the neonatal period and in which thyroid hormone must be present for developmental growth and "maturation" of the CNS; if thyroid hormone is replaced after this period, it is considerably less effective, particularly on the CNS maturation process.

3) *influence upon carbohydrate metabolism.* T_3 and T_4 have multiple simulative effects on metabolism. Many of the effects are dosage dependent; to generalize, low to moderate doses are anabolic, while higher doses tend to be catabolic. For simplicity, these effects are listed below:

- thyroid hormones increase the gut absorption of glucose, probably via general stimulation of metabolic activity. Plasma glucose will tend to be higher after a meal if T_3 and T_4 levels are elevated and this can lead to glucose intolerance;
- thyroid hormones enhance the effects of insulin on liver glycogen synthesis and glucose utilization;
- thyroid hormones when present in excess enhance the adenylcyclase response to epinephrine. Thus, the epinephrine-induced glycogenolysis and hyperglycemia are potentiated by thyroid hormones;
- large doses of T_3 and T_4 enhance glycogenolysis and promote gluconeogenesis by making precursors available. All this effects probably result from the overall increased levels of cellular metabolic enzymes caused by T_3 and T_4 elevations .

4) *influence lipid metabolism.* T_3 and T_4 stimulate all aspects of lipid metabolism, but the catabolic pathways appear to be stimulated to a greater extent. Some effects are listed below:

- T_3 and T_4 decrease the plasma concentration of cholesterol. Cholesterol synthesis and its removal from the circulation are increased (excretion

and synthesis of compounds using cholesterol as their precursor). The thyroid hormones increase the formation of LDL receptors; this action contributes to the decrease in plasma cholesterol concentration;

- T_3 and T_4 enhance lipolysis by stimulating cyclase of adipose cells and by enhancing the cyclase response to other hormones (epinephrine, glucagon).

5) *enhance responses to catecholamine.* Many of the actions of excess T_3 and T_4 (hyperthyroidism) are similar to those of E, NE and β-adrenergic activation:

- Many of the effects of hyperthyroidism (increased heart rate, sweating, calorigenesis, etc.) can be diminished by β-adrenergic blocking agents or sympathectomy. It is thought that T_3 and T_4 enhance or sensitize the cyclase to catecholamine;
- Epinephrine and norepinephrine have an immediate calorigenic effect and increase the metabolic rate; this effect is abolished if the thyroid is removed;
- The increased irritability and "nervousness" of hyper-thyroidism may be related to the enhancement of catecholamine effects on CNS function.

6) *Effects of T_3 and T_4 on the cardiovascular system.* The thyroid hormones has direct effect on the excitability of the heart by increasing the heart rate T_3 and T_4 are strengtheners of both heart beat and blood volume .

7) *Effect on the CNS.* The thyroid hormones increased the rapidity of cerebration. They are essential for CNS development. The hyperthyroid individual is likely to have extreme nervousness and many psychoneurotic tendencies (anxiety, extreme worry).

8) *Other effects of T_3 and T_4*

One of the most characteristic signs of hyperthyroidism is a fine muscle tremor (with frequency of 10 to 15 times per second), which caused by increased reactivity of the neuronal synapses in the areas of the cord that control muscle tone. Increased thyroid hormones increases the rates of secretion of most endocrine glands, but it also increases the need of the tissues for the hormones.

1.4. Some General Characteristics of T_3 and T_4 Excess and Deficiency

An alteration in thyroid function can represent a hypofunctional or a hyperfunctional state. Disorders of the thyroid may be caused by a congenital defect in thyroid development, or they may develop later in life, with a gradual or sudden onset.

Excess of T_3 and T_4:

- increased excitability of nervous system (irritability, restlessness, fine hand tremors, sweating, diarrhea, etc.)
- increased basal O_2 consumption (increased heat production, sensitivity to hot environment)
- increased sensitivity to catecholamine (increased heart rate, pupillary dilatation, etc.) catabolic effects (weight loss despite increased appetite)

Deficiency of T_3 and T_4:

- increased reflex time (primarily due to slow muscle relaxation), dry skin and hair (decreased activity of sebaceous glands), mental sluggishness and lethargy
- decreased basal O_2 consumption (decreased heat production, sensitivity to cold environment)
- thick, puffy, edematous skin due to accumulation of hyaluronic acid (binds H_2O); this condition is termed *myxedema*.

Chapter 11. Endocrine regulation of homeostasis and organism nonspecific adaptation

This chapter will review the main aspects of the hormonal function of the following endocrinal glands: pancreatic, parathyroid and adrenal glands.

PANCREATIC GLAND

The development of mechanisms for the storage of large amounts of metabolic fuel was an important adaptation in the evolution of complex organisms. The processes involved in the digestion, storage, and use of fuels require a high degree of regulation and coordination. The pancreas, which plays a vital role in these processes, consists of two functionally different groups of cells. Cells of the *exocrine pancreas* produce and secrete digestive enzymes and fluids into the upper part of the small intestine. The *endocrine pancreas*, an anatomically small portion of the pancreas (1 to 2 % of the total mass), produces hormones involved in regulating fuel storage and use.

Three polypeptide hormones play important roles in the regulation of carbohydrate, lipid and protein metabolism and homeostasis. These are growth hormones, secreted by the anterior pituitary lobe and two hormones secreted by the pancreas, insulin and glucagon. Cells in the islets of Langerhans synthesize and secrete at least 4 different polypeptide or peptide hormones. Insulin is the hypoglycemic hormone secreted by the β-cells and glucagon is the hyperglycemic-glycogenolytic hormone secreted by the α-cells. Somatostatin is found in the δ-cells of islet tissue. Somatostatin is capable of inhibiting the release of insulin and glucagon. A fourth hormone present is pancreatic polypeptide (PP); it is produced by the F cells and is thought to decrease the enzymes in pancreatic juice.

I. INSULIN

Control of Insulin Secretion

Insulin is secreted into the portal circulation and approximately 50 % is removed by the liver before reaching peripheral tissues. The $t_{1/2}$ (half-life)

for circulating insulin is about 5 minutes. The rapid removal of insulin from the blood means that the $P_{insulin}$ level is highly responsive to the rate of insulin secretion and the factors which control it. Insulin secretion responds to a large number of metabolic and endocrine signals which are listed below. Establishing which of these factors are important is sometimes difficult because response to any particular agent is often influenced by previous stimulus history. In addition, secretion of the different pancreatic hormones can be shown (*in vitro*) to be regulated by paracrine effects of other islet hormones. For example, insulin secretion is stimulated by glucagon and inhibited by Somatostatin. Conversely, glucagon secretion is inhibited by insulin. Some of these *in vitro* observations are of uncertain physiological relevance.

Insulin secretion is regulated by:

Plasma glucose. This is known to be a direct effect on the secreting cells since perfusion of the isolated pancreas, islets of β-cells with more than 100 mg/dl of glucose will give an immediate increase in secretion rate which peaks at 4-5 minutes ("acute phase" or "acute release"; Fig. 11–2). The secretion rate then declines before undergoing a gradual and sustained increase ("second phase"). The acute phase although rapid and transient involves a relatively small proportion o f the total insulin in the pancreas. It is not dependent on protein synthesis, the duration of the glucose stimulus or its prestimulus level. The pancreas still contains large amounts of insulin at the start ofthe second phase but in spite of this continued secretion requires insulin synthesis early in the second phase. This is achieved by increased rates of transcription for the insulin gene and translation of insulin mRNA.

The stimulatory effect of glucose on insulin biosynthesis is independent of its effect on insulin secretion and occurs at lower glucose levels. The complex effects of $P_{glucose}$ on insulin secretion and biosynthesis probably have the net result of preventing hypoglycemia due to excessive insulin secretion and preventing hyperglycemia by ensuring replenishment of insulin stores between stimulatory episodes. The human pancreas contains about 180 units of insulin (1 mg = 25.4 units) but only secretes between 25–40 units/day in a normal stimulated state (i e. with meals). In the absence of meals the basal secretion rate is about 10 units/day. The natural and physiologically most important stimulus for insulin release is, in part, the rise in plasma glucose which occurs following a meal. The clinical "oral glucose tolerance test" duplicates, in a standard manner, the effect of a meal.

The mechanism by which β-cells are able to respond to changes in glucose levels (i.e. "the glucose sensor") has been intensively investigated and current views on this topic are summarized here and in the Fig. 11–3:

- glucose is taken up into the β-cells by a low affinity glucose transporter,
- metabolism of glucose via glycolysis increases ATP levels,
- ATP-sensitive K^+ channel closes leading to depolarization of the plasma membrane,
- depolarization opens Ca^{2+} channels,
- intracellular $[Ca^{2+}]$ increases leading to exocytosis of secretory granules.

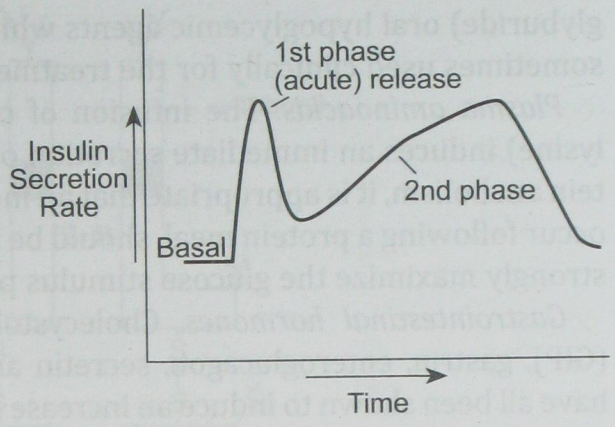


Fig. 11–2. Biphasic insulin response to a constant glucose stimulus.

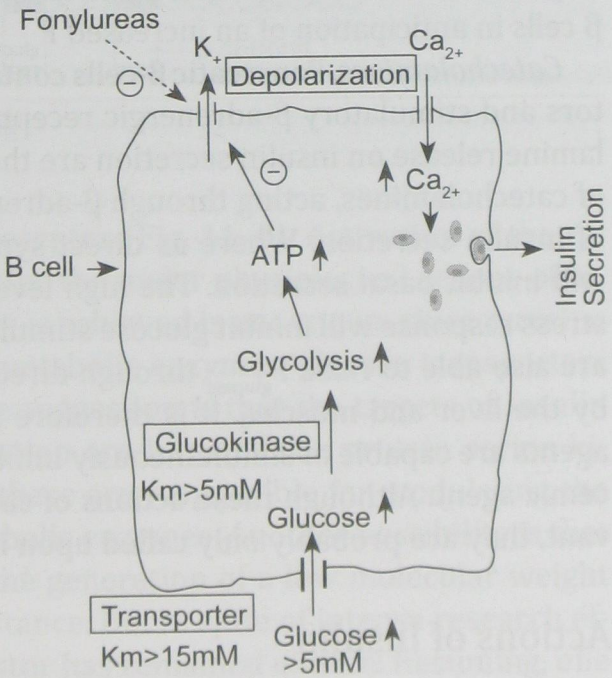


Fig. 11–3. Schematical illustration of the pancreatic β-cells response to plasma glucose level changing.

The ability of this mechanism to respond continuously to the changes in $P_{glucose}$ occurring between fasting and immediate postprandial i.e. approx 5 mM to 10 mM (90–180 mg/dl) is thought to be due to two components of the β-cell with relatively low affinity for glucose; the glucose transporter (K_m = 15-20 mM) and glucokinase (K_m for glucose = 5-20 mM. Drugs which interfere either with glucose uptake or metabolism will inhibit insulin release. By contrast, the ATP-sensitive K^+ channel is the target for the sulfonylureas (e.g. tolbutamide, glipizide, and

glyburide) oral hypoglycemic agents which stimulate insulin secretion and are sometimes used clinically for the treatment of type II diabetes.

Plasma aminoacids. The infusion of certain aminoacids (e.g., arginine and lysine) induces an immediate secretion of insulin. Since insulin stimulates protein anabolism, it is appropriate that an increase in plasma aminoacids, as would occur following a protein meal, should be an adequate stimulus. The aminoacids strongly maximize the glucose stimulus potential for insulin secretion.

Gastrointestinal hormones. Cholecystokinin (CCK), gastrointestinal peptide (GIP), gastrin, enteroglucagon, secretin and vasoactive intestinal peptide (VIP) have all been shown to induce an increase in insulin secretion, but GIP is the most functionally relevant one. These hormones are released following the ingestion of food and may account for the fact that the same amount of glucose (or aminoacids), given orally (e.g,. food), induces a much greater secretion of insulin than if given intravenously. Of the above G.I. hormones, GIP is the functionally most important one for insulin secretion. Its role may be to "sensitise" the pancreatic β cells in anticipation of an increased $P_{glucose}$ resulting from intestinal absorption.

Catecholamines. Pancreatic β-cells contain both inhibitory α-adrenergic receptors and stimulatory β-adrenergic receptors. The effects of changes in catecholamine release on insulin secretion are therefore complex. In general, low levels of catecholamines, acting through β-adrenergic receptors, maintain basal levels of insulin secretion. Where as direct sympathetic stimulation of the pancreas will inhibit basal secretion. The high levels of circulating epinephrine found in stress response will inhibit glucose stimulated insulin secretion. Catecholamines are also able to raise $P_{glucose}$ through direct effects on carbohydrate metabolism by the liver and muscles. It is therefore appropriate that these hyperglycemic agents are capable of simultaneously inhibiting the release of insulin, a hypoglycemic agent. Although these actions of catecholamines are physiologically relevant, they are probably only called upon in extreme hypoglycemia.

Actions of Insulin

The receptor for insulin is a heterotetramer composed of two α- (130 kDa) and two β- (95 kDa) subunits, which are formed by proteolytic cleavage and glycosylation of a 210 kDa proreceptor (Fig. 11–4). The-α subunits are on the exterior surface of the plasma membrane and are disulfide-linked to the β-subunits which span the membrane. The α-subunits contain the insulin-binding site and the β-subunits — a cytoplasmically oriented tyrosine kinase activity. Binding of insulin to the α-subunit results in increased tyrosine activity and autophos-

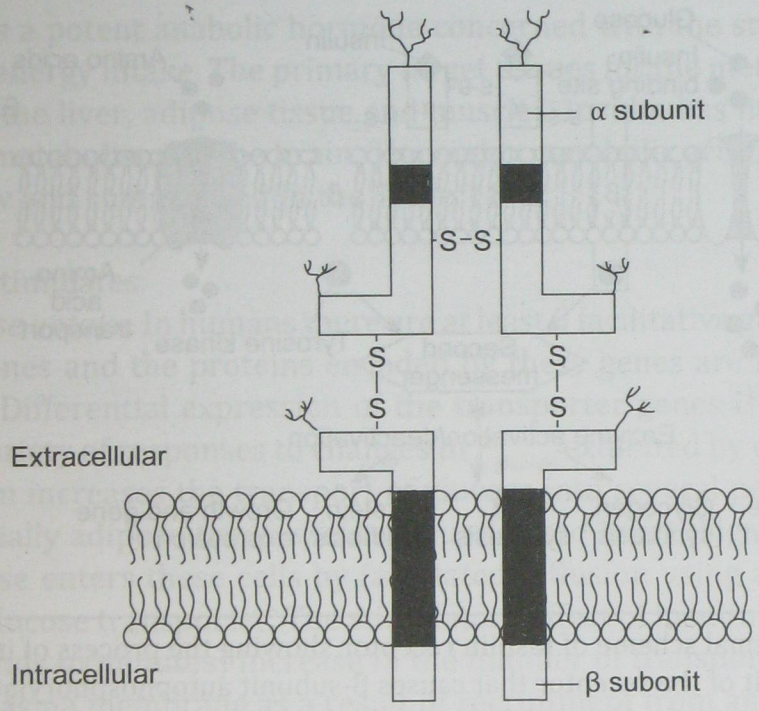


Fig. 11–4. The structure of insulin receptor.

pho-rylation of the β-subunit of the receptor (Fig. 11–5). Activation of the tyrosine kinase appears to be essential for the major physiological actions of insulin but the mechanisms by which this is achieved is not known since tyrosine phosphorylation of insulin-sensitive metabolic enzymes, glucose transporters etc. have not been demonstrated. One suggestion is that the targets of insulin receptor tyrosine kinase phosphorylation are intracellular protein serine kinases and/or phosphatases and that these are responsible for modulating the activities of the insulin-sensitive metabolic enzymes. Another possibility is that tyrosine kinase activation results in the generation of a low molecular weight mediator or "second messenger" substance. But in spite of intense research effort, the chemical nature of this mediator has remained elusive. Resuming, one could conclude that Tyrosine phosphorylation, in turn, activates a cascade of intracellular signaling proteins that mediate the effects of glucose on insulin, fat, and protein metabolism (Fig. 11–5).

It is reasonably certain that increased levels of other well-established second messengers such as cAMP, IP_3, diacylglycerol and Ca^{2+} do not play a major role in insulin action. However, some of the actions of insulin probably result from its demonstrated ability to decrease cAMP levels in many cells.

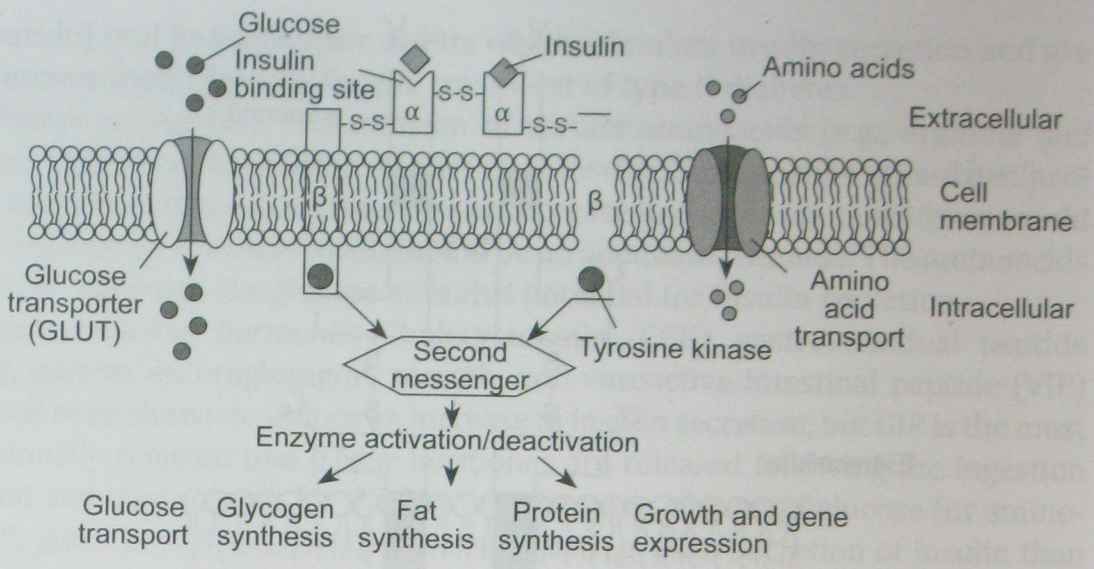


Fig. 11–5. Functional scheme of insulin receptor showing the process of its binding with the α-subunit of the receptor that causes β-subunit autophosphorylation that in turn initiates a cascade of intracellular biochemical processes mediate glucose influence on insulin, fat and protein metabolism.

The exact mechanism for this is not known but may involve both inhibition of adenylate cyclase and stimulation of cAMP phosphodiesterase. Patients with various types of inherited insulin resistance ("insulin resistance" results from the inability of tissues to respond to circulating insulin and has many causes) often have mutations in the insulin receptor gene which leads to the expression of a defective receptor.

One complication in elucidating the molecular mechanisms of insulin action is the large number of responses and the wide time scale over which they occur:

1. Translocation of glucose transporter to the membrane with increased glucose uptake (seconds to minutes)
2. Increased aminoacid uptake (seconds to minutes)
3. Changes in metabolic enzyme activities (minutes)
4. Modulation of major metabolic pathways (minutes to hours)
5. Induction and repression of enzyme synthesis (hours)
6. Insulin/receptor internalization (hours to days)
7. Cell division and growth (hours to days)

Insulin is a potent anabolic hormone concerned with the storage of energy at times of energy intake. The primary target tissues for the metabolic effects of insulin are the liver, adipose tissue and muscles. Insulin has little direct effect on energy metabolism by the brain. The major metabolic actions of insulin are given below and summarized in the Table 11–1:

Insulin stimulates:

1. *Glucose uptake.* In humans there are at least 5 facilitative glucose transporter genes and the proteins encoded by these genes are functionally distinct. Differential expression of the transporter genes is responsible for the variety of responses to changes in $P_{glucose}$ exhibited by different tissues. Insulin increases the transport of glucose into several peripheral tissues especially adipose *tissue, skeletal muscles and heart* (Fig. 11–6).
   Glucose enters these cells by facilitated diffusion using an insulin-sensitive glucose transporter. The stimulatory effect of insulin is due to a rapid (seconds to minutes) increase in the number of transporter molecules in the plasma membrane as a result of recruitment from an intracellular location. Thus, the rate of glucose uptake is increased allowing these tissues to take advantage of elevated $P_{glucose}$ following a meal. Conversely, in the absence of insulin the use of plasma glucose by these tissues is restricted when it is in short supply during a fast or starvation.
   Insulin does not have a direct stimulatory effect on glucose uptake in several tissues including liver and brain. These tissues contain insulin-insensitive glucose transporters and any increase in glucose uptake (especially by the liver) when insulin levels are elevated is secondary to increased glucose utilization. The liver and pancreatic β cell transporters have a relatively high K_m for glucose and "respond" to high $P_{glucose}$ either by using more of it or by secreting more insulin (β cells). By contrast, the transporters in brain cells and erythrocytes have a relatively low K_m for glucose and their glucose utilization will be relatively independent of changes in $P_{glucose}$. This is consistent with the dependence of latter cell types on glucose as the main energy source. Insulin does not increase active transport of glucose by the gut or kidney.
2. *Conversion of glucose to glycogen.* Insulin promotes glycogen synthesis in liver and skeletal muscle (Fig. 11–6) as a result of its ability to reduce cAMP levels. This leads to decrease in the activity of cAMP dependent protein kinase. One substitute for this kinase is glycogen synthetase which becomes inactive upon phosphorylation. Another substitute for cAMP-dependent protein kinase is phosphorylase b kinase which in

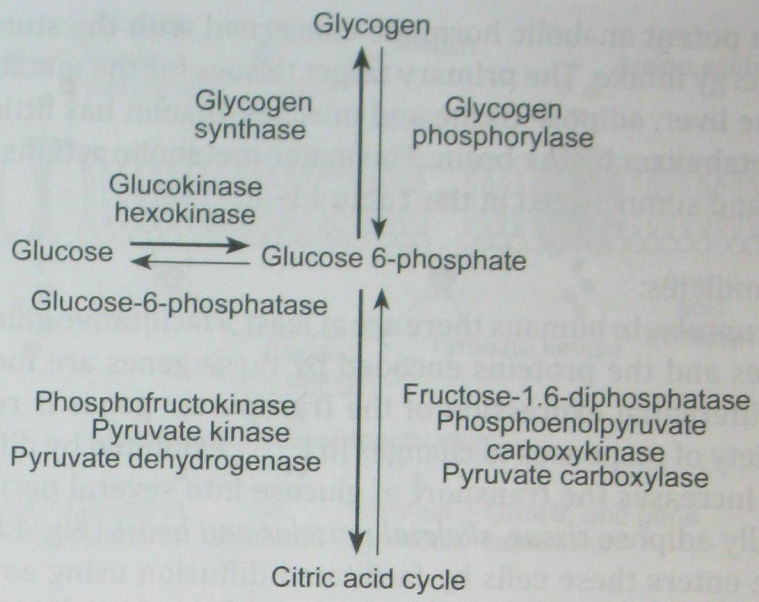


Fig. 11–6. Insulin stimulation of glycogen synthesis and glucose metabolism. Insulin promotes glucose uptake into target tissues, stimulates glycogen synthesis, and inhibits glycogenolysis. In addition, it promotes glycolysis in its target tissues. Thick arrows indicate processes stimulated by insulin; thin arrows indicate processes inhibited by insulin.

turn activates phosphorylase. The net result of a decrease in cAMP level, therefore, is to increase the activity of glycogen synthetase while simultaneously decreasing the activity of phosphorylase.

3. *Conversion of glucose into lipid.* In the liver and adipose tissue insulin acts at several points to stimulate the metabolic flux through two pathways:

- production of acetyl CoA from glucose is increased in result of activation of a number of glycolytic enzymes, i.e. glucokinase, phosphofructokinase, pyruvate kinase and pyruvate dehydrogenase;
- conversion of acetyl CoA into fatty acids as a result of activation of acetyl CoA carboxylase.

The net result of this is to divert excess dietary glucose into production of triglycerides, the main form of stored metabolic energy. In adipose tissue the lipid so produced is stored in large lipid droplets inside the adipocytes (Fig. 11–7).

4. *Aminoacid incorporation into protein.* Insulin increases amino acid uptake and protein synthesis in many tissues (Fig. 11–8). The effect on protein

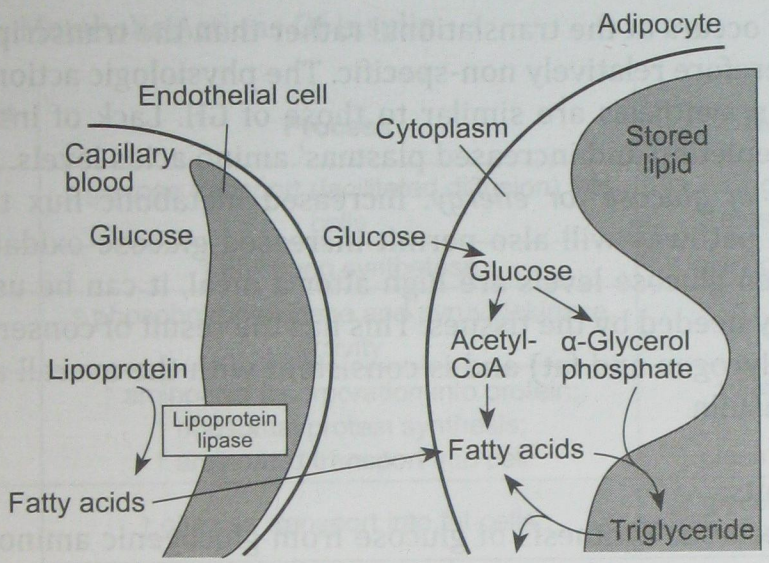


Fig. 11–7. Effects of insulin on lipid metabolism in adipocytes. Insulin promotes the accumulation of lipid (triglycerides) in adipocytes by stimulating the processes shown by the heavy arrows and inhibiting the processes shown by the light arrows. Similar stimulatory and inhibitory effects occur in liver cells.

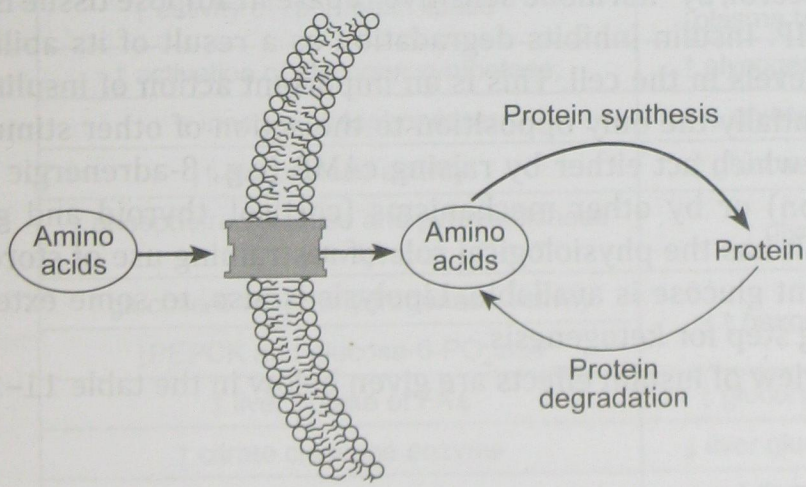


Fig. 11–8. Effects of insulin on protein synthesis and protein degradation. Insulin promotes the accumulation of protein by stimulating (heavy arrows) amino acid uptake and protein synthesis and by inhibiting (light arrows) protein degradation in liver, skeletal muscle, and adipose tissue.

synthesis occurs at the translational rather than the transcriptional level and is therefore relatively non-specific. The physiologic actions of insulin on protein synthesis are similar to those of GH. Lack of insulin causes protein depletion and increased plasmas' amino acids levels.

5. *Oxidation of glucose for energy.* Increased metabolic flux through the glycolytic pathway will also permit increased glucose oxidation to CO_2. Thus, when glucose levels are high after a meal, it can be used to meet the energy needed by the tissues. This has the result of conserving stored energy (glycogen and fat) and is consistent with the overall anabolic actions of insulin.

Insulin inhibits:

1. *Gluconeogenesis.* Synthesis of glucose from glucogenic amino acids (e.g., alanine) occurs mainly in the liver and is inhibited as a result of insulin action. Two key enzymes in the gluconeogenic pathway are sensitive to insulin : phosphoenolpyruvate-carboxykinase (PEPCK) and glucose 6-phosphatase. The purpose of this effect of insulin on gluconeogenesis is to inhibit the catabolic conversion of aminoacids/protein to glucose when sufficient glucose is available either from a meal or from stored glycogen.
2. *Lipolysis and ketogenesis.* Degradation of triglycerides, to fatty acids (FA) and glycerol, by "hormone sensitive" lipase in adipose tissue is dependent on cAMP. Insulin inhibits degradation as a result of its ability to lower cAMP levels in the cell. This is an important action of insulin because it is essentially the only opposition to the action of other stimulatory hormones which act either by raising cAMP (e.g., β-adrenergic stimuli and glucagon) or by other mechanisms (cortisol, thyroid and growth hormone) It has the physiological role of restraining use of stored fat when sufficient glucose is available. Lipolysis is also, to some extent, the rate limiting step for ketogenesis.

The overview of insulin effects are given below in the table 11–1.

Insulin Deficiency and Insulin Resistance

Insulin deficiency or resistance is responsible for the disease diabetes mellitus. It is a fairly common disease with > 6 million diagnosed and an estimated 4-6 million undiagnosed in the United States alone. It is indirectly responsible for about 40,000 deaths/year. Two major types are recognized; Type I (also

Table 11–1. Metabolic Actions Of Insulin

Major Target Tissue	Process	Net Effect
Muscle	↑ glucose transport (facilitated diffusion) into cells	↑ glycolysis ↓plasma glucose
Muscle	↑ glycogen synthetase	↑ glycogen synthetase
Muscle	↑ phosphofructokinase and pyruvatekinase activity	↑ glycolysis
Muscle	↑ aminoacid incorporation into protein; ↑ ribosomal protein synthesis; ↑ aminoacid transport into cell	↑ tissue protein; ↓ plasma amino acids
Adipose	↑ glucose transport into fat cells;	↑ glycogen utilization; ↑ lipogenesis
	↑ glucose oxidation provides: ↑ glycerol-PO_4 to etherify FA's; prevents FA back inhibition of acetyl CoA carboxylase; ↑ NADPH (cofactor for FA synthesis); ↑ acetyl CoA (2 carbon fragment for FA synthesis); ↓ activity of lipase (via ↓ cAMP);	↑ triglyceride synthesis ↑ FA synthesis ↑ FA synthesis ↑ FA synthesis ↓ lipolysis, ↓plasma FFA,
	↑ activity of lipoprotein lipase	↓ ketogenesis ↓plasma triglycerides
Liver	↑ activation of glycogen synthetase;	↑ glycogen synthesis
	↓ inactive phosphorylase;	↓ glycogenolysis
	↑ glycokinase activity;	↑ glucose utilization
	↑ phosphofructokinase and pyruvatekinase activity;	↑ glycolysis
	↑ glucose-6-PO_4-dehydrogenase activity	↑ hexose shunt
	↓PEPCK and glucose-6-PO_4tase	
	↑ liver uptake of FA's	↓ gluconeogenesis
	↑ citrate cleavage enzyme	↓ liver glucose output
	↑ amino acid uptake	↑ lipogenesis
	↑ amino acid incorporation into protein	↑ protein synthesis

known as insulin-dependent diabetes mellitus or IDDM) and Type II (non-insulin-dependent diabetes mellitus or NIDDM):

Type I — This results from a complete absence of insulin secretion. It occurs primarily in the young and is often referred to as "juvenile onset diabetes". It accounts for about 20 % of the cases and is believed to be due to autoimmune destruction of β cells. The disease often has a sudden onset with the "classical" symptoms noted below which can be corrected by injections of insulin.

Type II — This results from an inability of the pancreatic β-cells to compensate for insulin resistance by increasing insulin secretion. It occurs primarily in older individuals and is often referred to as "maturity onset diabetes". The cause of the insulin resistance and the declining insulin secretory capacity is not known but the occurrence of Type II diabetes has been linked to obesity. The disease has a slow onset and, in contrast to type I, patients are often asymptomatic at diagnosis. The metabolic abnormalities that contribute to hyperglycemia in people with type II diabetes include (1) impaired insulin secretion, (2) peripheral insulin resistance, and (3) increased hepatic glucose production (Fig. 11–9). Insulin resistance initially stimulates insulin secretion from the β- cells in the pancreas to overcome the increased demand to maintain a normoglycemic state. In time, the insulin response by the beta cells declines because of exhaustion. This results in elevated postprandial blood glucose levels. During the evolutionary phase, an individual with type II diabetes may not produce sufficient amounts of insulin levels because of β-cell failure. Because people with type II diabetes do not have an absolute insulin deficiency, they are less prone to ketoacidosis than are people with type I diabetes.

Apparently, sufficient insulin secretion and/or insulin sensitivity of adipose tissue remains to permit suppression of lipolysis. Many cases of type II diabetes can be treated by diet (to reverse insulin resistance due to obesity) and oral hypoglycemic agents (to stimulate insulin secretion by pancreatic β-cells). However, many do not respond and require insulin.

While diabetes began as a single deficiency, the pathophysiological processes described above require that therapy include sodium and potassium salts, glucose, bicarbonate and fluid as well as insulin. There are two categories of antidiabetic agents: insulin and oral medications. Because people with type I diabetes are deficient in insulin, they are in need of exogenous insulin replacement therapy from the start.

People with type II diabetes have increased hepatic glucose production; decreased peripheral utilization of glucose; decreased utilization of ingested carbohydrates; and over time, impaired insulin secretion from the pancreas (Fig. 11–10). The oral antidiabetic agents used in the treatment of type II diabetes exert their action in one or sometimes all of these areas. If good glycemic

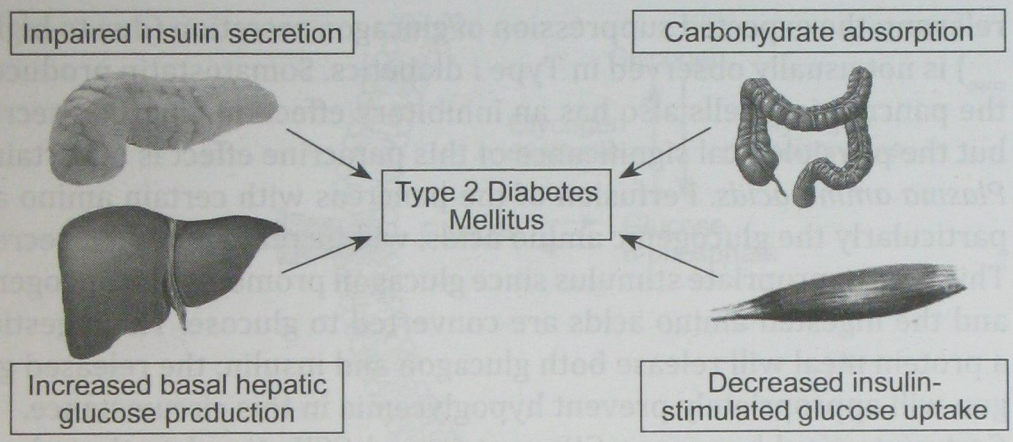


Fig. 11–10. Mechanisms of elevated blood glucose in type II diabetes mellitus.

control cannot be achieved with a combination of oral agents, insulin can be used with the oral agents or by itself.

II. GLUCAGON

Control of Glucagon Secretion

As with insulin glucagon is rapidly removed from the circulation by the liver ($t_{1/4}$ = 10 min) and the levels of glucagon are essentially determined by the rate of secretion. The factors which control the secretion and the metabolic actions of these hormones are essentially opposite. The insulin/glucagon ratio is therefore continually changing in response to the requirement for glucose storage or production, e.g. large carbohydrates meal, insulin/glucagon ratio = 70; small meal, 7; overnight fast, 2.3; starvation, 0.4. The metabolic changes observed in these situations reflect the balance between the opposing actions of the two hormones.

Glucagon secretion is regulated by:

1. *Plasma glucose.* A decrease in plasma glucose concentration is the primary stimulus for glucagon release. This is consistent with the primary physiological role of glucagon to maintain $P_{glucose}$ levels. Conversely, increased $P_{glucose}$ will inhibit glucagon secretion. The latter is a paracrine effect of insulin secreted by neighbouring β-cells and is probably physiologically

relevant; the expected suppression of glucagon secretion (due to high $P_{glucose}$) is not usually observed in Type I diabetics. Somatostatin produced by the pancreatic δ-cells also has an inhibitory effect on glucagon secretion but the physiological significance of this paracrine effect is uncertain.

2. *Plasma amino acids.* Perfusion of the pancreas with certain amino acids, particularly the glucogenic amino acids, will increase glucagon secretion. This is an appropriate stimulus since glucagon promotes gluconeogenesis, and the ingested amino acids are converted to glucose. The ingestion of a protein meal will release both glucagon and insulin; the released glucagon will appropriately prevent hypoglycemia in this circumstance.
3. *Gastrointestinal hormones.* GIP, gastrin and CCK stimulate the release of glucagon.
4. *Stimulation of sympathetic nerves to the pancreas.* Activation of β-adrenergic receptors on α-cells increases glucagon secretion.
5. *Exercise.* Exercise stimulates glucagon secretion, often increases four-to five folds. This is an appropriate stimulus enabling the body to respond to increase demand for energy by stimulating liver glycogenolysis (see below).

Actions of Glucagon:

The actions of glucagon on tissues (mainly liver and adipose tissue) are mediated through activation of the adenylate cyclase-cAMP system. It is the primary and continuous mechanism for prevention of the hypoglycemia which would result if the actions of insulin were unopposed. The net result of all its actions is to facilitate maintenance of normal $P_{glucose}$ levels in the absence of dietary glucose.

1. *Increases plasma glucose.* Glucagon is an important regulator of hepatic glycogen metabolism. Increased cAMP levels by glucagon activity activate the cAMP-dependent protein kinase which converts phosphorylase to its active form resulting in increased glycogen breakdown (Fig. 11–11). Simultaneously glycogen synthetase is phosphorylated by the same kinase resulting in an inactive form and decreased glycogen synthesis. In the liver (but not muscle) glucose is produced by the sequential action of phosphorylase, phosphoglucomutase and glucose 6-phosphatase. It can freely exit from the cell via the insulin-insensitive glucose transporter and is used to maintain $P_{glucose}$ levels.
2. Glucagon *increases hepatic glucose output* by stimulating gluconeogenesis. In addition, to promoting hepatic glucose production by stimulating glycogenolysis, glucagon stimulates hepatic gluconeogenesis (Fig. 11–12). It

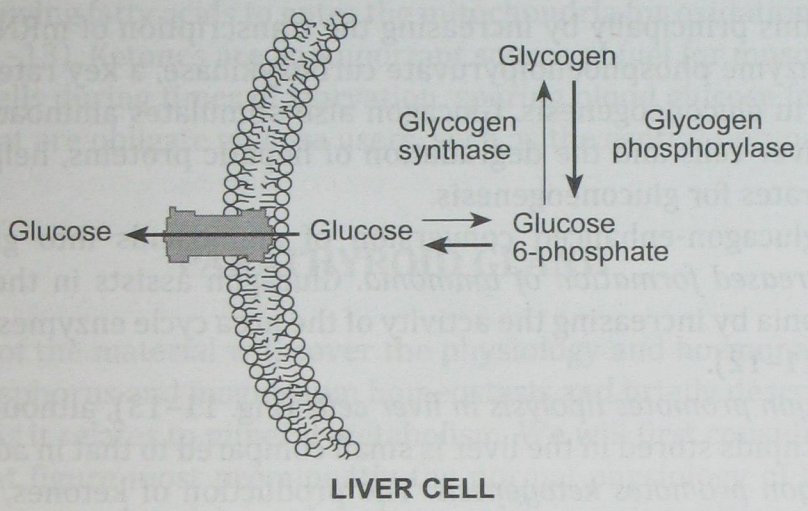


Fig. 11-11. The role of glucagon in glycogenolysis and glucose production in liver cells. Thick arrows indicate processes stimulated by glucagon; thin arrows indicate processes inhibited by glucagon.

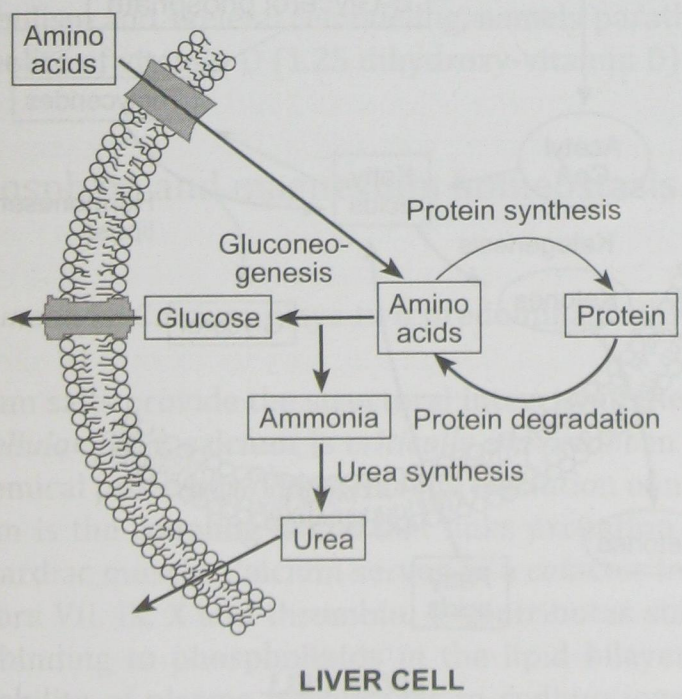


Fig. 11-12. The role of glucagon in gluconeogenesis and ureagenesis in liver cells. Thick arrows indicate processes stimulated by glucagon; thin arrow indicates processes inhibited by glucagon

does this principally by increasing the transcription of mRNA coding for the enzyme phosphoenolpyruvate carboxykinase, a key rate-limiting enzyme in gluconeogenesis. Glucagon also stimulates aminoacid transport into liver cells and the degradation of hepatic proteins, helping provide substrates for gluconeogenesis.

3. The glucagon-enhanced conversion of aminoacids into glucose leads to *increased formation of ammonia*. Glucagon assists in the disposal of ammonia by increasing the activity of the urea cycle enzymes in liver cells (Fig. 11–12).
4. *Glucagon promotes lipolysis in liver cells* (Fig. 11–13), although the quantity of lipids stored in the liver is small compared to that in adipose tissue.
5. *Glucagon promotes ketogenesis.* The production of ketones, by lowering the levels of malonyl-CoA, relieving an inhibition of palmitoyl transferase

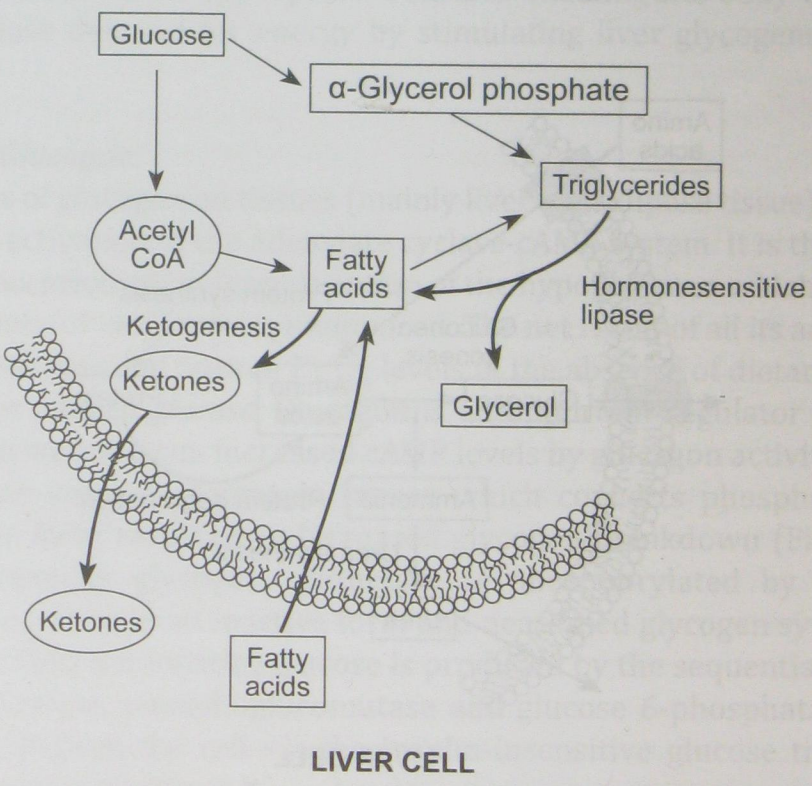


Fig. 11–13. The role of glucagon in lipolysis and ketogenesis in liver cells. Thick arrows indicate processes stimulated by glucagon; thin arrows indicate processes inhibited by glucagon.

and allowing fatty acids to enter the mitochondria for oxidation to ketones (Fig. 11–13). Ketones are an important source of fuel for muscle cells and heart cells during times of starvation, sparing blood glucose for other tissues that are obligate glucose users, such as the central nervous system.

PARATHYROID GLAND

This part of the material will cover the physiology and hormonal control of calcium, phosphorus and magnesium homeostasis and briefly describe skeletal remodeling as it relates to mineral metabolism. We will first consider the three elements that figure most prominently the normal physiology of the mineral metabolism — namely calcium, phosphate and magnesium. We consider them together primarily because similar conditions and hormones tend to control the absorption and excretion of these ions and they all have a role as major constituents of bone mineral. The major emphasis will be on calcium levels, which massive fluctuations of it can be life-threatening.

Finally, we will consider the three major hormones involved in the control of mineral metabolism and skeletal remodeling, namely parathyroid hormone, the active metabolite of vitamin D (1.25 dihydroxy-vitamin D) and calcitonin.

I. Calcium, phosphate and magnesium homeostasis

Calcium

Physiologic Functions: Calcium plays two predominant physiologic roles in the organism.

In *bone*, calcium salts provide the structural integrity of the skeleton.

In the *extracellular fluid*, calcium is *critically important* in the control of a variety of biochemical processes. These include excitation of nerve and muscle by which calcium is the coupling factor that links excitation and contraction in skeletal and cardiac muscle. Calcium serves as a cofactor in the coagulation cascade for factors VII, IX, X and thrombin. It contributes stability to plasma membranes by binding to phospholipids in the lipid bilayer and also regulates the permeability of plasma membranes to sodium ions. A reduction in ionized calcium increases sodium permeability and enhances the excitability of all excitable tissues; an increase in ionized calcium has the opposite effect. Calcium is essential for release of cellular secretory products. Calcium serves in

several second messenger systems, including the calmodulin, cAMP and phosphotidylinositol cell signaling systems.

Absorption: The intestinal absorption of calcium (and this also applies to magnesium and phosphate) provides the only supply of these minerals to meet the body's needs. The quantity of calcium absorbed by the gut is determined both by the amount present in the diet and the ability of the intestine to absorb it.

In general intestinal calcium absorption represents the sum of 2 processes:
1) saturable trans-cellular absorption that is physiologically regulated
2) nonsaturable paracellular absorption that depends on the concentration of calcium in the gut lumen.

Net intestinal calcium absorption is defined as the difference between dietary calcium intake and fecal calcium excretion. The relationship between net intestinal Ca^{2+} absorption and dietary calcium intake is complex. When dietary calcium intake is less than 200 mg/day, net intestinal calcium absorption is 0; that is, fecal excretion of calcium exceeds intestinal absorption. On average, healthy adults require daily calcium intakes of greater than 400 mg/day to maintain normal calcium balance. However, there is a great deal of individual variation in calcium absorption at any level of calcium intake. Also note that when dietary calcium intake exceeds 1000 mg/day, the amount of calcium absorbed tends to level off at an average of 300 mg/day (Fig. 11–14) .

Calcium absorption occurs primarily in the duodenum and jejunum. The active form of vitamin D, $1.25(OH)_2D$, is the only hormone controlling calcium absorption. Experimental data in humans have shown that in the absence of $1.25\ (OH)_2D$, only 7 % of a normal calcium intake is absorbed. Dietary calcium intake and $1.25(OH)_2D$ interact in a complex way to determine calcium absorption. Reduced absorption can occur when dietary calcium intake is low, $1\,25(OH)_2D$ levels are reduced or when the intestine is unable to respond to $1.25(OH)_2D$.

Conversely, increased calcium absorption can occur when die-tary calcium intake is increased or when $1.25(OH)_2D$ levels are increased.

Excretion/Balance: Among healthy adults in *steady-state* who are maintaining a diet which provide 800 mg of calcium a day, net intestinal calcium absorption averages about 20 % of dietary calcium intake or approximately 160 mg/day. Bone is continuously being remodelled but under normal circumstances, bone re-absorption and bone formation are tightly coupled both with respect to time and amount.

Thus, bone reabsorption equals bone formation and there is no net contribution to calcium balance from the skeleton in normal young adults. Therefore, in order to sustain a steady state situation, in which intake equals output and balance is zero, the kidneys must excrete 160 mg of calcium a day. The normal serum calcium averages 10 mg/dl of which approximately half is bound to albumin and thus cannot be filtered through the glomerulus's.

A normal adult with a glomerular filtration rate of 125 ml/min thus filters approximately 10,000 mg of calcium per day. Under normal circumstances, more than 98 % of filtered calcium is reabsorbed by the tubules and less than 2 % of filtered calcium is excreted into the urine. Normal daily urinary calcium excretion varies from 40 mg to 300 mg/day and averages 160 mg/day when a normal dietary calcium intake is maintained (600–1200 mg/dl).

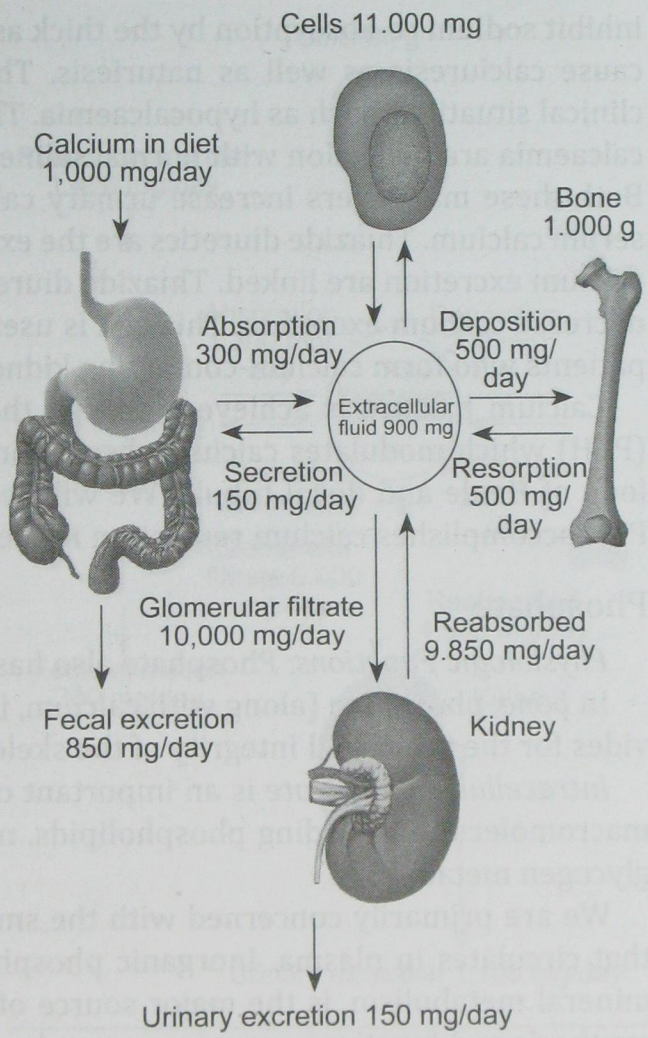


Fig. 11–14. Typical daily exchanges of calcium between different tissue compartments in a healthy adult.

The renal handling of calcium is very similar to the renal handling of sodium. In the case of both sodium and calcium, 65–75 % of the filtered load is reabsorbed in the proximal tubule, 20–25 % in the loop of Henley and 5–10 % in the distal nephron. There is good correlation between the fractional excretion of sodium and calcium; any maneuver that increases sodium excretion — such as an increase in dietary sodium intake or an expansion of plasma volume — will also increase calcium excretion. For example, *loop* diuretics that

inhibit sodium re-absorption by the thick ascending limb of the loop of Henley cause calciuresis as well as naturiesis. This correlation is useful in certain clinical situations such as hypocalcaemia. The mainstays of treatment of hypocalcaemia are hydration with normal saline and administration of furosemide. Both these maneuvers increase urinary calcium excretion and tend to lower serum calcium. Thiazide diuretics are the exception to the rule that sodium and calcium excretion are linked. Thiazide diuretics increase sodium excretion but decrease calcium excretion. This fact is useful in the therapy of hypercalciuric patients who form calcium-containing kidney stones.

Calcium balance is achieved through the actions of parathyroid hormone (PTH) which modulates calcium absorption in the thick ascending link of the loop of Henle and distal tubule. We will return to the specific ways in which PTH accomplishes calcium regulation in the sections to follow.

Phosphate

Physiologic Functions: Phosphate also has two major physiologic roles.

In *bone*, phosphate (along with calcium, in the form of hydroxyapatite) provides for the structural integrity of the skeleton.

Intracellular phosphate is an important component of a variety of essential macromolecules including phospholipids, nucleic acids, phosphoproteins and glycogen metabolism.

We are primarily concerned with the small amount of *inorganic phosphate* that circulates in plasma. Inorganic phosphate plays a crucial role in normal mineral metabolism, is the major source of phosphate for organic phosphate synthesis and functions as an important body buffer.

Absorption: The wide daily fluctuations in serum phosphate levels are due primarily to changes in dietary intake. Intestinal phosphate absorption like calcium absorption is dependent on both *passive* phosphate transport related to high intraluminal postprandial phosphate concentrations and on *active* phosphate transport that is located in the jejunum and stimulated by 1.25 dihydroxy-vitamin D. Under usual circumstances, phosphate absorption is directly related to intake, averaging about 2/3 of the ingested load. Since phosphate is a major constituent of all cells, dietary phosphate intake is 800-900 mg/day in an average American diet. A typical adult ingests approximately 1,400 mg/day of phosphorus (Fig. 11–15).

In marked contrast to calcium, most (1,300 mg/day) of this phosphorus is absorbed from the GI tract, typically as inorganic phosphate. There is an obligatory contribution of phosphorus to the contents of the GI tract (about 200

mg/day), much like that for calcium, resulting in a net uptake of phosphorus of 1,100 mg/day and excretion of 300 mg/day via the feces. Thus, the majority of ingested phosphate is absorbed from the GI tract and little passes through to the feces.

Excretion/Balance: Despite the large daily influxes in phosphate derived from diet, under steady state conditions, intake equals output and balance is zero. There is no net gain or loss of phosphate from the processes of bone remodeling and intracellular metabolism, under normal circumstances.

Therefore, phosphate homeostasis is largely controlled by alterations in urinary excretion. An increase in dietary phosphate is associated with an inhibition of renal tubular phosphate reabsorption and vice versa. Thus, phosphate excretion varies directly with phosphate absorption. The control of phosphate excretion is mediated primarily through the actions of parathyroid hormone.

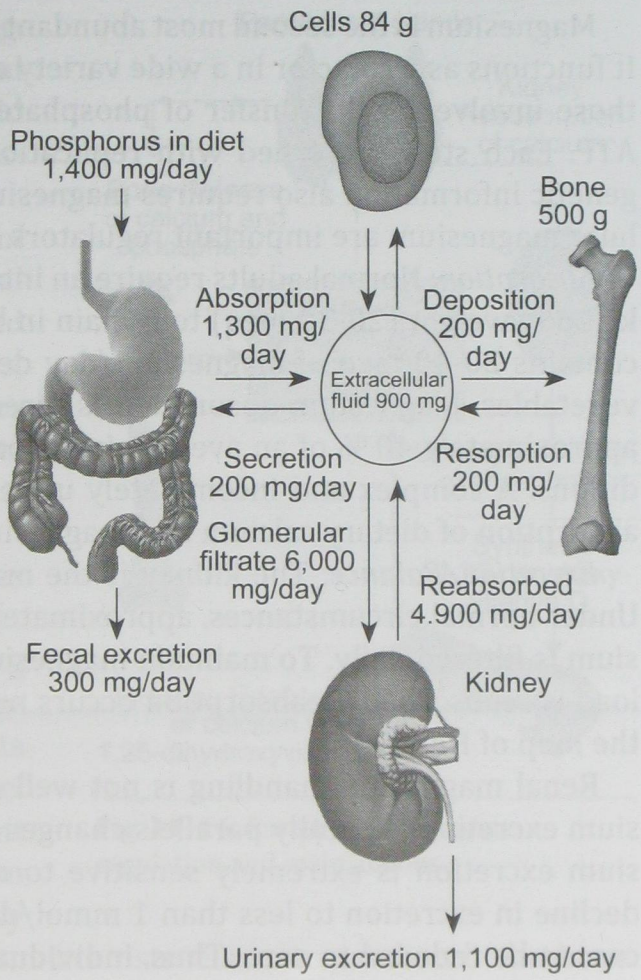


Fig. 11–15 Typical daily exchanges of phosphate between different tissue compartments in a healthy adult.

Magnesium

Physiologic Functions: Like calcium and phosphorus, most of the magnesium in the body is located in the skeleton. Unlike calcium and phosphorus, skeletal magnesium is not a part of the hydroxyapatite crystal structure but appears to be located on the crystal surface. Despite its location on the surface, bone magnesium is not freely exchangeable with ECF magnesium.

Magnesium is the second most abundant intracellular cation after potassium. It functions as a cofactor in a wide variety of enzymatic reactions, including all those involved with transfer of phosphate groups and all those that require ATP. Each step concerned with replication, transcription and translation of genetic information also requires magnesium. Both intracellular and extracellular magnesium are important regulators of neuromuscular excitability.

Absorption: Normal adults require an intake of 0.3–0.35 meq of magnesium/kg body weight (20-30 meq) to remain in balance. The average American diet contains 20-40 meq of magnesium/day derived mainly from meat and green vegetables. Magnesium absorption is directly related to dietary intake, with approximately 40 % of an average ingested load absorbed under normal conditions. A complex and incompletely understood relationship exists between absorption of dietary calcium and magnesium.

Excretion/Balance: The kidney is the main site of excretion of magnesium. Under normal circumstances, approximately 3000 mg or 125 mmol of magnesium is filtered daily. To maintain magnesium balance, 90–99 % of the filtered load is reabsorbed. Reabsorption occurs mainly in the thick ascending limb of the loop of Henle.

Renal magnesium handling is not well understood but changes in magnesium excretion generally parallels changes in calcium excretion. Renal magnesium excretion is extremely sensitive to deprivation states; there is a rapid decline in excretion to less than 1 mmol/day. However, magnesium excretion cannot be reduced to zero. Thus, individuals whose magnesium intake is low can become magnesium depleted. In contrast, the kidney has a vast capacity to excrete magnesium, so that even when magnesium intakes are extremely high, magnesium overload does not occur when renal function is normal.

II. Hormonal regulation of mineral metabolism

Three hormones — parathyroid hormone, $1.25(OH)_2D$ and to a lesser extent calcitonin — are the principal regulators of mineral metabolism. These 3 hormones act in concert to maintain the concentration of ionized calcium in the ECF within a remarkably narrow range despite widely varying dietary intakes.

Parathyroid hormone

Parathyroid hormone (PTH) is an 84 amino acid peptide secreted by the chief cells of the parathyroid glands. There are usually 4 parathyroid glands,

each weighing approximately 40 mg. They most often are located behind and at the 4 poles of the thyroid gland. Ectopic locations and accessory glands are frequent.

Target Organs and Biological Actions of PTH: The two principal target organs of PTH are bone and kidney (Fig. 11–16). In both tissues, PTH binds to cell surface receptors and acts via stimulation of adenylate cyclase and increasing cellular cyclic AMP levels.

In *bone* the major action of PTH is to stimulate bone reabsorption.

Paradoxically, bone reabsorbing osteoclasts do not have PTH receptors and PTH has no direct effects on isolated osteoclasts in vitro. However, PTH receptors *are* present on bone-forming osteoblasts. When osteoblasts are exposed to PTH, they secrete substances that in turn stimulate osteoclasts to re-absorb the bone. *The effect of PTH action on bone is thus to release calcium and phosphate ions into the extracellular fluid.*

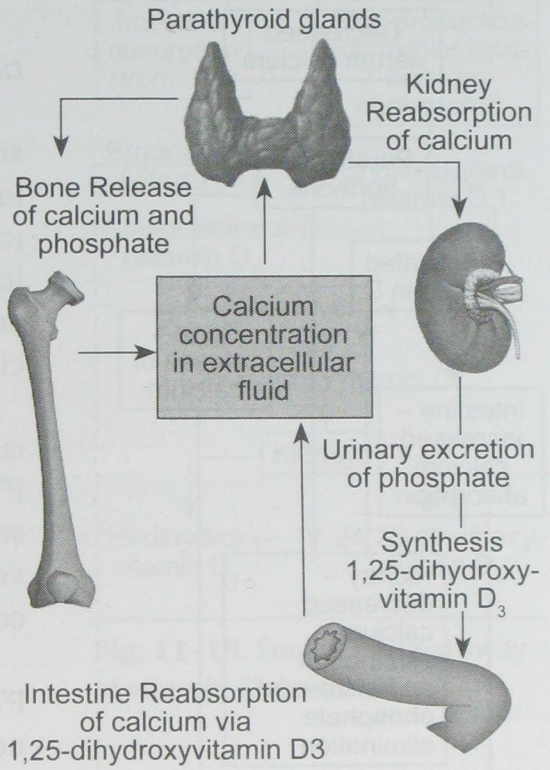


Fig. 11–16. Parathyroid hormone regulation and main actions.

In the *kidney*, PTH has three actions:

- PTH increases calcium reabsorption in the distal nephron.
- Conversely, PTH *decreases* phosphate absorption in the proximal tubule.
- PTH also stimulates the production of 1.25 $(OH)_2D$ by the kidney.

Thus, the ultimate effect of an in-crease in PTH secretion is to *increase serum* calcium level by virtue of an increase in *calcium* release from bone and an increase in renal calcium reabsorption (Fig. 11–17). The phosphate released into the circulation from the increased bone resorption is thus disposed of by renal losses resulting from the effect of PTH to decrease phosphate reabsorption in the kidney.

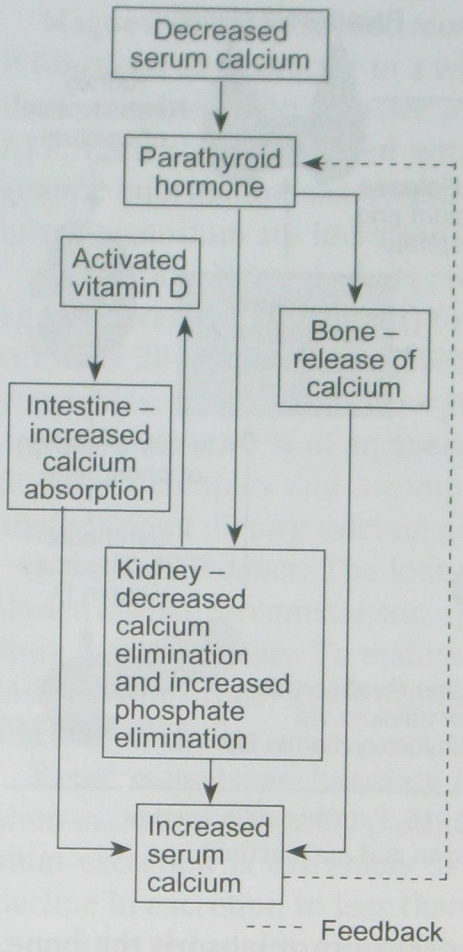


Fig. 11–17. Serum calcium level regulation by PTH.

Regulation of PTH Synthesis and Secretion:

The primary regulator of PTH synthesis and secretion is the concentration of ionized calcium in the serum. When the ionized calcium level falls below its physiological set point (1.3 mm), PTH synthesis and secretion is stimulated and serum calcium rises again (Fig. 11–18).

Conversely, a rise in the ionized calcium to greater than 1.3 mm suppresses PTH secretion with a resultant fall in the serum calcium (Fig. 11–18). However, even very high serum calcium levels cannot completely suppress PTH secretion.

The secretion, synthesis, and action of PTH are also influenced by magnesium. Magnesium serves as a cofactor in the generation of cellular energy and is important in the function of second messenger systems. Magnesium's effects on the synthesis and release of PTH are thought to be mediated through these mechanisms. Because of its function in regulating PTH release, severe and prolonged hypomagnesemia can markedly inhibit PTH levels.

PTH Clinical Correlations: Primary hyperparathyroidism is a very common disorder that most commonly is due to a benign adenoma of a parathyroid gland that overproduces PTH. This disorder is characterized by hypercalcemia, hypophosphatemia and bone and renal complications.

1.25 Dihydroxyvitamin D

In the past two decades, there have been major advances in our understanding of vitamin D. Conceptually vitamin D has come to be regarded as a classical steroid hormone rather than a fat soluble vitamin.

Sources and Metabolism of 1.25 Dihvdroxyvitamin D: The metabolically active form of vitamin D, 1.25 dihydroxyvitamin D is derived from 2 principal sources (Fig. 11–19).

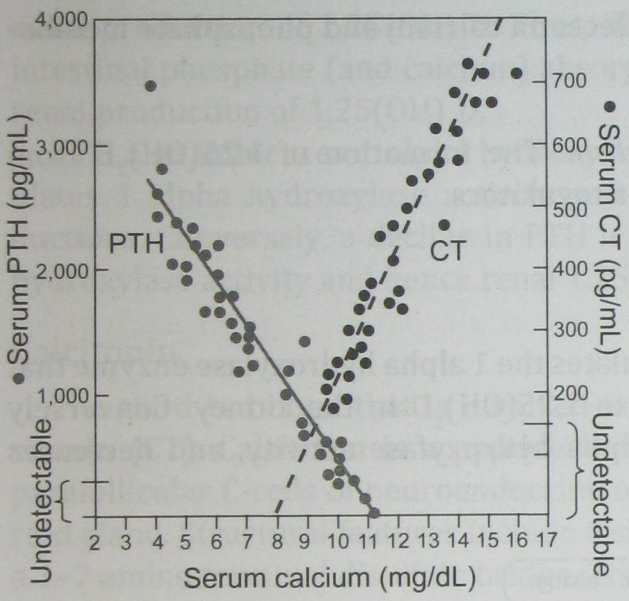


Fig. 11–18. Plasma calcium concentration changes influence on PTH and CT secretion.

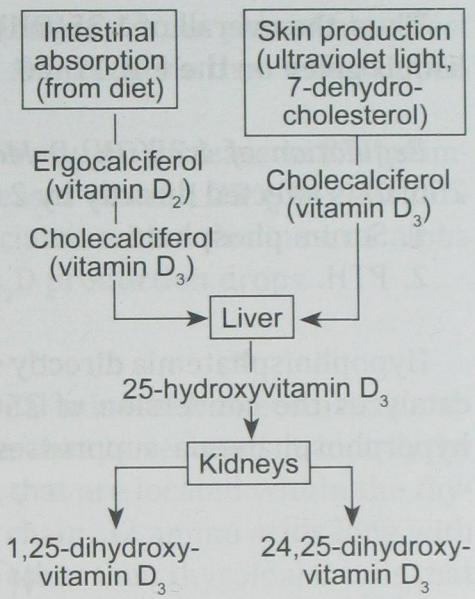


Fig. 11–19. Source and pathway of vitamin D activation.

Vitamin D_3 or *cholecalciferol* is formed by the action of ultraviolet light upon a precursor molecule present in skin, 7-dehydrocholesterol.

Vitamin D_2 or *ergocalciferol* is formed by irradiation of a precursor *ergosterol*, found primarily in plants.

To be available to humans, Vitamin D_2 or ergosterol must be ingested in the diet. When there is adequate exposure to ultraviolet light, the conversion of 7-dehydrocholesterol to cholecalciferol can provide adequate vitamin D without dietary supplementation.

Both vitamin D_2 and vitamin D_3 can be regarded as prohormones, that is, they require further modification to be metabolically active. In general, they are processed similarly and can be considered interchangeably. After vitamin D is formed in the skin (absorbed from the gut), it is transported to the liver bound to a specific vitamin D binding protein that binds vitamin D, as well as all other vitamin D metabolites.

In the liver, vitamin D is metabolized by hydroxylation at carbon 25 to yield 25 hydroxyvitamin D. This is the major circulating form of vitamin D and the serum level of 25 hydroxyvitamin D is a good indicator of body stores of this hormone. Its serum concentration is 25 ng/ml and its half life is 15 days.

Thus, the overall of $1.25(OH)_2D$ effects on calcium and phopsphate metabolism is given on the Fig. 11--20.

Regulation of $1.25(OH)_2D$ Metabolism: The formation of $1.25(OH)_2D$ from 250HD is affected directly by 2 major regulators:

1. Serum phosphate.
2. PTH.

Hypophoisphatemia directly stimulates the 1 alpha hydroxylase enzyme that catalyzes the conversion of 250HD to $1,25(OH)_2D$ in the kidney. Conversely hyperphosphatemia suppresses 1 alpha hydrpxyfase activity, and decreases

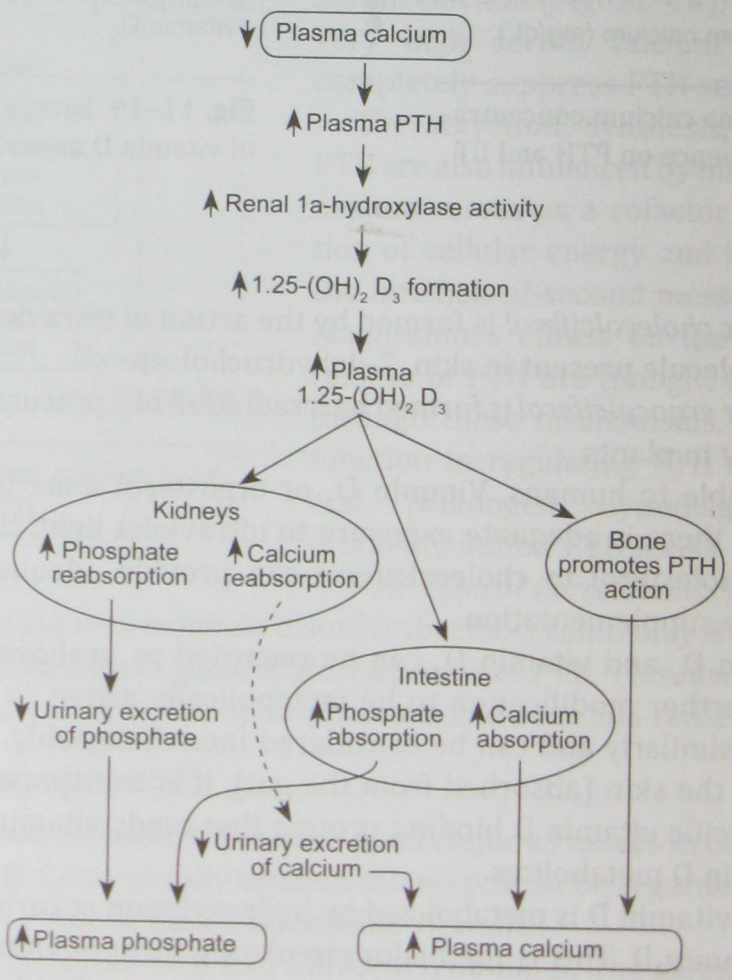


Fig. 11–20. Effect of $1.25(OH)_2D$ on calcium and phosphate metabolism.

renal production of $1,25(OH)_2D$. Thus, alterations in serum phosphate affect intestinal phosphate (and calcium) absorption *indirectly* via their effects upon renal production of $1,25(OH)_2D$.

PTH also influences 1 alpha hydroxylase activity. An increase in PTH stimulates 1 alpha hydroxylase activity and increases renal 1,25 (OH) 2D productions. Conversely, a decline in PTH is associated with a decline in 1 alpha hydroxylase activity and hence renal $1.25(OH)_2D$ production drops.

Calcitonin

The third hormone that is involved in normal mineral homeostasis is calcitonin (CT). Calcitonin is a peptide hormone that is secreted primarily by parafollicular C-cells of neuroendocrine origin that are located within the thyroid gland. Structural features include a single chain, 32 amino acids long with a 1–7 amino terminal disulfide bridge. Tissues other than thyroidal C cells that produce CT are pituitary cells and widely distributed neuroendocrine cells.

Target Organs and Biological Actions of CT: The main target organs of CT are bone and kidney (Fig. 11–21). The primary biological effect of CT *in bone* is to

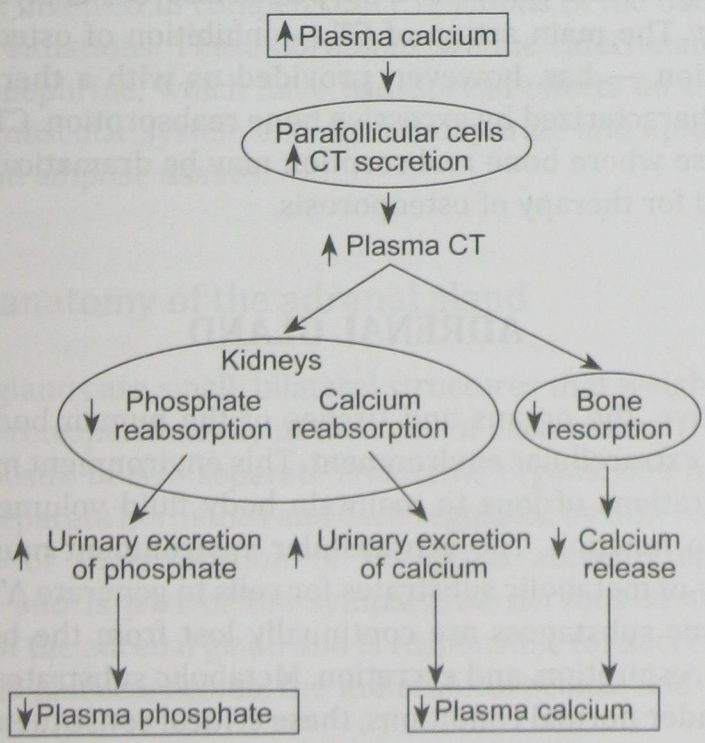


Fig. 11–21. Effect of calcitonin (CT) on calcium and phopsphate metabolism

inhibit osteoclast mediated bone reabsorption. Within minutes of administration, CT causes the osteoclast to shrink in size and decrease its bone reabsorbing activity.

In the kidney, calcitonin has been reported to increase urinary excretion of calcium and phosphorus — all at supraphysiologic concentrations. Both these effects of calcitonin are mediated by stimulation of adenylate cyclase and increased c- AMP synthesis.

Other extra skeletal effects of CT have been described — particularly in the pituitary and the CNS where it may function as a neurotransmitter. Certain clinical evidence suggests that CT may have an analgesic effect.

Regulation of Calcitonin Secretion: The serum ionized calcium level is the primary regulator of CT secretion. An acute rise in blood ionized calcium stimulates CT secretion; conversely an acute fall in blood ionized calcium inhibits CT secretion (Fig. 11–18). However, the effects of chronic hypercalcemia and hypocalcaemia are not well understood.

Gastrointestinal peptides (gastrin, cholecystokinin) also stimulate CT secretion when given in pharmacologic doses and this observation has led to the use of provocative testing with pentagastrin as a means of diagnosing CT-producing tumors such as medullary thyroid cancer — a neoplastic disorder of thyroidal C cells.

CT as a Drug: The main action of CT — inhibition of osteoclast mediated bone reabsorption — has, however, provided us with a therapy for certain disease states characterized by excessive bone reabsorption. CT is widely used in Paget's disease where bone reabsorption may be dramatically increased. It is also approved for therapy of osteoporosis.

ADRENAL GLAND

To remain alive, the organs and tissues of the human body must have a finely regulated extracellular environment. This environment must contain the correct concentrations of ions to maintain body fluid volume and to enable excitable cells to function. The extracellular environment must also have an adequate supply of metabolic substrates for cells to generate ATP. Salts, water, and other organic substances are continually lost from the body as a result of perspiration, respiration, and excretion. Metabolic substrates are constantly used by cells. Under normal conditions, these critical constituents of the body's extracellular environment are replenished by the intake of food and liquids. However, a person can survive for weeks on little else but water because the

body has a remarkable capacity for adjusting the functions of its organs and tissues to preserve body fluid volume and composition.

The *adrenal glands* play a key role in making these adjustments. This is readily apparent from the fact that an adrenalectomized animal, unlike its normal counterpart, cannot survive prolonged fasting. Its blood glucose supply diminishes, ATP generation by the cells becomes inadequate to support life, and the animal eventually dies. Even when fed a normal diet, an adrenalectomized animal typically loses body sodium and water over time, and eventually dies of circulatory collapse. Its death is caused by a lack of certain steroid hormones that are produced and secreted by the cortex of the adrenal gland.

The glucocorticoid hormones, *cortisol and corticosterone*, play essential roles in adjusting the metabolism of carbohydrates, lipids, and proteins in liver, muscle, and adipose tissues during fasting, which assures an adequate supply of glucose and fatty acids for energy metabolism despite the absence of food. The mineralocorticoid hormone *aldosterone* stimulates the kidneys to conserve sodium and, hence, body fluid volume. The glucocorticoids also enable the body to cope with physical and emotional traumas or stresses.

Hormones produced by the other endocrine component of the adrenal gland, the medulla, are involved in compensatory reactions of the body to trauma or life-threatening situations. These hormones are the catecholamines, epinephrine and norepinephrine, which have widespread effects on the cardiovascular system and muscular system and on carbohydrate and lipid metabolism in liver, muscle, and adipose tissues.

I. Functional anatomy of the adrenal gland

The adrenal glands are small, bilateral structures that weigh approximately 5 g each and lie retroperitoneally at the apex of each kidney (Fig. 11–22). It is actually a composite of two separate endocrine organs, one inside the other, each secreting separate hormones and each regulated by different mechanisms.

The medulla or inner portion of the gland secretes epinephrine and norepinephrine and is part of the sympathetic nervous system. The cortex forms the bulk of the adrenal gland and is responsible for secreting three types of hormones: the glucocorticoids, the mineralocorticoids, and the adrenal sex hormones.

In the adult human, the adrenal cortex consists of three histologically distinct zones or layers (Fig. 11–22). The outer zone, which lies immediately

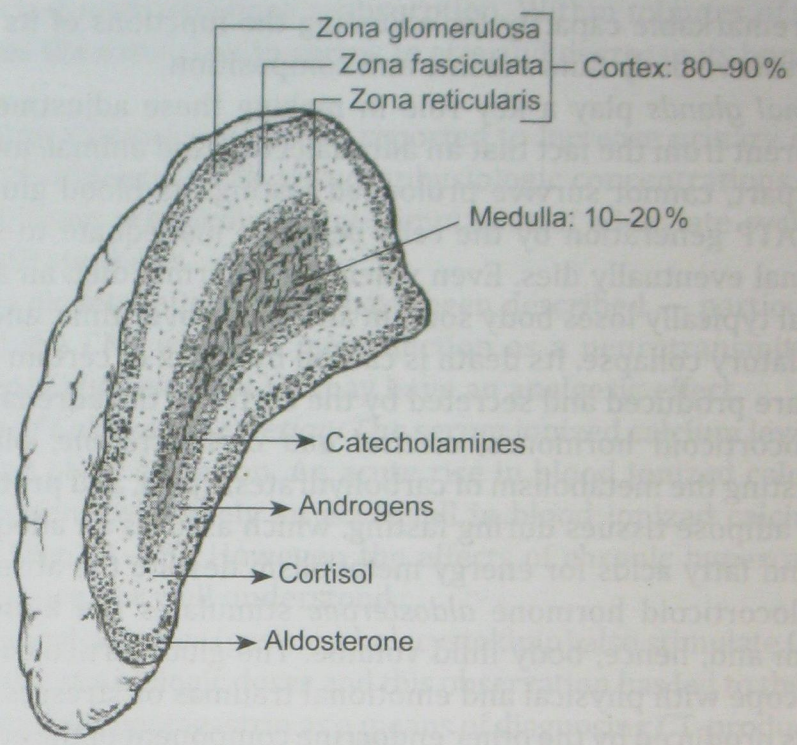


Fig. 11–22. Schematical representation of adrenal cortex three zones and corresponding hormones secretion.

under the capsule of the gland, is called the *zona glomerulosa* and consists of small clumps of cells that produce the mineralocorticoid aldosterone.

The *zona fasciculata* is the middle and thickest layer of the cortex and consists of cords of cells oriented radial to the center of the gland. The inner layer is comprised of interlaced strands of cells called the *zona ret icularis*. The zona fasciculata and zona reticularis both produce the physiologically important glucocorticoids, cortisol and corticosterone. These layers of the cortex also produce the androgen *dehydroepiandrosterone*, which is related chemically to the male sex hormone *testosterone*.

The cells of the adrenal cortex contain abundant lipid droplets. This stored lipid is functionally significant because cholesterol esters present in the droplets are an important source of the cholesterol used as a precursor for the synthesis of steroid hormones.

The adrenal medulla can be considered a modified sympathetic ganglion. The medulla consists of clumps and strands of *chromaffin cells* interspersed with

venous sinuses. Chromaffin cells, like the modified postganglionic neurons that receive sympathetic preganglionic cholinergic innervation from the splanchnic nerves, produce catecholamine hormones, principally epinephrine and norepinephrine.

II. The neuroendocrine control of the hypothalamic-adrenal-adrenal axis

The primary hormonal stimulus in the activation of the adrenal response to stress is corticotrophin-releasing hormone (CRH). This neurohormone is synthesized by specific neurons within the paraventricular nucleus of the hypothalamus. One subset of CRH-containing axons terminates in the median eminence, and CRH is released into the long portal vessels to reach the anterior pituitary and corticotrophic cells. The other subset of CRH-containing axons terminates in the lower brainstem and spinal cord.

Some CRH neurons also synthesize vasopressin (VP), which is compacted with CRH in the secreted granules, and released together with CRH. Vasopressin is known to maximize potential effects of CRH on the corticotrophic cells.

1. Control of CRH secretion

The control of CRH release is multifactored.

The primary stimulus to the hypothalamic-pituitary adrenal axis is stress. In 1936, a Canadian physiologist, Hans Selye, discovered that injections of cattle ovary extract into rats caused: 1) stimulation of growth to the adrenal cortex; 2) atrophy of lymphoid tissue in the spleen, lymph nodes and thymus, and 3) produced bleeding peptic ulcers.

In the response to stress, there are *three* stages:
1) the *alarm reaction* when the adrenal glands are activated;
2) the *stage of resistance* in which readjustment occurs;
3) the *stage of exhaustion*, which may lead to sickness and possible death.
Selye's concept of stress has been refined by subsequent researches.

Being the primary stimulus to the hypothalamic-pituitary adrenal axis, stress. activates CRH (and in certain circumstances) VP release. Adrenal axis responses to stress are blocked by appropriately placed hypothalamic lesions, anaesthetic agents and large doses of tranquillisers. A wide variety of stresses

are effective, such as for example emotional (anxiety, pain, fear, frustration), traumatic (injury, burn) or chemical (hypoglycemia, histamine).

1. Open-loop control elements

- *Diurnal rhythm*: CRH (and therefore ACTH and cortisol) is released in an episodic fashion, i.e., in highly irregular "bursts" of secretions;
- *Circadian rhythm*: early morning secretion rate is greater than that in the evening. Plasma levels of CRH and ACTH increase sharply after 4–5 hours of sleep, reach a peak at one hour after awakening and decline irregularly throughout the day to reach minimal values during the first few hours of sleep. Plasma cortisol levels vary accordingly (Fig. 11–23).

If the subject reverses her/his sleeping-waking habits, the ACTH-cortisol diurnal rhythm reverses, but only after several weeks. Traveling long distances rapidly through many longitudes (east-west) places the traveler's ACTH-cortisol secretion rhythm out-of-phase with the new sleeping-waking cycle; this may contribute to the "jet-lag" syndrome.

The presence of circadian rhythms has several important clinical implications:

1) Blood samples for plasma determinations must be taken at a standard time of the day (usually 9–10 a.m.); an afternoon value can be one-half that of a morning sample in the normal subject;

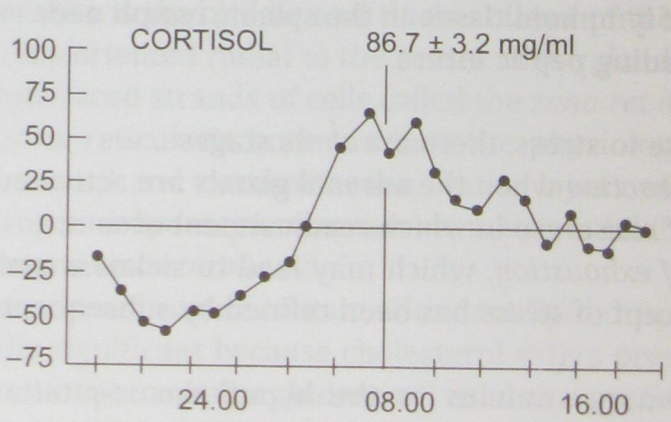


Fig. 11–23. Circadian rhythm in plasma concentrations of cortisol (from G. C. Lachelin et al. // J. Clin. Endocrine Metab. 49 : 892, 1979).

2) The diurnal rhythm is not present in patients with Cushing's syndrome and is therefore a diagnostic aid. Under these circumstances, one may wish to take a late evening blood sample.

2. Closed loop feedback element

- Long loop glucocorticoid negative feedback

Only the glucocorticoids participate in the negative feedback loop, and of the physiological steroids that are secreted; only cortisol has a significant effect. In the absence of cortisol, such as after adrenalectomy, there is enhanced secretion of CRH.

2. Actions of CRH

CRH activates the sympathetic nervous system and induces the behavioural responses characteristic of stress (see section on the sympathetic system);

CRH activates the corticotrope in the anterior pituitary to release ACTH and other POMC related peptides. Subsequently, it activates proopiomelanocortin (POMC; see below) gene transcription and increases POMC mRNA biosynthesis and concentration.

III. Adrenocorticotropic hormone (ACTH)

ACTH is a 39 amino acids peptide. The 1–20 sequence provides full physiologic activity. Variations in sequence among species occur only in the region of residues 25–32. The amino acids sequence 1–13 contacts α — and β — MSH and accounts for the intrinsic melanocyte-stimulating activity of ACTH. Synthetic ACTH is available for clinical use.

ACTH is synthesized as part of a large polypeptide pro-hormone called POMC. POMC undergoes extensive posttranslational processing; multiple peptides are produced from POMC by enzymatic cleavage. The cleavage pattern is tissue specific. In the anterior pituitary, POMC is cleaved to yield ACTH and β-lipotropin. A portion of β-lipotropin is cleaved further to yield γ-lipotropin and β-endorphin (Fig. 11–24).

1. Actions of ACTH

The main role of ACTH is to stimulate the adrenal glands endocrinal activity. Adrenals are composed of two distinct endocrine tissues which have separate embryologic origins.

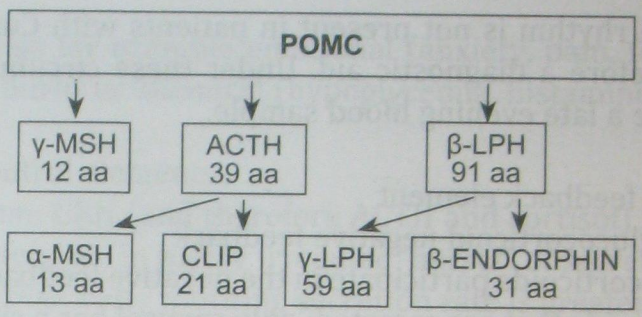


Fig. 11–24. Scheme of chemical structure of POMC and related molecules.

ACTH exerts both trophic and secretory effects:

1. Trophic Effects

ACTH increases the size and weight of the adrenal gland. The growth effect requires hours to days to be induced. This effect is due to a hypertrophy and hyperplasia mainly of the cells of the zona fasciculate and zona reticularis; the cells of the zona glomerulosa are affected to a much lesser extent In the hypoxic mammal, atrophy and hypo function of the adrenal cortex occur, the least amount of atrophy occurs in the zona glomerulosa which correlates with its ability to continue to secrete aldosterone in this circumstance. Adequate secretion of aldosterone can persist during years of ACTH deficiency.

2. Secretory Effects

The primary action of ACTH on the adrenal cortex is to increase cortisol secretion by increasing its synthesis. Cortisol storage in the adrenal is minimal. Androgen synthesis and secretion is also enhanced. These effects occur within minutes. ACTH acts by binding to specific receptors at the cell surface. About 3600 sites of the single class high affinity receptors are present on each adrenocortical cell for ACTH. The binding of ACTH promotes adenylate cyclase activation, which increases cyclic AMP (intracellular signal). cAMP activates protein kinase A and phosphorylation of a number of proteins.

Steroid synthesis induced by ACTH proceeds from cholesterol esters. The cholesterol that has been removed from the lipid droplets for steroid hormone biosynthesis is replenished in two ways (Fig. 11–25):

- increases cholesterol availability by activating cholesterol esterase, which release cholesterol from cholesterol esters depots in the cytoplasm of adrenocortical cells and by facilitating the uptake of low-density lipoproteins (LDL), which carry cholesterol, from plasma. A 400-kDa protein molecule called *apoprotein* B_{100} is also present on the surface of the LDL

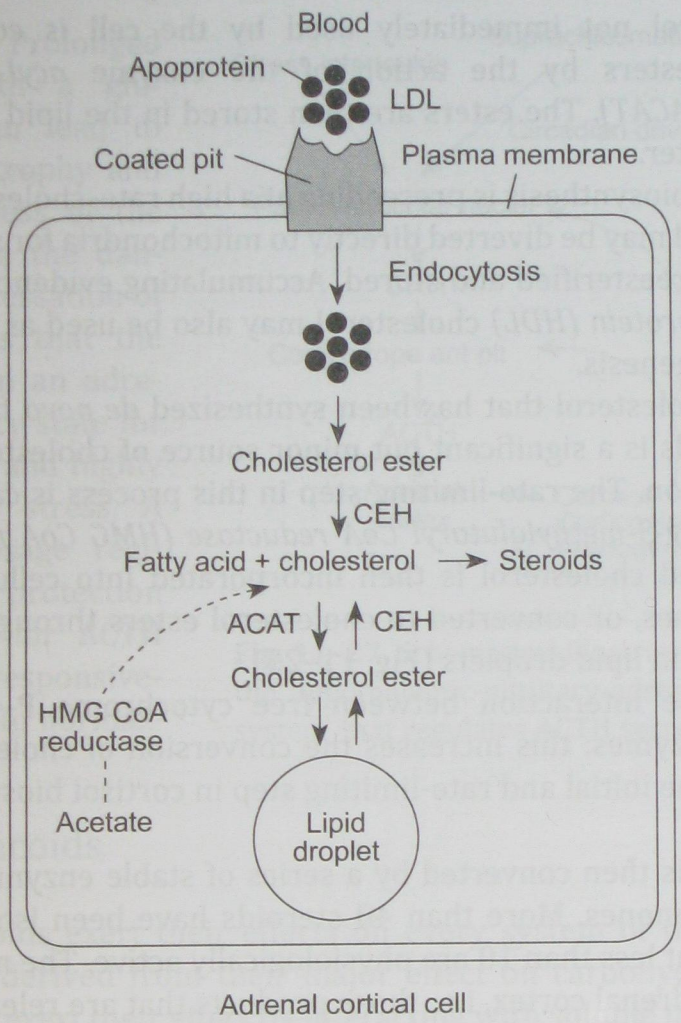


Fig. 11–25. Sources of cholesterol for steroid biosynthesis by the adrenal cortex. Most cholesterol comes from low-density lipoprotein (LDL) particles in the blood, which bind to receptors in the plasma membrane and are taken up by endocytosis. The cholesterol in the LDL particle is used directly for steroidogenesis or stored in lipid droplets for later use. Some cholesterol is synthesized directly from acetate. CEH, cholesterol ester hydrolase; ACAT, acyl-CoA:cholesterol acyltransferase; HMG, 3-hydroxy-3-methylglutaryl.

particle; it is recognized by LDL receptors localized to coated pits on the plasma membrane of adrenal cortical cells (Fig. 11–25). Hence, one of the actions of ACTH is to cause a rapid fall in the adrenal content of cholesterol, by mobilizing cholesterol from the adrenal lipid droplets for steroid genesis.

Any cholesterol not immediately used by the cell is converted again to cholesterol esters by the action of the enzyme *acyl-CoA:cholesterol acyltransferase (ACAT).* The esters are then stored in the lipid droplets of the cell to be used later.

When steroid biosynthesis is proceeding at a high rate, cholesterol delivered to the adrenal cell may be diverted directly to mitochondria for steroid production rather than reesterified and stored. Accumulating evidence suggests that *high-density lipoprotein (HDL)* cholesterol may also be used as a substrate for adrenal steroidogenesis.

In humans, cholesterol that has been synthesized *de novo* from acetate by the adrenal glands is a significant but minor source of cholesterol for steroid hormone formation. The rate-limiting step in this process is catalyzed by the enzyme *3-hydroxy-3-methylglutaryl CoA reductase (HMG CoA reductase*). The newly synthesized cholesterol is then incorporated into cellular structures, such as membranes, or converted to cholesterol esters through the action of ACAT and stored in lipid droplets (Fig. 11–25).

- activates the interaction between free cytochrome P-450 side-chain cleavage enzymes: this increases the conversion of cholesterol to pregnenolone, the initial and rate-limiting step in cortisol biosynthesis.

Pregnenolone is then converted by a series of stable enzymes into the adrenal steroid hormones. More than 40 steroids have been isolated from the adrenal cortex, but less than 10 are physiologically active. The major secretory products of the adrenal cortex, i.e., those products that are released in physiologically significant amounts into the peripheral circulation are: (1) cortisol; (2) corticosterone; (3) aldosterone; and (4) dehydroepiandrosterone (DHEA) and androstenedione, two adrenal androgens.

1. Control of ACTH secretion

Net ACTH release is dictated by the relative strength of the stimulatory signal (mainly CRH) and of the inhibitory signal (the negative feedback loop). Only the glucocorticoids participate in the negative feedback control, and, of these, only *cortisol* has a significant effect. The site of feedback is dual: on the pituitary (cortisol suppresses transcription of the POMC gene in the corticotrope), and in the hypothalamus (cortisol decreases synthesis of CRH) (Fig. 11–27).

It is important to note that the active part in the negative feedback loop is cortisol free. In plasma, about 90 % of cortisol is bounded, mainly to corticosteroid binding protein (CBG), synthesized in the liver. CBG acts as a reservoir for cortisol, protecting it from clearance and degradation.

Important! Prolonged treatment with a glucocorticoid can lead to pronounced atrophy and unresponsiveness of the adrenal cortex; the danger of abrupt cessation of therapy means that the individual is in an adrenal insufficiency state for several weeks and highly susceptible to stress. A decreasing dosage regimen provides protection and the time for ACTH and adrenal responsiveness to return to normal.

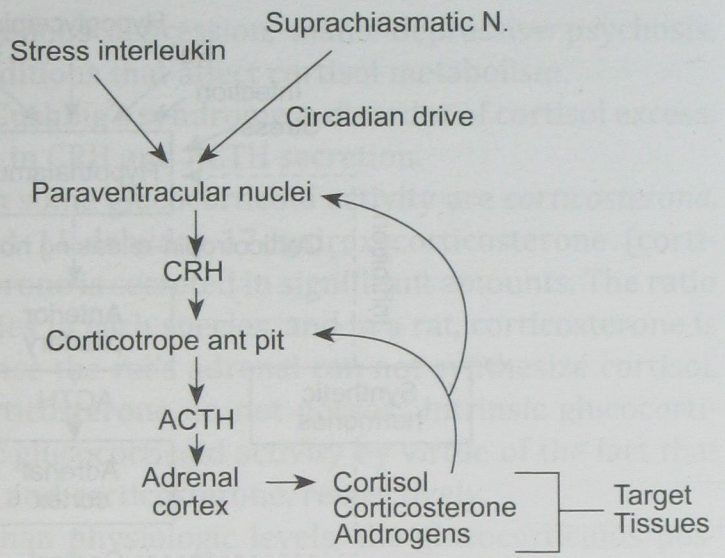


Fig. 11–27. Schematical illustration of the hypothalamic-pituitary-adrenal feedback system that regulates ACTH secretion.

V. Glucocorticoids

Glucocorticoids exert their effects on every system of the body, although their name is derived from their major effect on carbohydrate metabolism. These steroids exert their effect by interacting with soluble intracellular receptor proteins.

The glucocorticoid hormones, mainly cortisol, are synthesized in the zona fasciculata and the zona reticularis of the adrenal gland. The blood levels of these hormones are regulated by negative feedback mechanisms of the hypothalamic-pituitary-adrenal system (Fig. 11–28). The quantitatively and biologically most important glucocorticoid secreted in a human is *cortisol*.

Cortisol levels increase as ACTH levels rise and decrease as ACTH levels fall. There is considerable diurnal variation in ACTH levels, which reach their peak in the early morning (around 6 to 8 a.m. and decline as the day progresses (Fig. 11–29).

This appears to be attributable to rhythmic activity in the CNS, which causes bursts of corticotropin-releasing hormone (CRH) secretion and, in turn, ACTH secretion. This diurnal pattern is reversed in people who work during the night and sleep during the day. The rhythm also may be changed by physical and

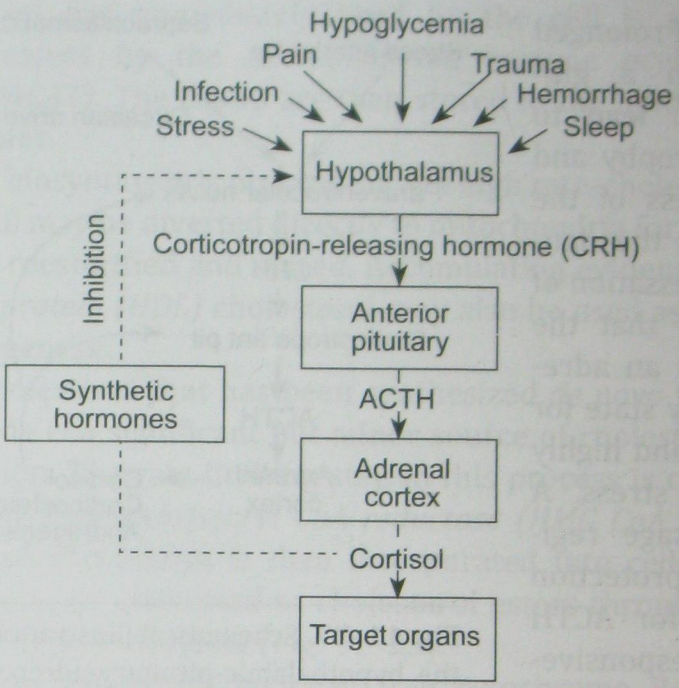


Fig. 11–28. The hypothalamic-pituitary-adrenal (HPA) feedback system that regulates glucocorticoid (cortisol) levels.

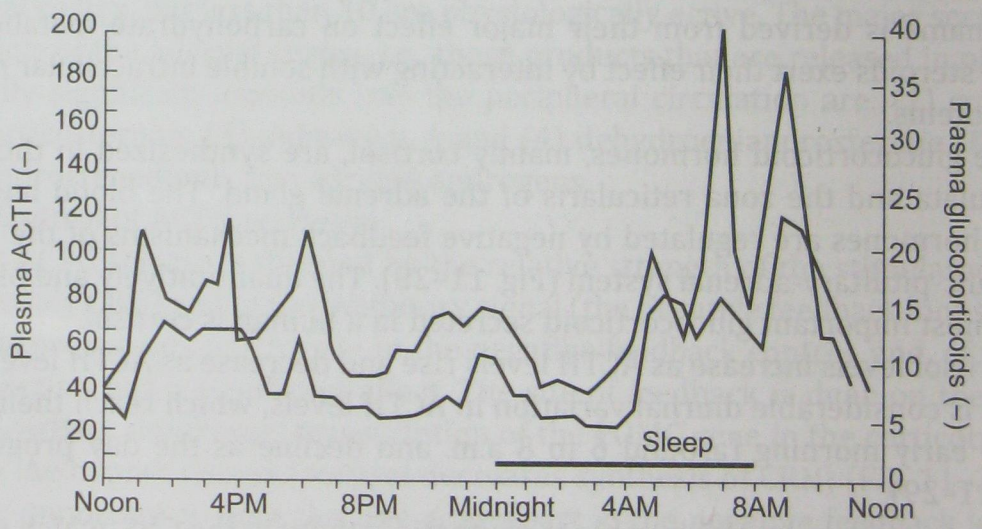


Fig. 11–29. Diurnal changes of the ACTH and glucocorticoids plasma levels throughout the 24 hrs period.

psychological stresses, endogenous depression, manic-depressive psychosis, and liver disease or other conditions that affect cortisol metabolism.

One of the earliest signs of Cushing's syndrome, a disorder of cortisol excess, is the loss of diurnal variation in CRH and ACTH secretion.

Other adrenal steroids with some glucocorticoid activity are *corticosterone*, 11-dehydrocorticosterone and 11-dehydro-17-hydroxycorticosterone (cortisone); of these, only corticosterone is secreted in significant amounts. The ratio of cortisol: corticosterone varies in each species, and in a rat, corticosterone is a major secretory product, since the rat's adrenal can not synthesize cortisol. Cortisone and 11-dehydrocorticosterone do not possess intrinsic glucocorticoid activity but have biologic glucocorticoid activity by virtue of the fact that they are converted to cortisol and corticosterone, respectively.

When present at greater than physiologic levels, the glucocorticoids possess anti-inflammatory activity, and the anti-inflammatory potency parallels the glucocorticoid potency.

Actions of cortisol

Cortisol is one of the few hormones essential to life. For example, although carbohydrates and pure mineralocorticoids or NaCl can postpone death in the adrenalectomized patient, human beings cannot survive total adrenalectomy for long without glucocorticoid replacement. This is not the result specifically of one action of cortisol, but probably the overall result of a constellation of action, which cortisol exerts.

1. ***Effects on metabolism***

A. Carbohydrate metabolism

Cortisol is a hyperglycemic agent. Hypoglycemia is a characteristic finding in the adrenalectomized animal and in the patient with adrenal insufficiency. The hypoglycemia following adrenalectomy is accompanied by a reduction in liver glycogen, particularly if food is restricted. The administration of cortisol tends to increase plasma glucose; the latter is difficult to demonstrate in the normal human, however, because of the high insulin reserve, but can be seen readily where insulin reserve is limited. Adrenalectomy reduces hyperglycemia in diabetic laboratory animals. Glucocorticoid treatment restores the hyperglycemia of adrenalectomized or animals with adrenal insufficiency. As in laboratory animals, the administration of high doses of glucocorticoid hormone to the human can induce temporary diabetes which is highly insulin resistant. Administration of a glucocorticoid prior to the glucose tolerance test will require

a greater insulin release to return the plasma glucose concentration to normal; this has been used as a screening test for the detection of "prediabetic subjects".

The hyperglycemic effect of cortisol results from several actions on carbohydrates metabolism:

- Cortisol decreases the uptake and phosphorylation of glucose by peripheral tissues. This action is similar to that of GH, but opposite to that of insulin.
- *Cortisol increases liver gluconeogenesis.* This effect represents the major action of glucocorticoids on carbohydrates metabolism. The ability of glucocorticoids to increase liver gluconeogenesis depends upon a constellation of specific actions, but rests primarily upon their ability to induce a net protein catabolism and to provide the amino acid precursors to the liver.

Thus, liver gluconeogenesis is increased by cortisol for the following reasons

1. Cortisol inhibits protein anabolism and accelerates protein degradation. The resultant net protein catabolism provides an increase in the circulating glycogenic amino acids which are available to the liver.
2. Cortisol directly activates key hepatic gluconeogenic enzymes, such as transaminases, glucose-6-phosphatase and phosphoenal pyruvate carboxykinase (PEPCK). It is because of gluconeogenesis that glucocorticoids can simultaneously increase liver glycogen and liver glucose output. The effectiveness and importance of limited food intake are indicated by the finding that cortisol can increase liver glycogen content in the *fasting* animal; the increase in liver glycogen is accompanied by an increase in urinary NH_3 (increased net protein catabolism) which quantitatively parallels the increase in liver glycogen and blood glucose.
3. Cortisol increases liver glycogen synthetase to account for its "liver glycogen deposition effect". Greatly enlarged, glycogen-packed livers may be found in patients with diseases of cortisol excess.

The glucocorticoids' multiple and crucial stimulatory influences on gluconeogenesis explain the finding that glucocorticoid hormones have a "permissive action" on gluconeogenesis, i.e., some glucocorticoid must be present to permit the full expression of the gluconeogenic action of other hormones (epinephrine, glucagons and GH). Note that all of the gluconeogenic hormones

(epinephrine, glucagons, GH and cortisol) are released by decreased plasma glucose and stress, conditions which accompany fasting and starvation; appropriately, insulin release would be reduced. Insulin inhibits gluconeogenesis by rapidly inhibiting the synthesis of PEPCK and by promoting protein anabolism. Thus, *starvation is characterized by increased gluconeogenesis, glycogenolysis, lipolysis, and keto-acid formation.*

B. Protein Metabolism

Glucocorticoids inhibit (depress) protein anabolism in many tissues (e.g. lymphoid, muscular, connective tissues). The result is a *net protein catabolism* and negative nitrogen balance. The released amino acids provide amino acid precursors for gluconeogenesis. Normally, cortisol is balanced with anabolic hormones, but cortisol maybe present in excess (as in Cushing's syndrome), or if normal levels become unopposed by loss of anabolic hormones (loss of estrogen during menopause), the net protein catabolism is readily evident. The net protein catabolism is reflected outwardly by the muscular wasting and weakness, by the loss of protein connective tissue fibers from the skin and subcutaneous areas to produce the striae (stretch marks) and by the loss of bone in the Cushing's syndrome.

C. Fat Metabolism

Glucocorticoids at high dosages promote lipolysis and at low dosages have a permissive effect for the lipolytic action of other hormones. Like GH, cortisol requires intact and functioning protein synthesis machinery for its lipolytic action and may be involved in the synthesis of the hormone-sensitive lipase. If insulin release is not adequate, cortisol tends to produce ketosis. When insulin release is adequate, the lipolytic and ketosis-inducing action may not be evident. In Cushing's syndrome, there is a characteristic deposition of fat in the face (moon face), in the suprascapular area (buffalo hump) and on the trunk but not in the extremities. The moon face, due to fat deposition and some salt and water retention, is a clinical sign of glucocorticoid over dosage.

2. ***Effects on the inflammatory process and on immunological function***

1. Glucorticoids, at supra-physiological doses, inhibit the *normal inflammatory process* in response to tissue trauma, chemical irritants and infection. They inhibit the movement of cells and fluid from the intravascular compartment that characterizes the basal inflammatory response. Possible mechanisms for the effect reflect the following phenomena:

- The inhibition of prostaglandin synthesis.
- The inhibition of plasminogen activators, serine proteases that convert plasminogen into plasmin. Plasmin cleaves the plasma globulin, kininogen, into potent kinins that cause vasodilatation and increased capillary permeability.
- The inhibition of fibroblasts, thereby preventing fibrotic isolation of the infected area. The granuloma formation in response to a foreign body or bacteria is greatly inhibited by glucocorticoids.

Glucocorticoid therapy reduces the local undesirable effects (edema, pain, etc.) but weakens the normal bodily defense mechanisms for the isolation and/or destruction of the causative factor. Thus, the spread and growth of the infective agent will be promoted, and concomitant antibacterial therapy may be indicated.

The glucocorticoids possess anti-allergenic actions. They depress allergic reactions probably by interfering with the intracellular production or release of histamine or histamine-like substances by the sensitized cells.

2. Glucorticoids, at supra-physiological doses, suppress the *immunological response*. Within 4 hrs of administration, there is a decrease in peripheral lymphocyte numbers, mostly T-cell (thymus-derived). This results from the redistribution of lymphocytes from the intravascular ccompartment to the spleen, lymph nodes, thoracic duct and bone marrow. A similar redistribution of monocytes occurs. Granulocytes, on the contrary, leave the bone marrow and enter circulating blood, but glucocorticoids inhibit the accumulation of neutrophils at inflammatory sites. Thus, clinically, there is an increase in circulating neutrophils, but a suppression of local inflammatory responses.

T-cell function

Lymphocytes are particular targets for glucocorticoid action, and their immunological function may be affected: in vitro, glucocorticoids inhibit proliferation. In rodents, glucocorticoids induce lysis of T-cells and thymocytes. Mature human T-cells appear to be less sensitive, although, in Cushing's syndrome, there may be atrophy of the thymus, spleen and lymph nodes. Overall, the loss of lymphoid elements decreases the bodily immune defense mechanisms.

B-cell function

Glucocorticoids modulation of B-cell function may be both direct and mediated by effects on T-cell subpopulations. Early events, such as B-cell

activation and proliferation in response to T-cell derived growth factors (interleultin-4) are inhibited by glucocorticoids. Terminal B-cell differentiation is less sensitive to glucocorticoids, and thus production of immunoglobulin is not influenced in a major way.

Monocyte and macrophage function

Glucocorticoids inhibit the proliferative response of monocytes to colony-stimulating factor and their differentiation into macrophages. They also inhibit the phagocytic and cytotoxic functions of macrophages.

3. ***Effects on fluid and electrolyte homeostasis***

Cortisol has weak mineralocorticoid activity and slight Na^+ retaining properties. Patients with glucocorticoid excess usually have hypertension, which often occurs without elevation of plasma mineralocorticoids or evidence of functional mineralocorticoids excess (hypokalemia, rennin activity suppression). Whether this reflects entirely the mineralocorticoid activity on glucocorticoids is not known.

Additional mechanisms may include the effect that glucocorticoids exert on maintaining the responsiveness of vascular smooth muscle to the presser amines, epinephrine and norepinephrine, to ensure adequate perfusion of tissues. After adrenalectomy, the arterioles become unresponsive to epinephrine and mechanical stimulation, and there is dilatation of the capillary bed with increased permeability to colloidal dyes. The administration of a glucocorticoid returns the vascular response to epinephrine to normal and augments vasomotion (*permissive effect*). Enhanced vascular sensitivity to the presser effects of norepinephrine and angiotensin II is seen in Cushing's syndrome. Patients with glucocorticoid deficiency may be hypertensive.

Glucocorticoids maintain the renal tubular and collecting duct membranes impermeable to water in the absence of ADH; thus, the glucocorticoids are necessary to demonstrate a water diuresis (*permissive effect*). In the absence of glucocorticoids, the renal membranes become permeable to water, back diffusion of H_2O occurs and there develops a defect in the handling of a water load, i.e., it takes a longer time to excrete a water load than in a normal subject; water intoxication may occur. This was formerly used as a test (water load test) for adrenal insufficiency. Only a glucocorticoid hormone can restore the handling of a water load to normal. Since glucocorticoids render the membranes less permeable to water, they have been shown in dogs and humans to increase free-water clearance.

In summary, the glucocorticoids protect the organism against emotional and traumatic stress. By making energy available (lipolysis, glycogen stores, gluconeogenesis, glucose for the CNS) in times of great stress or need, the glucocorticoids help the organism to survive. They also maintain vascular reactivity to support perfusion pressure to tissues in times of stress. In the absence of glucocorticoids, circulatory collapse (decreased blood pressure) and extreme muscular fatigue occur in response to a severe stress. All cortisol effects are summarized in the Table 11–2.

Table 11–2. Cortisol effects in an organism

Major Influence	Effect on Body
Glucose metabolism	Stimulates gluconeogenesis Decreases glucose use by the tissues
Protein metabolism	Increases breakdown of proteins Increases plasma protein levels
Fat metabolism	Increases mobilization of fatty acids Increases use of fatty acids
Anti-inflammatory action (pharmacologic levels)	Stabilizes lysosomal membranes of the inflammatory cells, preventing the release of inflammatory mediators Decreases capillary permeability to prevent inflammatory edema Depresses phagocytosis by white blood cells to reduce the release of inflammatorymediators Suppresses the immune response Causes atrophy of lymphoid tissue Decreases eosinophils content in the blood Decreases antibody formation Decreases the development of cell-mediated immunity Reduces fever Inhibits fibroblast activity
Psychic effect	May contribute to emotional instability
Permissive effect	Facilitates the response of the tissues to humoral and neural influences, such as that of the catecholamines, during trauma and extreme stress

VI. THE MINERALOCORTICOIDS

The major mineralocorticoid is aldosterone, which is synthesized exclusively in the zona glomerulosa. The mineralocorticoids play an essential role in regulating potassium and sodium levels and water balance. Its precursors, 18-OH corticosterone, corticosterone and DOC have also weak mineralocorticoid activities. [Cortisol also has a modest mineralocorticoid activity, which may become significant during the hypercortisolemia of Cushing's syndrome].

Actions of aldosterone

1. Stimulates reabsorption of Na^+ from the distal renal tubule. As a result, net urinary Na^+ excretion is reduced, and this vital ion is conserved. Because water is passively reabsorbed with Na^+, extracellular fluid expands in an isotonic fashion.
2. Stimulates active secretion of K^+ from the distal tubular cell. Aldosterone increases the permeability of the renal luminal membrane to K^+ excretion. Thus, the effect of aldosterone is to increase the Na^+/K^+ ratio in plasma.
3. Enhances renal tubular excretion of H^+.

Regulation of aldosterone secretion

Aldosterone secretion is governed by multiple factors with complex regulatory interactions. Renin and its biochemical derivative angiotensin II is the important physiological regulator of aldosterone secretion, not ACTH. Other factors, such as an increase in serum potassium, can also stimulate aldosterone secretion, but normally, they play only a secondary role. Therefore, major regulators are the following:

1. The renin-angiotensin system

Renin is synthesized by the glomerular zone of the renal cortex. Its secretion is controlled by renal arteriolar pressure, Na^+ concentration of tubular fluid, and renal sympathetic activity. Plasma renin levels increase with factors that decrease renal blood levels (hemorrhage, dehydration, salt restriction, upward posture). Renin is an enzyme that releases angiotensinogen (originating from the liver) to produce angiotensin I (Fig. 11–30). The latter is cleaved in the lungs and other tissues to yield angiotensin II and angiotensin III.

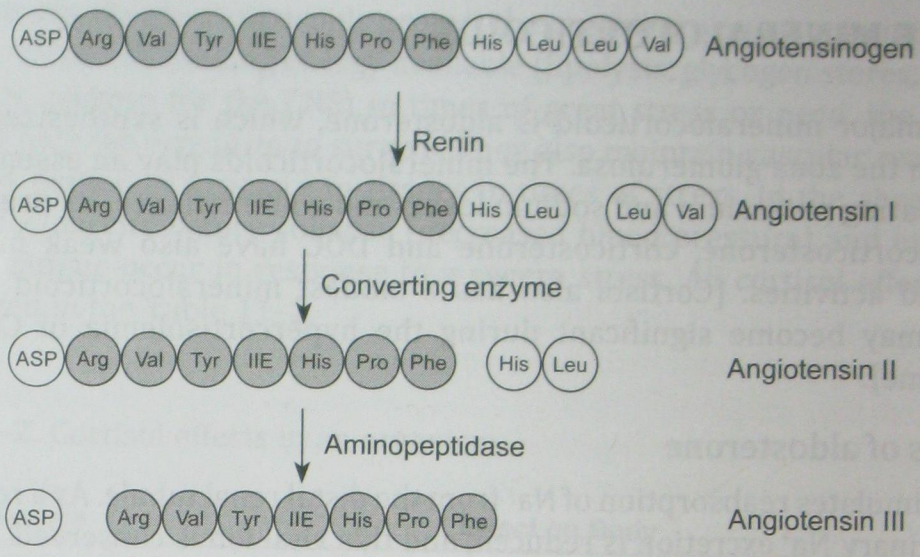


Fig. 11–30. Renin-catalyzed cascade of biochemical reactions aimed to angiotensin I, II and III formation from angiotensinogen.

Actions of the latter compounds on the adrenal glomerulosa are mediated by high-affinity cell surface receptors. The primary intracellular signal transduction mechanism is the activation of phospholipase C. Several events, including the influx of extracellular Ca^{2+}, then lead to the stimulation of conversion of cholesterol into pregnenolone (in a mechanism similar to that of ACTH in the zona fasciculata) and conversion of corticosterone to aldosterone. Angiotensin II does not activate adenylate cyclase.

The primary mechanism is shown in Fig. 11–31. The stimulation of aldosterone synthesis is initiated when angiotensin II binds to its receptors on the plasma membranes of zona glomerulosa cells. The signal generated by the interaction of angiotensin II with its receptors is transmitted to phospholipase C (PLC) by a G-protein, and the enzyme becomes activated. The PLC then hydrolyzes phosphatidylinositol 4,5 bisphosphate (PIP2) in the plasma membrane, p roducing the intracellular second messengers inositol trisphosphate (IP3) and diacylglycerol (DAG). The IP3 mobilizes calcium, which is bound to intracellular structures, increasing the calcium concentration in the cytosol.

This increase in intracellular calcium and DAG activates protein kinase C (PKC). The rise in intracellular calcium also activates calmodulin-dependent protein kinase (CMK). These enzymes phosphorylate proteins, which then become involved in initiating steroidogenesis.

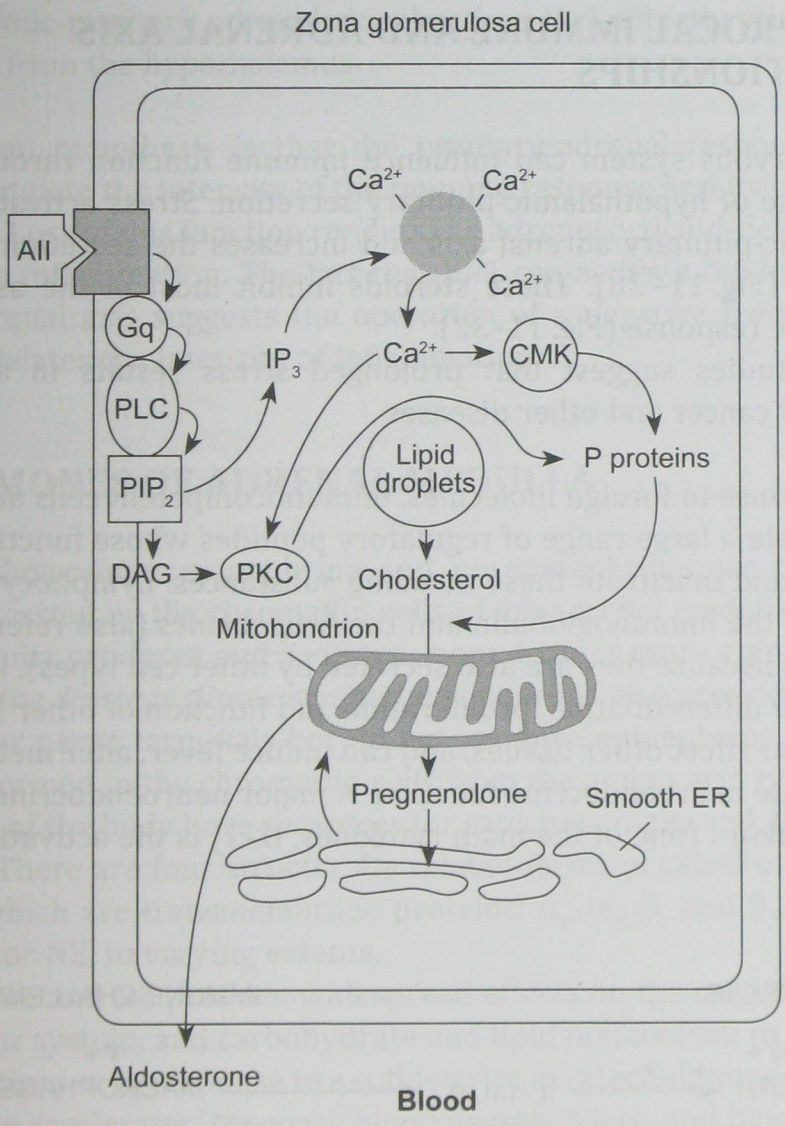


Fig. 11–31. The influence of angiotensin II on aldosterone synthesis.

2. Potassium

There is a reciprocal relation between serum K^+ and aldosterone concentrations. K^+ directly increases aldosterone secretion by the adrenal cortex and aldosterone lowers serum K^+ by stimulating K^+ excretion by the kidney.

3. Minor modulators

ACTH, vasopressin, dopamine and others.

VII. RECIPROCAL IMMUNE AND ADRENAL AXIS RELATIONSHIPS

1. The nervous system can influence immune function through the classic route of hypothalamic-pituitary secretion. Stress activates the hypothalamic-pituitary-adrenal axis and increases the secretion of glucocorticoids (Fig. 11–28). These steroids inhibit most of the aspects of the immune response (Fig. 11–32).

Indeed, studies suggest that prolonged stress results in an increased incidences of cancer and other diseases .

2. In response to foreign molecules, immunocompetent cells are stimulated to secrete a large range of regulatory peptides whose function is to neutralize and inactivate these invading substances. Lymphocyte secretions include the immunoglobulin and the *lymphokines* (also referred to as *cytokines*, because they are also secreted by other cell types). Lymphokines regulate differentiation, proliferation and function of other lymphocytes; they also affect other tissues, and can induce fever, alter metabolism, and modulate neuroendocrine function. A major neuroendocrine response to interleukin-l (one of the main cytokines, IL-1) is the activation of the hy-

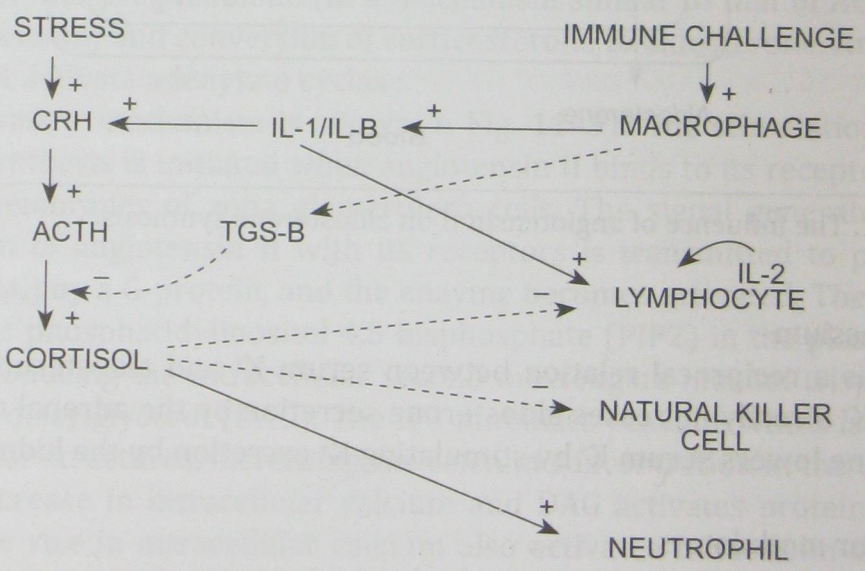


Fig. 11–32. Schematical illustration of glucocorticoids immune effects modulation.

pothalamic-pituitary-adrenal axis, because IL-1 directly stimulates CRH release from the hypothalamus.

The present hypothesis is that the pituitary-adrenal response to stress serves to modulate the intensity of the immune response and its inflammatory components. Loss of this function renders the adrenocorticoid-deficient animal vulnerable to inflammation. The fact that IL-1 can activate the hypothalamic-pituitary adrenal axis suggests the operation of a negative feedback control loop that regulates the intensity of inflammation.

VIII. HORMONES OF ADRENAL MEDULLA

The catecholamines, epinephrine and norepinephrine, are the two hormones synthesized by the chromaffin cells of the adrenal medulla. The human adrenal medulla produces and secretes about 4 times more epinephrine than norepinephrine. Postganglionic sympathetic neurons also produce and release NE from their nerve terminals but do not produce epinephrine. Epinephrine and NE are formed in the chromaffin cells from the amino acid tyrosine.

Most cells of the body have receptors for catecholamines and, thus, are their target cells. There are four structurally related forms of catecholamine receptors, all of which are transmembrane proteins: α_1, α_2, β_1 and β_2. All can bind epinephrine or NE, to varying extents.

Epinephrine and NE produce widespread effects on the cardiovascular system, muscular system, and carbohydrate and lipid metabolism in liver, muscle, and adipose tissues. In response to a sudden rise in catecholamines in the blood, the heart rate accelerates, coronary blood vessels dilate, and blood flow to the skeletal muscles is increased as a result of vasodilation (but vasoconstriction occurs in the skin). Smooth muscles in the airways of the lungs, gastrointestinal tract, and urinary bladder relax. Muscles in the hair follicles contract, causing piloerection. Blood glucose level also rises. This overall reaction to the sudden release of catecholamines is known as the *'fight-or-flight response'*.

Catecholamines secreted by the adrenal medulla and NE released from sympathetic postganglionic nerve terminals are key agents in the body's defense against hypoglycemia. Catecholamine release usually starts when the blood glucose concentration falls to the low end of the physiological range. A further decline in blood glucose concentration into the hypoglycemic range produces marked catecholamine release.

Catecholamines act on the liver to stimulate glucose production. They activate glycogen phosphorylase, resulting in the hydrolysis of stored glycogen, and stimulate gluconeogenesis from lactate and amino acids. Catecholamines also activate glycogen phosphorylase in skeletal muscle and adipose cells by interacting with β receptors, activating adenylyl cyclase and increasing camp in the cells.

In adipose cells, the rise in cAMP produced by catecholamines activates hormone-sensitive lipase, causing the hydrolysis of triglycerides and the release of fatty acids and glycerol into the bloodstream. These fatty acids provide an alternative substrate for energy metabolism in other tissues, primarily skeletal muscle, and block the Phosphorylation and metabolism of glucose.

During profound hypoglycemia, the rapid rise in blood catecholamine levels triggers some of the same metabolic adjustments that occur more slowly during fasting. During fasting, these adjustments are triggered mainly in response to the gradual rise in the ratio of glucagon to insulin in the blood. The ratio also rises during profound hypoglycemia, reinforcing the actions of the catecholamines on glycogenolysis, gluconeogenesis, and lipolysis.

Chapter 12. Physiology of male and female sexual glands

This chapter will review the main physiologic aspects of the male (I) and female (II) physiology.

I. PHYSIOLOGY OF MALE REPRODUCTIVE FUNCTION

The male genitourinary system is composed of the paired gonads, or testis, genital ducts, accessory organs, and penis. The dual function of the testis is to produce male sex androgens (male sex hormones), mainly testosterone, and spermatozoa (male germ cells). The internal accessory organs produce the fluid constituents of semen, and the ductile system aids in the storage and transport of spermatozoa. The penis functions in urine elimination and sexual function.

The testes have two primary functions, spermatogenesis, the process of producing mature sperm and steroidogenesis, the synthesis of testosterone.

1. Anatomic-structural overview of the male reproductive system

During embryonic stages of development the testes lie attached to the posterior abdominal wall. As the embryo elongates, the testes move to the inguinal ring. Between the seventh month of pregnancy and birth, the testes descend through the inguinal canal into the scrotum. The location of the testes in the scrotum is important for spermatogenesis which is optimal at 2 to 3 °C lower than core body temperature. Prolonged exposure of the testes to elevated temperature, fever or thermoregulatory dysfunction can lead to temporary or permanent sterility as a result of a failure of spermatogenesis, whereas steroidogenesis is unaltered.

Spermatogenesis refers to the sperm generation. It begins at an average age of 13 years and continues throughout the reproductive years of a man's life. Each human testis contains hundreds of tightly packed seminiferous tubules,

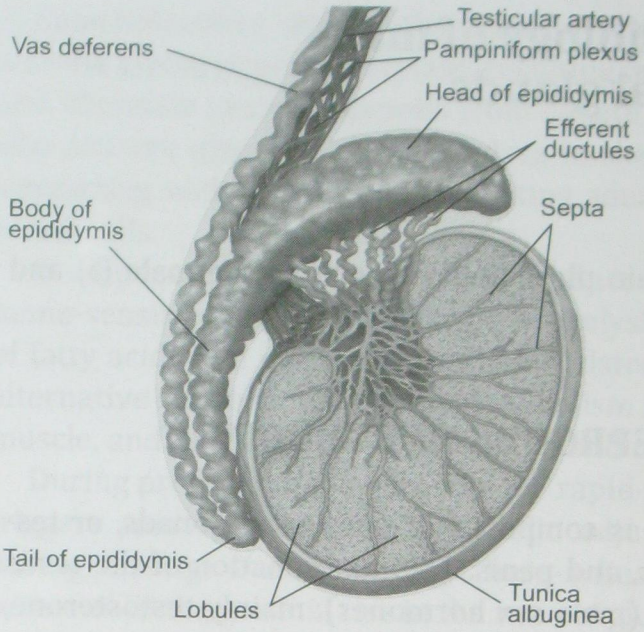


Fig. 12–1. The cross-section and structural parts of the testis and epididymis.

ranging from 150 to 250 μm in diameter and from 30 to 70 cm long. This is the place of spermatogenesis that occurs exactly in the seminiferous tubules of the testes (Fig. 12–1).

Internally, the testes are composed of several hundred compartments or lobules. Each lobule contains one or more coiled seminiferous tubules.

There are different testis morphological compartmentalization. Sperm production is carried out in the *avascular seminiferous tubules,* whereas testosterone is produced by the *Leydig cells*, which are scattered in a *vascular*, loose connective tissue between the seminiferous tubules in the interstitial compartment.

Each seminiferous tubule is composed of two somatic cell types (myoid cells and Sertoli cells) and germ cells. The seminiferous tubule is surrounded by a basement membrane (basal lamina) with myoid cells on its perimeter which define its outer limit. On the inside of the basement membrane are large, irregularly shaped *Sertoli cells*, which extend from the basement membrane to the lumen (Fig. 12–2).

Sertoli cells are attached to one another near their base by tight junctions (Fig. 12–3). The tight junctions divide each tubule into a *basal compartment*, whose constituents are exposed to circulating agents, and an *adluminal compartment*, which is isolated from bloodborne elements. The tight junctions limit the transport of fluid and macromolecules from the interstitial space into the tubular lumen, forming the *blood-testis barrier.*

Located between the nonproliferating Sertoli cells are germ cells at various stages of division and differentiation. Sertoli's cells secrete a special fluid that contains nutrients to bathe and nourish the immature germ cells; they provide digestive enzymes that play a role in spermiation (conyerting the spermatocytes to sperm); and they are thought to play a role in shaping the head and tail of

the sperm. Sertoli's cells also secrete several hormones, including the principal feminizing sex hormone, estradiol, which seems to be required in the male for spermatogenesis, and inhibin, which controls the function of Sertoli's cells through feedback inhibition of follicle-stimulating hormone (FSH) from the anterior pituitary gland.

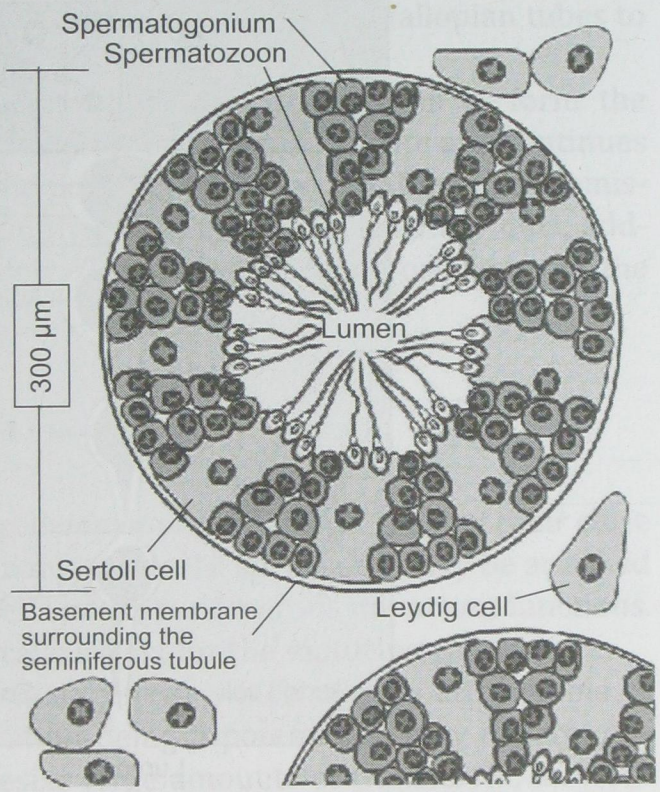


Fig. 12–2. The cross-sectional view of testis.

Mitosis of the *spermatogonia* (diploid progenitors of spermatozoa) occurs in the basal compartment of the seminiferous tubule (Fig. 12–3). The early meiotic cells (primary spermatocytes) move across the junctional complexes into the adluminal compartment, where they mature into *spermatozoa* or *gametes* after meiosis. The adluminal compartment is an immunologically privileged site. Spermatozoa that develop in the adluminal compartment are not recognized as "self" by the immune system. Consequently, males can develop antibodies against their own sperm, resulting in infertility. Sperm antibodies neutralize the ability of sperm to function. Sperm antibodies are often present after vasectomy or testicular injury and in some autoimmune diseases where the adluminal compartment is ruptured, allowing sperm to mingle with immune cells from the circulation.

After the spermatozoa are formed in the seminiferous tubules, they travel through the efferent tubules to the epididymis, which is the final site for sperm maturation.

Because the spermatozoa are not motile at this stage of development, peristaltic movements of the ductal walls of the epididymis aid in their movement. The spermatozoa continue their migration through the ductus deferens, also called the *vas deferens*. The ampulla of the vas deferens serves as a storage reservoir for sperm. Sperm are stored in the ampulla until they are released

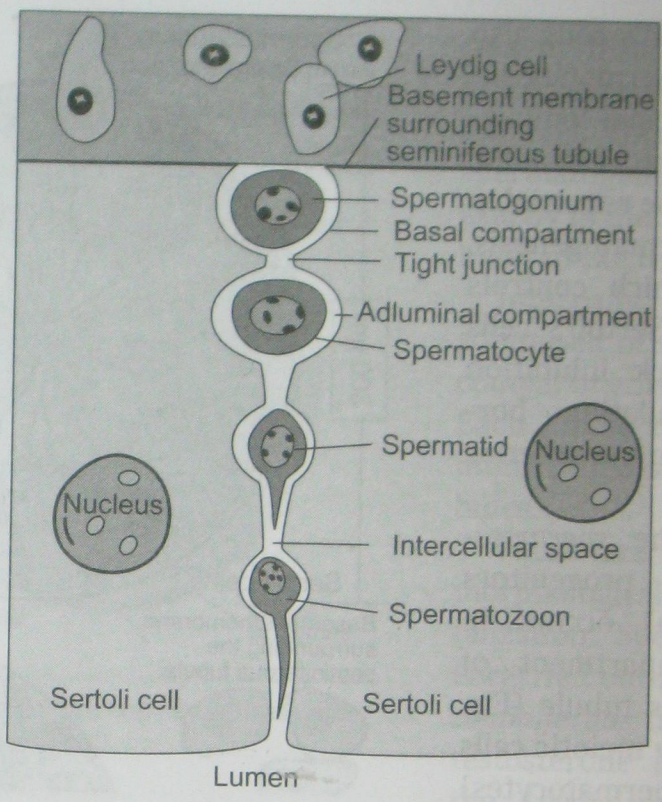


Fig. 12–3. Co-localization of the Sertoli and Leydig cells with the basal and adluminal compartments outlining.

through the penis during ejaculation. Spermatozoa can be stored in the genital ducts for as long as 42 days and still maintain their fertility. Surgical disconnection of the vas deferens in the scrotal area (*i.e.*, vasectomy) serves as an effective method of male contraception. Because sperm are stored in the ampulla, men can remain fertile for 4 to 5 weeks after performance of a vasectomy.

The seminal vesicles, the prostate gland, and the bulbourethral glands form the accessory reproductive structures of the male genitourinary system. The spermatozoa plus the secretions from the genital ducts and accessory organs make up the semen. The seminal vesicles consist of two highly tortuous tubes that secrete fluid for the semen. Each of the paired seminal vesicles is lined with secretory epithelium containing an abundance of fructose, prostaglandins, and several other proteins. The fructose secreted by the seminal vesicles provides the energy for sperm motility. The prostaglandins are thought to assist in fertilization by making the cervical mucus more receptive to sperm and

by causing reverse peristaltic contractions in the uterus and fallopian tubes to move the sperm toward the ovaries.

Each seminal vesicle joins its corresponding vas deferens to form the ejaculatory duct, which enters the posterior part of the prostate and continues through until it ends in the prostatic portion of the urethra. During the emission phase of coitus, each vesicle empties fluid into the ejaculatory duct, adding bulk to the semen. Approximately 70 % of the ejaculate originates in the seminal vesicles.

2. Sertoli cells

Sertoli cells are critical to germ cell development, as indicated by their close contact with germinative cells. As many as 6 to 12 spermatids may be attached to a Sertoli cell (Fig. 12–2). These cells perform numerous important functions.

1. Sertoli cells provide mechanical support for the maturing gametes.
2. They also have a role in spermatogenesis which may be attributable to their high glycogen concentration being a potential energy source. Additionally, Sertoli cells synthesize large amounts of transferrin, an iron-transport protein important for sperm development.
3. In the fetal testis, the Sertoli cells secrete H-Y-antigen, which is a product of the testis-organizing genes and is the cell surface glycoprotein responsible for the induction of testicular organogenesis. H-Y antigen has been found in all cell membranes from normal XY males except the cell membranes of immature germ cells.
4. The Sertoli cells secrete Müllerian duct inhibiting factor (MIF), the glycoprotein that causes regression of the Müllerian duct system.
5. With the onset of fetal testicular differentiation induced by H-Y antigen (which is under control of the Y-chromosome) and the incorporation of primitive germ cells into the seminiferous tubules, the Sertoli cells secrete a meiosis-inhibiting factor which suppresses germ cell proliferation and differentiation beyond the spermatogonial stage.
6. Sertoli cells participate in the release of spermatids by enveloping the residual lobules of spermatid cytoplasm. Thus, the residual bodies are not cast off into the lumen but are retained within the epithelium throughout the spermiation process.
7. Sertoli cells secrete a watery, K^+- and HCO_3^- rich fluid into the seminiferous lumen. These cells actively pump ions into the intercellular spaces

to create a standing osmotic gradient that moves water from the Sertoli cell base to the free surface of the lumen. This fluid movement provides a driving force for conveying sperm from the testis to the epididymis, where most of this isoosmotic fluid is reabsorbed.

8. Sertoli cells synthetize estradiol from androgenic precursors.
9. Sertoli cells also produce and secrete an androgen-binding protein (ABP) and inhibin. FSH stimulates the production of ABP and plasminogen activator, increases secretion of inhibin, and induces aromatase activity for the conversion of androgens to estrogens.

Inhibin suppresses FSH release from the pituitary gonadotrophs. The pituitary gonadotrophs and testicular Sertoli cells form a classical negative-feedback loop in which FSH stimulates inhibin secretion and inhibin suppresses FSH release. Inhibin also functions as a paracrine agent in the testis. Activin stimulates the release of FSH.

Shortly after birth, Sertoli cells cease proliferating, and throughout life, the number of sperm produced is directly related to the number of Sertoli cells. At puberty, the capacity of Sertoli cells to bind FSH and testosterone increases. Receptors for FSH, present only on the plasma membranes of Sertoli cells, are glycoproteins linked to adenyl-cyclase via G proteins (Fig. 12–4). FSH exerts

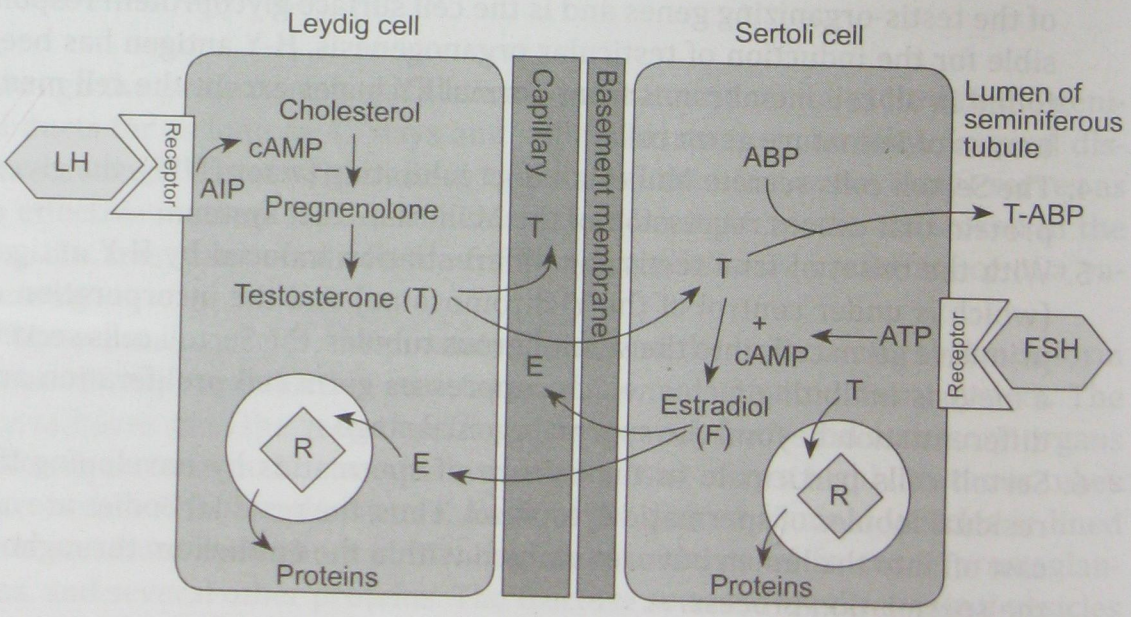


Fig. 12–4. Schematic illustration of Sertoli and Leydig cells interactions.

multiple effects on the Sertoli cell, most of which are mediated by cAMP and protein kinase A.

3. Leydig cells

Leydig cells are located between the seminiferous tubules. These cells also are called interstitial cells. They are equipped to produce steroids because they have numerous mitochondria, a prominent smooth endoplasmatic reticulum and conspicuous lipid droplets.

Leydig cells do not have FSH receptors (Fig. 12–4) but FSH can increase the number of developing Leydig cells by stimulating the production of growth stimulators from Sertoli cells that subsequently enhance the growth of the Leydig cells. Androgens stimulate the proliferation of developing Leydig cells. Estrogen receptors are present on Leydig cells and they reduce the proliferation and activity of these cells.

Leydig cells have LH receptors, and the major effect of LH is to stimulate androgen secretion via a cAMP-dependent mechanism (Fig. 12–4). The main product of Leydig cells is testosterone but two other androgens of less biological activity, dehydroepiandrosterone (DHEA) and androstenedione, are also produced.

There are bidirectional interactions between Sertoli and Leydig cells (Fig. 12–4). The Sertoli cell is incapable of producing testosterone but contains testosterone receptors as well as FSH-dependent aromatase. The Leydig cell does not produce estradiol but contains receptors for it, and estradiol can suppress the response of the Leydig cell to LH. Testosterone diffuses from the Leydig cells, crosses the basement membrane, enters the Sertoli cell, and binds to ABP. As a result, androgen levels can reach high local concentrations in the seminiferous tubules. Testosterone is obligatory for spermatogenesis and the proper functioning of Sertoli cells. In Sertoli cells, testosterone also serves as a precursor for estradiol production. The daily role of estradiol in the functioning of Leydig cells seems to be the LH responses modulation.

4. Spermatogenesis

Spermatogenesis is the process of transformation of male germ cells into spermatozoa. This process can be divided into three distinct phases.

1. **Mitotic Phase**. Spermatogonia are the representative cells. Spermatogonia undergo several rounds of mitotic division prior to entering the meiotic phase.
2. **Meiotic Phase**. Spermatocytes are the representative cells. The first meiotic division of *primary spermatocytes* gives rise to ***diploid*** (2n chromosomes) *secondary spermatocytes*. The second meiotic division produces ***haploid*** (1 set of chromosomes) cells called *spermatids*. Of every four spermatids emanating from a primary spermatocyte, two contain X chromosomes and two have Y chromosomes.
3. **Morphogenetic Phase**. *Spermaids* are the representative cells. Spermatozoa are the mature, differentiated product of this phase.

Spermatogenesis is initiated shortly before puberty, under the influence of the rising levels of gonadotropins and testosterone, and continues throughout life, with a slight decline during old age.

The time required to produce mature spermatozoa from the earliest stage of spermatogonia is 65 to 70 days. Because several developmental stages of spermatogenic cells occur during this time frame, the stages are collectively known as the *spermatogenic cycle*.

There is synchronized development of spermatozoa within the seminiferous tubules, and each stage is morphologically distinct. A spermatogonium becomes a mature spermatozoon after going through several rounds of mitotic divisions, a couple of meiotic divisions, and a few weeks of differentiation.

Spermatogenesis occurs along the length of each seminiferous tubule in successive cycles. New cycles are initiated at regular time intervals (every 2 to 3 weeks) before the previous ones are completed.

Approximately 200 million spermatozoa are produced daily in the adult human testis, which is about the same number of sperm present in a normal ejaculate.

Spermatogenesis requires high intratesticular levels of testosterone, secreted from the LH-stimulated Leydig cells (Fig. 12–4). Testosterone diffuses across the basement membrane of the seminiferous tubule (Fig. 12–2), crosses the blood-testis barrier, and complexes with ABP. Upon entering meiosis, spermatogenesis appears to depend on the availability of FSH and testosterone. In human males, FSH is thought to be required for the initiation of spermatogenesis before puberty. When adequate sperm production has been achieved, LH alone (through stimulation of testosterone production) or testosterone alone is sufficient to maintain spermatogenesis.

5. Steroidogenesis in testis

Following spermatogenesis, the second primary function of the testes is steroidogenesis. Steroidogenesis is the production of the steroid hormones. The male sex hormones are called *androgens*. The testes secrete several male sex hormones, including *testosterone, dihydrotestosterone, androstenediol, androstenedione dehydroepiandrosterone* (DHEA).

Testosterone, which is the most abundant of these hormones, is considered the main testicular hormone. The adrenal cortex also produces androgens, although in much smaller quantities (appr. 5% of the total male androgens) than those produced in the testes. The testes also secrete small quantities of *estradiol* and *estrone*.

Testosterone is produced and secreted by the interstitial Leydig's cells in the testis from cholesterol, 27-carbon (C27) steroid. Steroid hormones are produced from cholesterol by the adrenal cortex, ovaries, testes, and placenta. Each organ uses a similar steroid biosynthetic pathway, but the relative amount of the final products depends on the particular subset of enzymes expressed in that tissue and the trophic hormones (LH, FSH, ACTH) stimulating specific cells within the organ.

Cholesterol from low-density and high-density lipoproteins is released in the Leydig cell and transported from the outer mitochondrial membrane to the inner mitochondrial membrane — a process regulated by steroidogenic acute regulatory protein. Under the influence of LH, with cAMP as a second messenger, cholesterol is converted to pregnenolone (C21) by cholesterol side-chain cleavage enzyme (CYP11A1), which removes 6 carbons attached to the 21 position. Pregnenolone is a key intermediate for all steroid hormones in various steroidogenic organs. Pregnenolone is transported out of mitochondria by specific transport proteins. The pregnenolone then moves by diffusion to the smooth ER, where the remainder of sex hormone biosynthesis takes place.

Pregnenolone can be converted to testosterone via two pathways, the **delta-5-pathway** and the **delta-4-pathway**. In the delta 5 pathway, the double bond is in ring B; in the delta 4 pathway the double bond is in ring A.

The delta-5 intermediates include 17α-hydroxypregnenolone, DHEA, and androstenediol, while the delta-4 intermediates are progesterone, 17α-hydroxyprogesterone and androstenedione.

The conversion of C21 steroids (the progestins) to androgens (C19 steroids) proceeds in two steps: first, 17α-hydroxylation of pregnenolone (to form

17α- hydroxypregnenolone) and second, C17,20 cleavage; thus, two carbons are removed to form DHEA. This hydroxylation and cleavage is accomplished by a single enzyme, 17α-hydroxylase or 17,20-lyase (CYP17). DHEA is converted to androstenedione by another two-step enzymatic reaction: dehydrogenation in position 3 (catalyzed by 3β-hydroxysteroid dehydrogenase [3β-HSD]) and shifting of the double bond from ring B to ring A (catalyzed by delta-4,5-ketosteroid isomerase); these two may be the same enzyme. The final reaction yielding testosterone is carried out by 17-ketosteroid reductase (17β- hydroxysteroid dehydrogenase), which substitutes the keto group in position 17 with a hydroxyl group. Unlike all the preceding enzymatic reactions, this is a reversible step but tends to favour testosterone.

Although estrogens are only minor products of testicular steroidogenesis, they are normally found in low concentrations in men. Androgens (C19) are converted to estrogens (C18) by the action of the enzyme complex aromatase (CYP19). Aromatization involves the removal of the methyl group in position 19 and the rearrangement of ring A into an unsaturated aromatic ring. The products of aromatization of *testosterone* and *androstenedione* are *estradiol* and *estrone*, respectively.

In the testes, the Sertoli cell is the main site of aromatization, which is stimulated by FSH; however, aromatization may also occur in peripheral tissues that lack FSH receptors (e.g., adipose tissue).

In the bloodstream, testosterone exists in a free (unbound) or a bound form. The bound form is attached to plasma proteins, including albumin and the sex hormone-binding protein produced by the liver. Only approximately 2% of circulating testosterone is unbound and therefore able to enter the cell and exert its metabolic effects.

Plasma testosterone levels and rate of secretion have age-dependent manner in the normal male. Fig. 12–5 depicts the relative plasma LH and testosterone concentrations throughout the life of the normal human male. The number refer to the below listed descriptions.

1. *Fetal life.* The development of male and female internal and external structures depends on the fetal hormonal environment. The Wolffian and Müllerian ducts are initially present in both male and female fetuses. If there is no hormonal input (the situation in the normal female fetus), female internal and female external structures develop (Müllerian ducts develop, Wolffian ducts regress).

Normal male development requires the presence of three hormones: testosterone, dihydrotestosterone and the MIF.

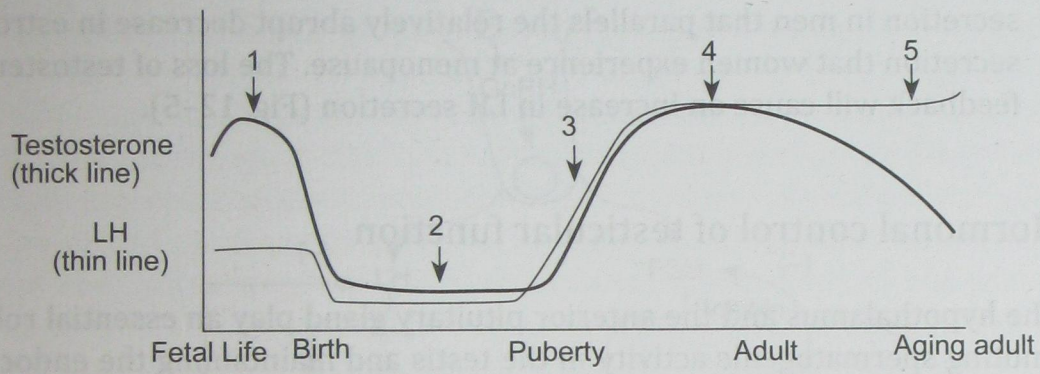


Fig. 12–5. Age-related changes in LH and testosterone secretion in the normal male.

MIF prevents the development of the Müllerian ducts which otherwise would differentiate into female internal structures. In the absence of MIF, the Müllerian ducts develop. Thus, in addition to normal structures, an uterus will be present.

Wolffian ducts differentiate into the majority of male internal structures; namely, epididymis, vasa deferentia and seminal vesicles. In the absence of testosterone, the Wolffian ducts regress.

Dihydrotestosterone induces the urogenital sinus and genital tubercle to differentiate into the external scrotum, penis and prostate gland. In the absence of dihydrotestosterone, female external structures develop.

2. *Childhood.* Within a few months after birth LH and testosterone drop to low levels and remain low until puberty. The cause of this prolonged quiescence of reproductive hormone secretion during childhood is not established. Interestingly, LH secretion remains low in spite of low testosterone.
3. *Puberty.* Near the onset of puberty, the amplitude of the LH becomes greater, driving the mean level of LH higher. Early in puberty this potentiation of the LH pulses is especially pronounced during sleep. This increased LH stimulates the Leydig cells again to secrete testosterone.
4. *Adult.* During adulthood, LH secretion drives testosterone secretion. Thus, it is not surprising that relative levels of these two hormones parallel one another.
5. *Aging Adult.* Testosterone secretion decreases with age. Men aged 70 generally secrete only 60–70 % as much testosterone as do men in their 20s. Nevertheless, there is no abrupt decrease in testosterone

secretion in men that parallels the relatively abrupt decrease in estrogen secretion that women experience at menopause. The loss of testosterone feedback will cause an increase in LH secretion (Fig. 12–5).

6. Hormonal control of testicular function

The hypothalamus and the anterior pituitary gland play an essential role in promoting spermatogenic activity in the testis and maintaining the endocrine function of the testis by means of the gonadotropic hormones. The synthesis and release of the gonadotropic hormones from the pituitary gland are regulated by gonadotropin-releasing hormone (GnRH), which is synthesized by the hypothalamus and secreted into the hypothalamo-hypophysial portal circulation (Fig. 12–6).

Two gonadotropic hormones are secreted by the pituitary gland: FSH and LH. LH regulates the secretion of testosterone by the Leydig cells and FSH, in synergy with testosterone, regulates the production of spermatozoa.

FSH binds selectively to Sertoli's cells surrounding the seminiferous tubules, where it functions in the initiation of spermatogenesis. Under the influence of FSH, Sertoli's cells produce androgen-binding protein, plasminogen activator and inhibin.

Androgen-binding protein binds testosterone and serves as a carrier of testosterone in Sertoli's cells and as a storage site for testosterone. Although FSH is necessary for the initiation of spermatogenesis, full maturation of the spermatozoa requires testosterone. Androgen-binding protein also serves as a carrier of testosterone from the testes to the epididymis. Plasminogen activator, which converts plasminogen to plasmin, functions in the final detachment of mature spermatozoa from Sertoli's cells.

LH binds membrane receptors on the surface of Leydig cells (Fig. 12–6). The activation of LH and FSH receptors on Leydig and Sertoli cells increases the intracellular second messenger cAMP. The two gonadotropin receptors are linked to G-proteins and adenyl-cyclase for the production of cAMP from ATP. For the most part, cAMP can account for all of the actions of LH and FSH on testicular cells. cAMP binds to protein kinase A, which activates transcription factors. These factors activate the promoter region of the genes of steroidogenic enzymes that control testosterone production by Leydig cells. Similar signal-transducing events occur in Sertoli cells that regulate the production of estradiol.

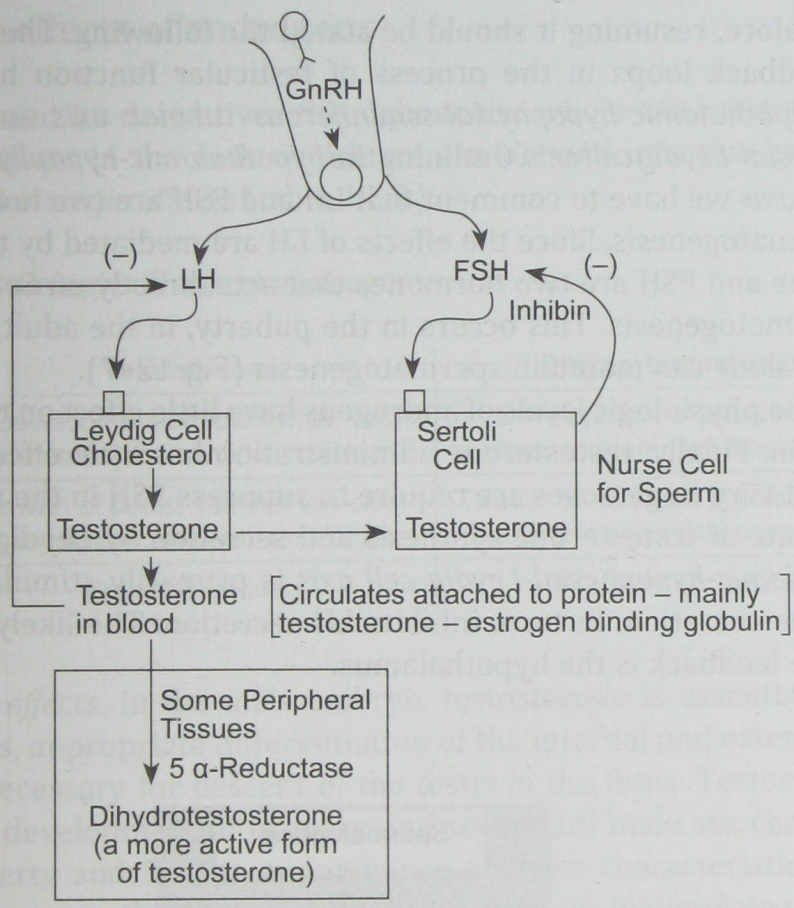


Fig. 12–6. Hypothalamic-pituitary feedback control of spermatogenesis and testosterone levels in the male.

Circulating levels of the gonadotropic hormones are regulated in a negative feedback manner by testosterone. High levels of testosterone suppress LH secretion through a direct action on the pituitary and an inhibitory effect on the hypothalamus.

FSH is thought to be inhibited by a substance called *inhibin*, produced by Sertoli's cells. Inhibin suppresses FSH release from the pituitary gland. The pituitary gonadotropic hormones and Sertoli's cells in the testis form a classic negative feedback loop in which FSH stimulates inhibin, and inhibin suppresses FSH. Unlike the cyclic hormonal pattern in the female, in the male, FSH, LH, and testosterone secretion and spermatogenesis occur at relatively unchanging rates during adulthood.

Therefore, resuming it should be stated the following. There are two negative feedback loops in the process of testicular function hormonal regulation: *hypothalamic-hypophysial-seminiferous tubular axis* and *hypothalamic-hypophysial-Leydig cell axis*. Outlining the *hypothalamic-hypophysial-seminiferous tubular axis* we have to comment that LH and FSH are two hormones required for spermatogenesis. Since the effects of LH are mediated by testosterone, testosterone and FSH are two hormones that acts directly on Sertoli cells to promote gametogenesis. This occurs in the puberty, in the adult, however, testosterone alone can maintain spermatogenesis (Fig. 12–7).

Plasma physiologic levels of androgens have little effect on the FSH secretion inhibition. Finally, testosterone administration has little effect on FSH secretion, and very large doses are require to suppress FSH in the male.

The rate of testosterone synthesis and secretion by Leydig cells within the *hypothalamic-hypophysial-Leydig cell axis* is primarily stimulated by LH. Testosterone secretion, in turn, inhibits LH secretion. The likely major target of negative feedback is the hypothalamus.

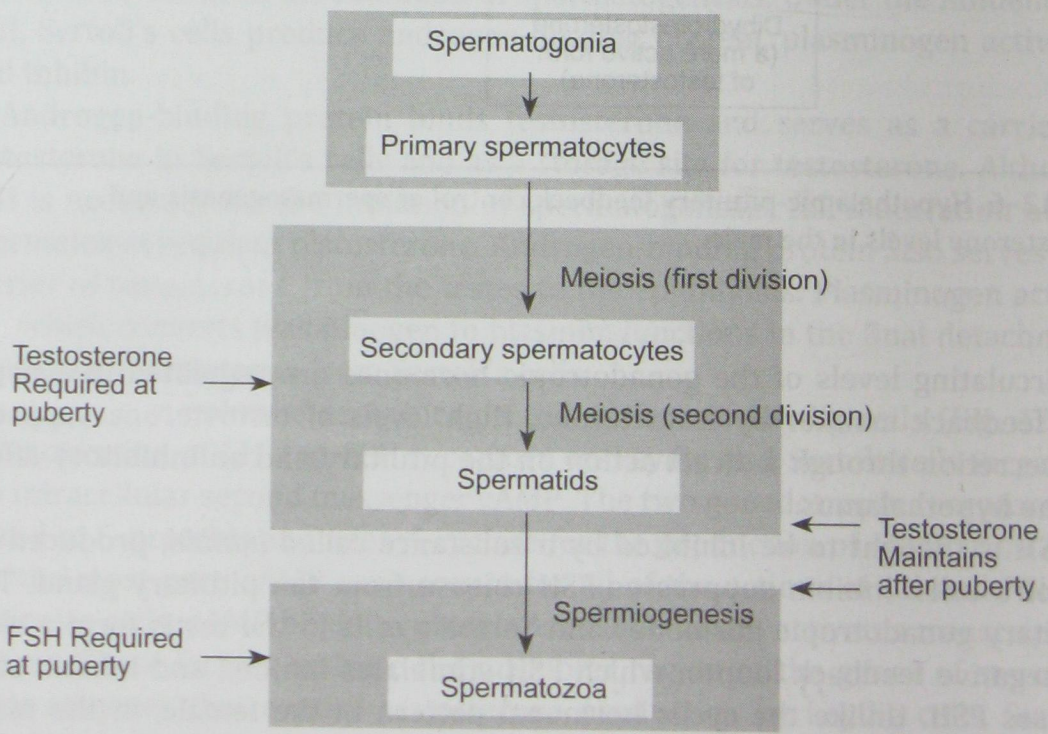


Fig. 12–7. The scheme of spermatogenesis endocrine control.

7. Physiologic effects of androgens

Testosterone and its derivatives throughout the whole life exert a variety of physiologic effects in the male which we can divided into two groups — *reproductive* and *biologic* effects (table 12–1).

Table 12–1. Physiologic effects of androgens

Life stage	Testosterone	Dihydrotestosterone
Fetal period	Development of epididymis, vas deferens and seminal vesicles	Development of penis, penile urethra, scrotum and prostate gland
Puberty	Growth of penis, seminal vesicles, musculature, skeleton and larynx	Growth of scrotum, prostate gland, pubic hair and sebaceous glands
Adulthood	Spermatogenesis	Prostatic secretions

Reproductive effects. In the male embryo, testosterone is essential for the spermatogenesis, appropriate differentiation of the internal and external genitalia, and it is necessary for descent of the testis in the fetus. Testosterone is essential to the development of primary and secondary male sex characteristics during puberty and for the maintenance of these characteristics during adult life. It causes growth of pubic, chest, and facial hair; it stimulate growth of the larynx that result in the male bass voice and promote haemoglobin synthesis (males have higher haemoglobin levels than females) and bone growth. It increases also the thickness of the skin and the activity of the sebaceous glands, predisposing to acne.

Biologic effects. All or almost all of the actions of testosterone and other androgens result from cell division stimulation and increased protein synthesis in target tissues. Androgens function as anabolic agents in males and females to promote metabolism together with muscles and skeletal growth. It promotes the erythropoiesis that additionally to its reproductive effects resulted in the bigger male plasma erythrocytes and haemoglobin levels pertaining the same in females.

In the adolescent, androgens produce linear growth, muscular development and retention of nitrogen, potassium and phosphorus. Testosterone also accelerates epiphyseal fusion of the long bones.

Testosterone and the androgens have a great effect on the development of increasing musculature during puberty, with boys averaging approximately 50% more of an increase in muscle mass than do girls. Testosterone stimulates

differentiation of the Wolffian duct system into the epididymis, vas deferens and seminal vesicles (Table 12–1). Dihydrotetosterone stimulates organogenesis of the urogenital sinus and tubercle into the prostate gland, penis, urethra and scrotum. Androgens are responsible for libido.

8. Male sexual function, semen composition and fertility

The approximate volume of the human ejaculate (semen) is 1.5 to 5.0 ml. The bulk of this fluid (45 % to 80 %) is produced by the seminal vesicles, and 15% to 30% is contributed by the prostate. the seminal vesicles are the source of prostaglandins, fructose, ascorbic acid and phosphorylcholine. Metabolism of fructose provides energy for sperm motility. The prostate gland is the origin of the citric acid, acid phosphatase, zinc and spermine.

Sperm represents less than 10 % of the ejaculate volume. There are usually between 100 and 300 million spermatozoa in it. Normal human semen values are summarized in table 12–2.

Table 12–2. Semen analysis

Characteristic	Referent value
Volume of ejaculate	1.5–5.0 ml
Sperm count	40–250 million/ml
Sperm motility Percentage of motile forms: 1 hour after ejaculation 3 hours after ejaculation	 70% or more 60% or more
Leukocyte count	0–2,000/ml
pH	7.2–7.8
Fructose concentration	1.5–6.0 mg/ml

Euspermia is defined as greater than 20 million sperm/ml of ejaculate.

A sperm concentration below about 20 million per ml is termed *oligospermia* and is associated with decreased fertility. A total sperm count below about 50 million per ejaculation is clinically significant in male infertility.

Azoospermia is defined as 5 million sperm/ml of ejaculate.

The male sex act involves erection, emission, ejaculation and detumescence. The physiology of these functions involves a complex interaction between autonomic-mediated spinal cord reflexes, higher neural centers, and the vascular system.

Penile erection is the first effect of male sexual stimulation, whether psychological or physical. It involves increased inflow of blood into the corpora cavernosa due to relaxation of the trabecular smooth muscle that surrounds the sinusoidal spaces and compression of the veins controlling outflow of blood from the venous plexus.

Erection is mediated by parasympathetic impulses that pass from the sacral segments of the spinal cord through the pelvic nerves to the penis.

Parasympathetic stimulation results in release of nitric oxide, a nonadrenergic-noncholinergic neurotransmitter, which causes relaxation of trabecular smooth muscle of the corpora cavernosa (Fig. 12–8). This relaxation permits inflow of blood into the sinuses of the cavernosa at pressures approaching those of the arterial system.

Nitric oxide released by parasympathetic axons and produced by the endothelial cells of penile blood vessels activates guanylate cyclase in the vascular smooth muscle cells.

Guanylate cyclase catalyzes the production of cyclic GMP (cGMP), which closes Ca^{2+} channels in the plasma membrane (Fig. 12–8) that decreases the cytoplasmic Ca^{2+} concentration, causing smooth muscle relaxation (chapter # 2).

The penile blood vessels thereby dilate to increase the blood flow into the erectile tissue, producing an erection. As the erectile tissues become engorged with blood and the penis becomes turgid, venous outflow of blood is partially occluded, thus aiding erection. The term emission refers to the movement of semen into the urethra, and ejaculation refers to the forcible expulsion of semen from the urethra out of the penis. Emission and ejaculation are stimulated by sympathetic nerves, which cause peristaltic contractions of the tubular system, contractions of the seminal vesicles and prostate, and contractions of muscles at the base of the penis. Sexual function in the male thus requires the synergistic action of both the parasympathetic and sympathetic systems.

Erection is controlled by two portions of the central nervous system — the hypothalamus in the brain and the sacral portion of the spinal cord. Conscious sexual thoughts originating in the cerebral cortex act via the hypothalamus to control the sacral region, which in turn increases parasympathetic nerve activity to promote vasodilation and erection of the penis. Conscious thought is not required for erection, however, because sensory stimulation of the

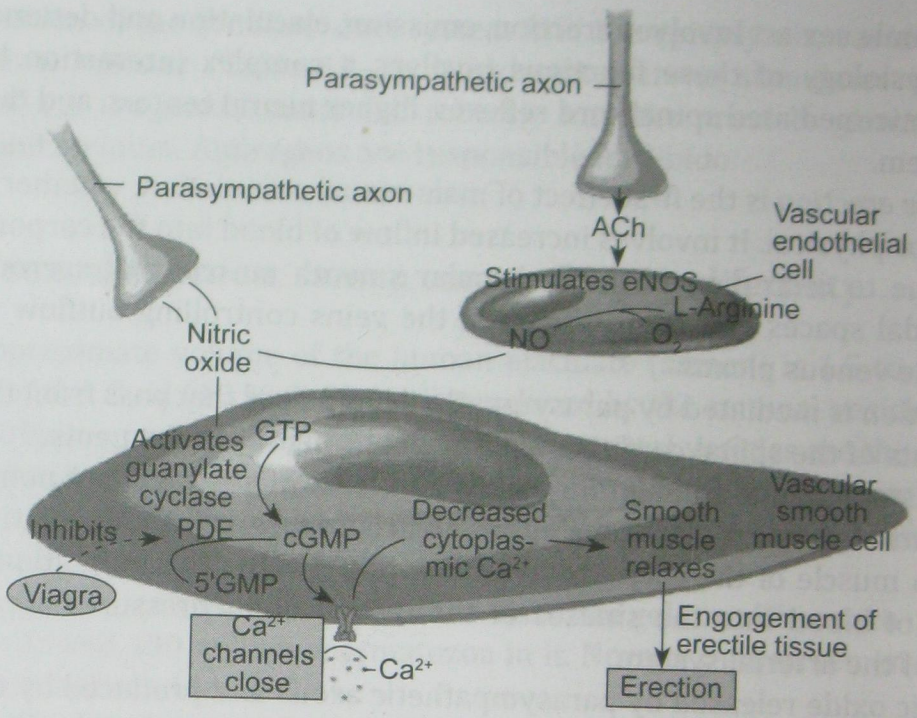


Fig. 12–8. Role of nitric oxide in penile erection. Notes: eNOS = endothelial nitric oxide synthase. The enzyme phosphodiesterase (PDE) breaks down cGMP, and it is this enzyme that is inhibited by the drug Viagra.

penis can more directly activate the sacral region of the spinal cord and cause an erection than antagonistic action of the parasympathetic and sympathetic systems.

Emission and ejaculation are mediated by the sympathetic nervous system. Erectile dysfunction is defined as the inability to achieve and maintain an erection sufficient to permit satisfactory sexual intercourse. It can be due to psychogenic factors, organic disorders or mixed psychogenic and organic conditions.

9. Androgens excess and deficiency

Decreased release of androgens can be due to a *lack of GnRH*. Even *non-pulsatile GnRH secretion* stimulates androgen formation inadequately. Both can occur with damage to the hypothalamus (tumor, radiation, abnormal perfusion, genetic defect) as well as psychological or physical stress (Fig. 12–9, 1).

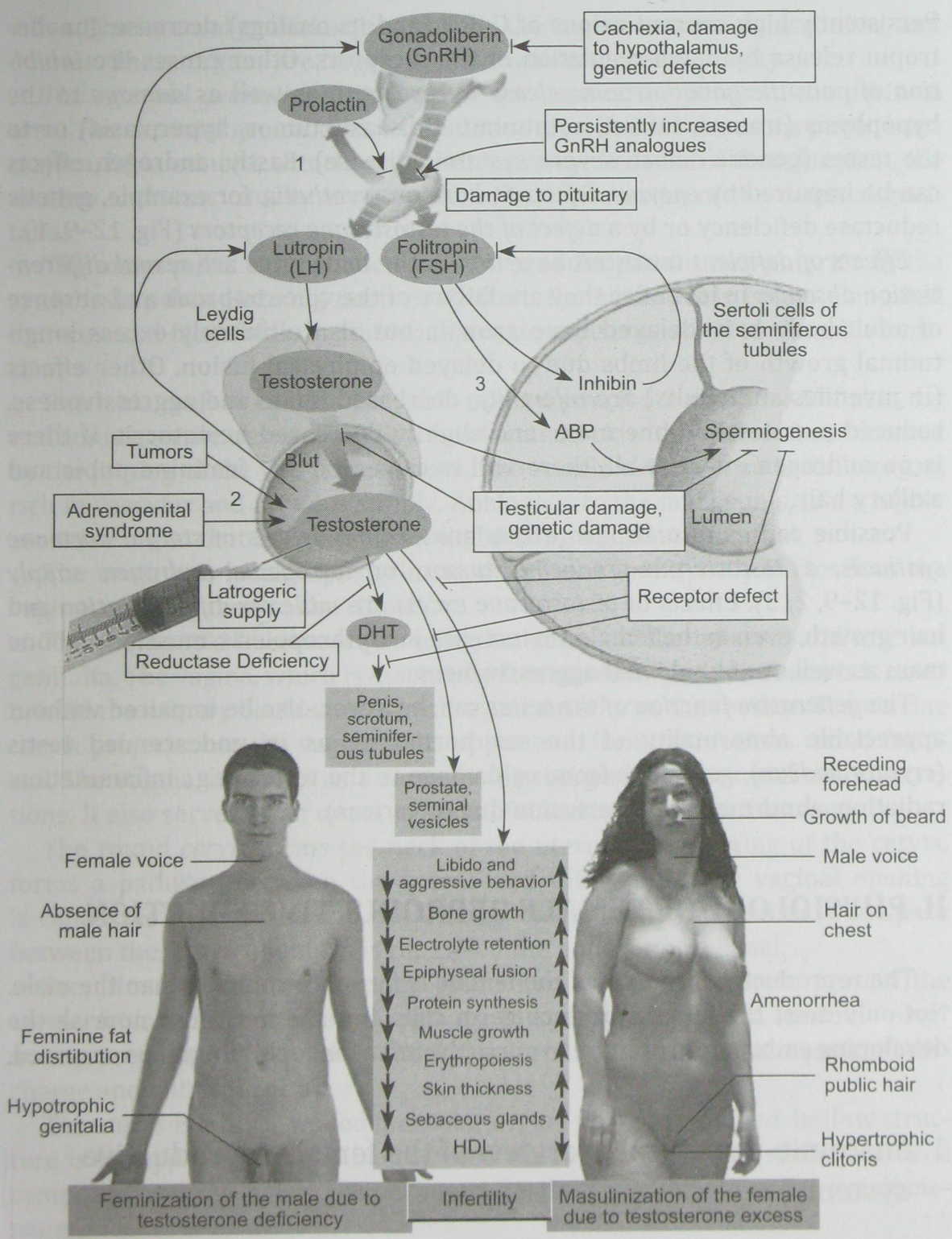


Fig. 12–9. Causes and sequences of male sexual hormones excess and deficiency.

Persistently high concentrations of GnRH (and its analogs) decrease gonadotropin release by down-regulation of the receptors. Other causes are *inhibition of pulsatile gonadotropin release* by prolactin as well as *damage* to the hypophysis (trauma, infarct, autoimmune disease, tumor, hyperplasia) or to the testes (genetic defect, severe systemic disease). Lastly, androgen effects can be impaired by *enzyme defects in hormone synthesis*, for example, genetic reductase deficiency or by a *defect of the testosterone receptors* (Fig. 12–9, 1).

Effects of deficient testosterone action in the male fetus are *sexual differentiation absence;* in juveniles they are failure of the voice to break and absence of adult body hair, delayed bone growth, but also ultimately excess longitudinal growth of the limbs due to delayed epiphyseal fusion. Other effects (in juveniles and adults) are *infertility*, decreased libido and aggressiveness, reduced muscle and bone mass, and slightly decreased hematocrit. If there is no androgen effect at all, there will not even be any feminine pubic and axillary hair.

Possible causes of *androgen excess* are *enzyme defects in steroid hormone synthesis*, a *testosterone-producing tumor*, or *iatrogenic androgen supply* (Fig. 12–9, 2, 3). Effects of *testosterone excess* are *male sex differentiation* and hair growth, even in the female, an increase in erythropoiesis, muscle and bone mass as well as of libido and aggressiveness.

The *generative function of the testes* can, however, also be impaired without appreciable abnormality of the sex hormones, as in undescended testis (*cryptorchidism*), genetic defects, or damage to the testes (e.g., inflammation, radiation, abnormal blood perfusion due to varices).

II. PHYSIOLOGY OF FEMALE REPRODUCTIVE FUNCTION

The reproductive function of the female is far more complex than the male. Not only must the female produce germ cells, but she must also nourish the developing embryo and prepare to nurse the infant once childbirth has occurred.

1. Anatomic-structural overview of the female reproductive system

The female genitourinary system consists of the external and internal genital organs. The external sex organs of the female are referred to as the genitalia or

vulva. The internal genital organs include the vagina, uterus, uterine tubes, and ovaries. These organs are largely located within the pelvic cavity.

The external genitalia are located at the base of the pelvis in the perineal area. The external genitalia, also called the vulva, include the mons pubis, labia majora, labia minora, clitoris, and perineal body. Because of their location, the urethra and anus usually are considered in a discussion of the external genitalia.

The mons pubis is a rounded skin-covered fat pad located anterior to the symphysis pubis. Running posteriorly from the mons pubis is two elongated hair-covered fatty folds, the labia majora that are analogous to the male scrotum.

The labia majora enclose the labia minora which are composed of skin, fat and some erectile tissue. The clitoris is located below the clitoral hood formed by the joining of the two labia minora. The female clitoris is an erectile organ, rich in vascular and nervous supply. Analogous to the male penis, it is a highly sensitive organ that becomes distended during sexual stimulation.

The urethra is located posterior to the clitoris and usually is closer to the vaginal opening than to the clitoris.

The *vagina* is a fibromuscular tube that connects the external and internal genitalia. The vagina, which is essentially free of sensory nerve fibers, is located behind the urinary bladder and urethra and anterior to the rectum. The uterine cervix projects into the vagina at its upper end, forming recesses called fornices. The vagina functions as a route for discharge of menses and other secretions. It also serves as an organ of sexual fulfillment and reproduction.

The round *cervix* forms the neck of the uterus. The opening of the cervix, forms a pathway between the uterus and the vagina. The vaginal opening is called the external os and the uterine opening, the internal os. The space between these two openings is the called the endocervical canal.

Secretions from the columnar epithelium of the endocervix protect the uterus from infection, alter receptivity to sperm, and form a mucoid "plug" during pregnancy. The endocervical canal provides a route for menstrual discharge and entry of sperm.

The uterus is a thick-walled muscular organ. This pear-shaped, hollow structure is located between the bladder and the rectum. The wall of the uterus is composed of three layers: the perimetrium, the myometrium and the endometrium.

The endometrial cycle can be divided into three phases: proliferative, secretory, and menstrual. The proliferative, or preovulatory, phase is the period

during which the glands and stroma of the superficial layer grow rapidly under the influence of estrogen. The secretory, or postovulatory, phase is the period during which progesterone produces glandular dilation and active mucus secretion and the endometrium becomes highly vascular and edematous. The menstrual phase is the period during which the superficial layer degenerates and sloughs off.

The fallopian, or uterine, tubes are slender cylindrical structures attached bilaterally to the uterus and supported by the upper folds of the broad ligament. The end of the fallopian tube nearest the ovary forms a funnel-like opening with fringed finger-like projections, called fimbriae, that pick up the ovum following its release into the peritoneal cavity after ovulation. The fallopian tubes are formed of smooth muscle and lined with a ciliated, mucus-producing epithelial layer. The beating of the cilia, along with contractile movements of the smooth muscle, propels the nonmobile ovum toward the uterus. Besides providing a passageway for ova and sperm, the fallopian tubes provide for drainage of tubal secretions into the uterus.

In the adult, *the ovaries* are flat, almond-shaped structures that are 4.0 · 2.5 · 1.5 cm in dimension. They are located on either side of the uterus below the ends of the two fallopian tubes. The ovaries are attached to the posterior surface of the broad ligament and to the uterus by the ovarian ligament. They are covered with a thin layer of surface epithelium that is continuous with the lining of the peritoneum. The integrity of this covering is periodically broken at the time of ovulation. The mature ovary has two major functions: the maturation of germ cells and steroidogenesis.

2. Folliculogenesis, oogenesis

Unlike the male gonads, which produce sperm throughout a man's reproductive life, the female gonads contain a fixed number of ova at birth that diminishes throughout a woman's life. Without germ cells, the ovary does not develop. The germs cells — *oogonia* — actively divide by mitosis. Oogonia undergo mitosis only during the prenatal period. By birth, the ovaries contain a finite number of oocytes, estimated to be about 1 million. Most of them will undergo atresia. However, some will develop into mature follicles, produce steroids, and ovulate. By puberty, only 200,000 oocytes remain; by age 30, only 26,000 remain; and by the time of menopause, the ovaries are essentially devoid of oocytes.

When oogonia finish the process of mitosis, they are called (primary) *oocytes* enclosing 46 chromosomes. At that time they enter the meiotic cycle (or meiosis, to prepare for the production of a haploid ovum), become arrested in prophase of the first meiotic division, and remain arrested in that phase until they either die or grow into mature oocytes at the time of ovulation. During meiosis, each primary oocyte produces a single haploid gamete. If the secondary oocyte is fertilized, it forms a second polar body and its nucleus fuses with that of the sperm cell to become a zygote. Thus, the second meiotic division is completed only by an oocyte that has been fertilized.

The process of *oogenesis* begins during the sixth week of fetal life and proceeds to the development of the primary oocytes, which become surrounded by a single layer of granulosa cells. The primary oocytes with their surrounding granulosa cells are referred to as primordial follicles (Fig. 12–10).

It is critically important to imagine that oogenesis and folliculogenesis are two important internal processes that occurred simultaneously throughout the different stages of the female organism development (Table 12–3). To illustrate this we'll continue.

Being the primitive germ cells, primordial follicles provide the 1 to 2 million oocytes that are present in the ovaries at birth. Throughout childhood, the granulosa cells provide nourishment for the ovum and secrete an inhibiting factor that keeps the ovum suspended in a primordial state.

After puberty, when FSH and LH from the anterior pituitary begin to be secreted in sufficient amounts, the entire ovaries, together with some of the follicles within them, begin to grow. As follicles mature, oocytes will also mature by entering meiosis, which produces the proper number of chromosomes in preparation for fertilization.

The process of follicle's mature defines as folliculogenesis.

Follicles might be in one of the following physiological states: resting, growing, degenerating, or ready to ovulate. During each menstrual cycle, the ovaries produce a group of growing follicles of which most will fail to grow to maturity and will undergo follicular atresia (death) at some stage of development. However, one domi-nant follicle generally emerges from the cohort of developing follicles and it will ovulate, releasing a mature haploid ovum.

Primordial follicles are generally considered the nongrowing resting pool of follicles, which gets progressively depleted throughout life; by the time of menopause, the ovaries are essentially devoid of all follicles. Primordial follicles are located in the ovarian cortex (peripheral regions of the ovary) beneath the tunica albuginea.

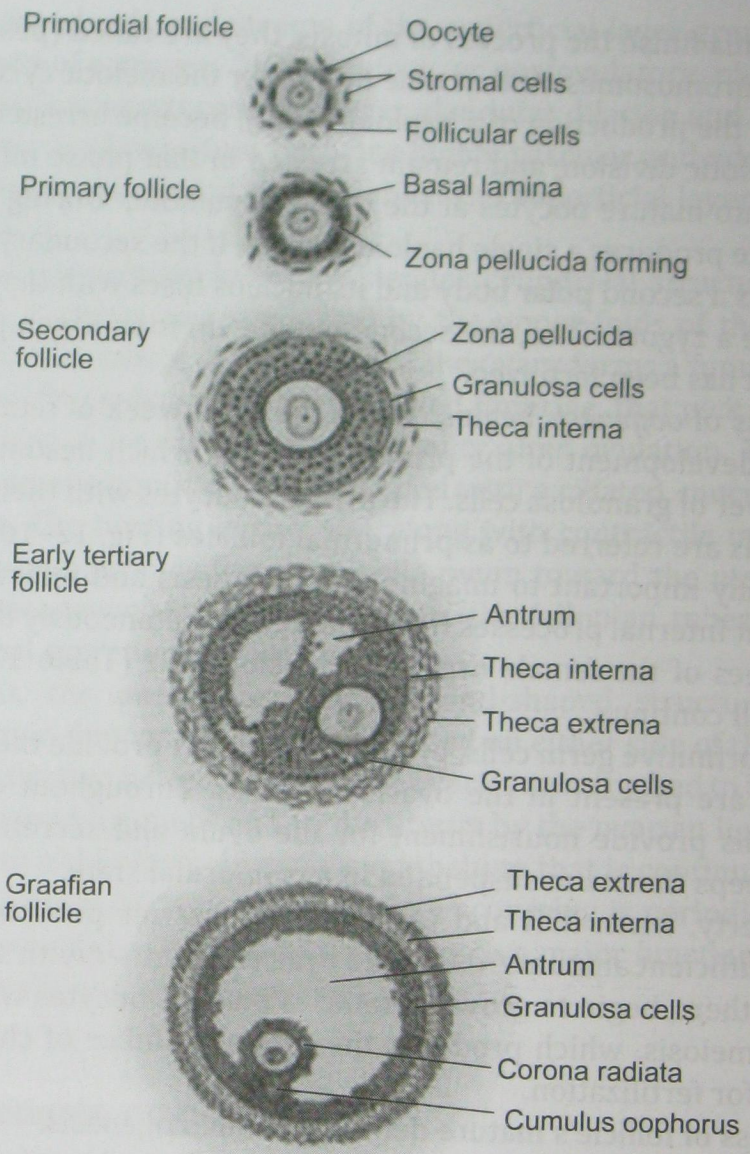


Fig. 12–10. The developing follicle, from primordial through graafian.

Progression from primordial to the next stage of follicular development, the primary stage, occurs at a relatively constant rate throughout fetal, juvenile, prepubertal, and adult life. Once primary follicles leave the resting pool, they are committed to further development or atresia. Most become atretic, and typically only one fully developed follicle will ovulate. The conversion from primordial to primary follicles is believed to be independent of pituitary gonado-

Table 12–3. Oogenesis and folliculogenesis comparative expression in the female organism

Stage	Process	Ovum	Follicle
Fetal life	Migration	Primordial germ cells	
	Mitosis	Oogonia	Primordial follicle
	First meiotic division begins	Primary oocyte	Primary follicle
Birth	Arrest in prophase		
	Growth of oocyte and follicle		
Puberty	Follicular maturation		Secondary follicle
Cycle			Antral follicle
Ovulation	Resumption of meiosis	Secondary oocyte	Graafian follicle
	Emission of first polar body		
	Arrest in metaphase		
Fertilization	Second meiotic division complete	Zygote	Corpus luteum
	Emission of second polar body		
Implantation	Mitotic divisions	Embryo	
	Blastocyst		
Parturition	Body Patterning	Fetus	Corpus albicans

tropins. The exact signal that recruits a follicle from a resting to a growing pool is unknown; it could be programmed by the cell genome or influenced by local ovarian growth regulators.

The first sign that a primordial follicle is entering the growth phase is a morphological change of the flattened pregranulosa cells into cuboidal granu-

losa cells. The cuboidal granulosa cells proliferate to form a single continuous layer of cells surrounding the oocyte, which has enlarged from 20 μm in the primordial stage to 140 μm in diameter. At this stage, a glassy membrane, the *zona pellucida*, surrounds the oocyte and serves as means of attachment through which the granulosa cells communicate with the oocyte. This is the *primary follicular stage* of development, consisting of one layer of cuboidal granulosa cells and a basement membrane (Fig. 12–10).

The follicle continues to grow, mainly through proliferation of its granulosa cells, so that several layers of granulosa cells exist in the *secondary follicular stage* of development (Fig. 12–10).

During the development of the secondary follicle, the primary oocyte increases in size, and the surrounding granulosa cells proliferate to form a multilayered wall around it. In addition, cells from the surrounding ovarian interstitium align themselves to form a cellular wall called the *theca*. The cells of the theca become differentiated into two layers: an inner — *theca interna*, which lies adjacent to the follicular cells, and an outer — *theca externa*. The cells in the *theca interna* take on epithelioid characteristics similar to those of the granulosa cells and develop the ability to secrete additional sex hormones (estrogen and progesterone). The *theca externa* develops into a highly vascular connective tissue capsule that becomes the capsule for the developing follicle. As the follicle continues to grow, theca layers expand, and fluidfilled spaces or *antra* begin to develop around the granulosa cells. This early antral stage of follicle development is referred to as the *tertiary follicular stage* (Fig. 12–10).

The critical hormone responsible for progression from the preantral to the antral stage is FSH. Mitosis of the granulosa cells is stimulated by FSH. As the number of granulosa cells increases, the production of estrogens, the binding capacity for FSH, the size of the follicle, and the volume of the follicular fluid all increase significantly. As the antra increase in size, a single, large, coalesced antrum develops, pushing the oocyte to the periphery of the follicle and forming a large 2- to 2.5 cm-diameter *graafian follicle* (Fig. 12–10). Three distinct granulosa cell compartments are evident in the graafian follicle. Granulosa cells surrounding the oocyte are *cumulus granulosa cells* (i.e. *cumulus oophorus*). Those cells lining the antral cavity are called *antral granulosa cells* and those attached to the basement membrane are called *mural granulosa cells*. Mural and antral granulosa cells are more steroidogenically active than cumulus cells.

The monthly series of events associated with the maturation of the ovum and follicles is called the *ovarian cycle*. It consists of two phases: the *follicular phase* and the *luteal phase*. The *follicular phase*, typically days 1 to 14, is the

period of follicle growth. The *luteal phase*, days 14 to 28, is the period of corpus luteum activity. The typical ovarian cycle repeats at intervals of 28 days, with ovulation occurring midcycle. However, cycles as long as 40 days and as short as 21 days are not uncommon.

Usually by the 10th to 14th day after the 1st day of menstruation only one follicle has continued its growth to become a fully mature graafian follicle (Fig. 12–10). Other secondary follicles during that cycle regress and become *atretic* — a term that means "without an opening" in reference to their failure to rupture. Follicle atresia, or degeneration, is a type of apoptosis that results from a complex interplay of hormones and paracrine regulators.

The follicle that is protected from atresia and that develops into a graafian follicle becomes so large that it forms a bulge on the surface of the ovary. Under proper hormonal stimulation, this follicle will rupture — much like the popping of a blister — and extrude its oocyte into the uterine tube in the process of *ovulation* (Fig. 12–11). The released cell is a secondary oocyte. If

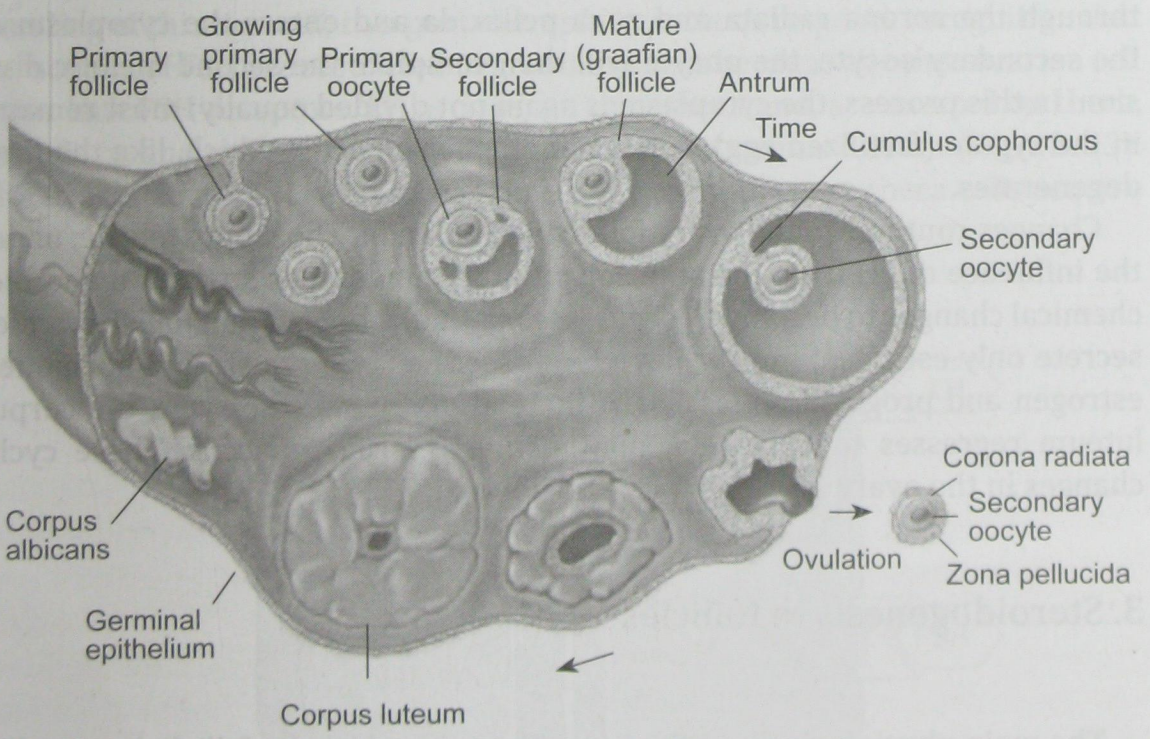


Fig. 12–11. Schematic diagram of an ovary, showing the sequence of events in the origin, growth and rupture of an ovarian follicle and the formation and retrogression of a corpus luteum.

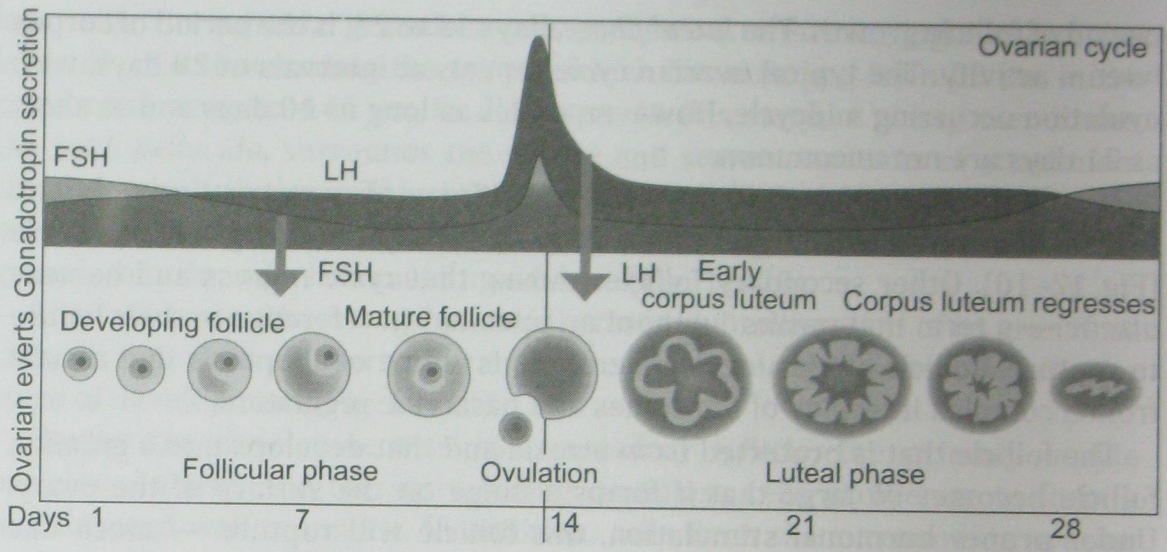


Fig. 12–12. The cycles of ovulation.

it is not fertilized, it will degenerate in a couple of days. If a sperm passes through the corona radiata and zona pellucida and enters the cytoplasm of the secondary oocyte, the oocyte will then complete the second meiotic division. In this process, the cytoplasm is again not divided equally; most remains in the zygote (fertilized egg), leaving another polar body which, like the first, degenerates.

Changes continue in the ovary following ovulation. The empty follicle, under the influence of LH from the anterior pituitary, undergoes structural and biochemical changes to become a *corpus luteum*. Unlike the ovarian follicles, which secrete only estrogen, the corpus luteum secretes two sex steroid hormones: estrogen and progesterone. Toward the end of a nonfertile cycle, the corpus luteum regresses to become a nonfunctional *corpus albicans*. These cyclic changes in the ovary are summarized below (Fig. 12–12).

3. Steroidogenesis in follicles

The main physiologically active steroid produced by the follicle is *estradiol*. *Steroidogenesis* — the process of steroid hormone production, depends on the availability of cholesterol, which originates from several sources and serves as the main precursor for all of steroidogenesis. Ovarian cholesterol can come

from plasma lipoproteins, *de novo* synthesis in ovarian cells, and cholesterol esters within lipid droplets in ovarian cells. For ovarian steroidogenesis, the primary source of cholesterol is low-density lipoprotein (LDL).

The conversion of cholesterol to *pregnenolone* by *cholesterol side-chain cleavage enzyme* is a rate-limiting step regulated by LH using the second messenger camp. LH binds to specific membrane receptors on theca cells, activates adenyl-cyclase through a G-protein, and increases the production of cAMP. cAMP increases LDL receptor mRNA, the uptake of LDL cholesterol and cholesterol ester synthesis. cAMP also increases the transport of cholesterol from the outer to the inner mitochondrial membrane, the site of pregnenolone synthesis using a unique protein called *steroidogenic acute regulatory protein*. Pregnenolone diffuses out of the mitochondria and enters the ER, the site of subsequent steroidogenesis (Fig. 12–13).

Two steroidogenic pathways may be used for subsequent steroidogenesis (Fig. 12–6). In *theca cells*, the *delta 5 pathway* is predominant; in *granulosa cells* and the *corpus luteum*, the *delta 4 pathway* is predominant.

Resuming, estradiol secretion by the follicle requires cooperation between granulosa and theca cells and coordination between FSH and LH. An understanding of this *two-cell, two-gonadotropin hypothesis* requires recognition that the actions of FSH are restricted to granulosa cells because all other ovarian cell types lack FSH receptors. LH actions, on the other hand, are exerted on theca, granulosa, and stromal (interstitial) cells and the corpus luteum (Fig. 12–13). The expression of LH receptors is time-dependent because theca cells acquire

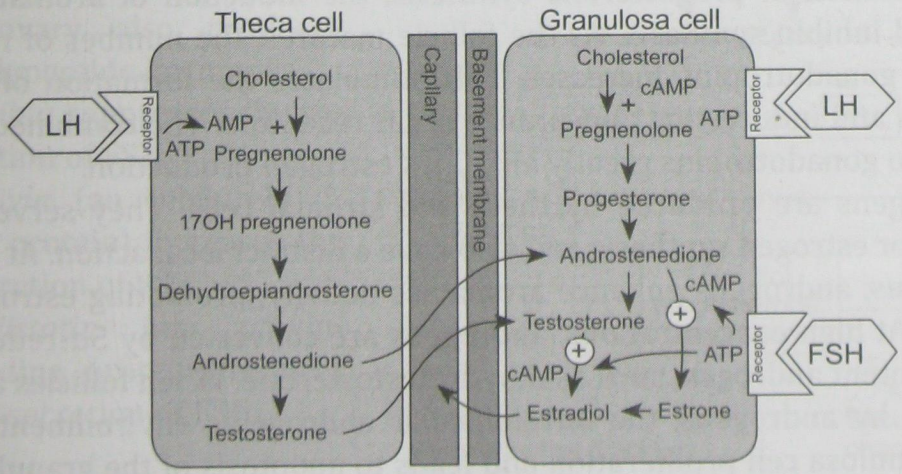


Fig. 12–13. The schematic explanation of the steroid hormones secretion by two types of the ovarian cells.

LH receptors at a relatively early stage, whereas LH receptors on granulosa cells are induced by FSH in the later stages of the maturing follicle.

The biosynthetic enzymes are differentially expressed in the two cells. Aromatase is expressed only in granulosa cells, and its activation and induction are regulated by FSH. Granulosa cells are deficient in 17α-hydroxylase and cannot proceed beyond the C-21 progestins to generate C-19 androgenic compounds (Fig. 12–13). Consequently, estrogen production by granulosa cells depends on an adequate supply of exogenous aromatizable androgens provided by theca cells. Under LH regulation, theca cells produce androgenic substrates, primarily androstenedione and testosterone, which reach the granulosa cells by diffusion. The androgens are then converted to estrogens by aromatization.

In follicles, theca and granulosa cells are exposed to different microenvironments. Vascularization is restricted to the theca layer because blood vessels do not penetrate the basement membrane. Theca cells, therefore, have better access to circulating cholesterol, which enters the cells via LDL receptors. Granulosa cells, on the other hand, primarily produce cholesterol from acetate, a less efficient process than uptake. In addition, granulosa cells are bathed in follicular fluid and exposed to autocrine, paracrine, and juxtacrine control by locally produced peptides and growth factors. "Juxtacrine" describes the interaction of a membrane- bound growth factor on one cell with its membrane-bound receptor on an adjacent cell.

FSH acts on granulosa cells by a cAMP-dependent mechanism and produces a broad range of activities, including increased mitosis and cell proliferation, the stimulation of progesterone synthesis, the induction of aromatase, and increased inhibin synthesis. As the follicle matures, the number of receptors for both gonadotropins increases. FSH stimulates the formation of its own receptors and induces the appearance of LH receptors. The combined activity of the two gonadotropins greatly amplifies estrogen production.

Androgens are produced by theca and stromal cells. They serve as precursors for estrogen synthesis and also have a distinct local action. At low concentrations, androgens enhance aromatase activity, promoting estrogen production. At high concentrations, androgens are converted by 5α-reductase to a more potent androgen, such as dihydrotestosterone. When follicles are overwhelmed by androgens, the intrafollicular androgenic environment antagonizes granulosa cell proliferation and leads to apoptosis of the granulosa cells and subsequent follicular atresia.

4. Hormonal control of female reproductive function

The female hormonal system consists of the brain, pituitary, ovaries, and reproductive tract. In the brain, the hypothalamus produces gonadotropin-releasing hormone (GnRH), which controls the secretion of LH and FSH (Fig. 12–14).

The mature ovary has two major functions: the maturation of germ cells and steroidogenesis. Each germ cell is ultimately enclosed within a follicle, a major source of steroid hormones during the menstrual cycle. At ovulation, the ovum or egg is released and the ruptured follicle is transformed into a corpus luteum, which secretes progesterone as its main product. FSH is primarily involved in stimulating the growth of ovarian follicles, while LH induces ovulation.

Both LH and FSH regulate follicular steroidogenesis and androgen and estradiol secretion, and LH regulates the secretion of progesterone from the corpus luteum. Ovarian steroids inhibit the secretion of LH and FSH with one exception: Just prior to ovulation (at midcycle), estradiol has a positive-feedback effect on the hypothalamic-pituitary axis and induces significant increases in the secretion of GnRH, LH, and FSH.

The ovary also produces three polypeptide hormones.

1. *Inhibin* suppresses the secretion of FSH.
2. *Activin* (an inhibin-binding protein) increases the secretion of FSH.
3. *Follistatin* (an activin-binding protein) reduces the secretion of FSH.

Shortly after fertilization, the embryo begins to develop placenta cells, which attach to the

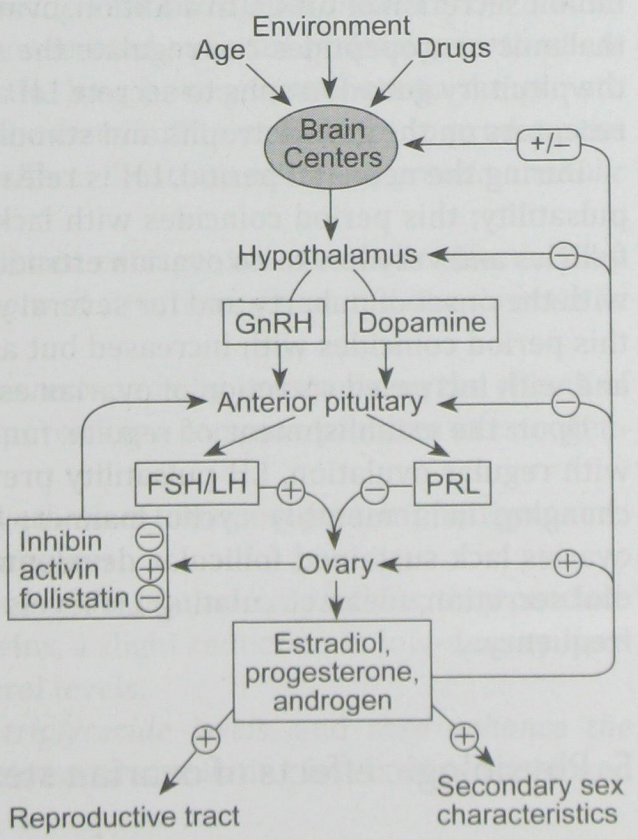


Fig. 12–14. Hormonal regulation of the female reproductive functions.

uterine lining and unite with the maternal placental cells. The *placenta* produces several pituitary-like and ovarian steroid-like hormones. These hormones support placental and fetal development throughout pregnancy and have a role in parturition. The mammary glands are also under the control of pituitary hormones and ovarian steroids, and provide the baby with immunological protection and nutritional support through lactation. Lactation is hormonally controlled by *prolactin* (PRL) from the anterior pituitary, which regulates milk production, and *oxytocin* from the posterior pituitary, which induces milk ejection from the breasts.

The hypothalamic-pituitary axis has an important role in regulating the menstrual cycle. GnRH produced in the hypothalamus and released in a pulsatile manner controls the secretion of LH and FSH through a portal vascular system.

The secretion of GnRH by the hypothalamus is regulated by neurons from other brain regions. Neurotransmitters, such as epinephrine and norepinephrine, stimulate the secretion of GnRH, whereas dopamine and serotonin inhibit secretion of GnRH. In addition, ovarian steroids and peptides and hypothalamic neuropeptides can regulate the secretion of GnRH. GnRH stimulates the pituitary gonadotrophs to secrete LH and FSH. GnRH binds to high-affinity receptors on the gonadotrophs and stimulates the secretion of LH and FSH.

During the neonatal period, LH is released at low and steady rates without pulsatility; this period coincides with lack of development of mature ovarian follicles and very low to no ovarian estradiol secretion. Pulsatile release begins with the onset of puberty and for several years is expressed only during sleep; this period coincides with increased but asynchronous follicular development and with increased secretion of ovarian estradiol.

Upon the establishment of regular functional menstrual cycles associated with regular ovulation, LH pulsatility prevails throughout the 24 hrs period, changing in a monthly cyclic manner. In postmenopausal women whose ovaries lack sustained follicular development and exhibit low ovarian estradiol secretion, mean circulating LH levels are high and pulses occur at a high frequency.

5. Physiologic effects of ovarian steroids

Estrogens are a family of structurally related female sex hormones synthesized and secreted by cells in the ovaries and, in small amounts, by cells in the adrenal cortex. Androgens can be converted to estrogens peripherally, espe-

cially in fat tissue. Three estrogens occur naturally in humans: *estrone*, *estradiol* and *estriol*. Of these, *estradiol* is the most biologically potent and the most abundantly secreted product of the ovary. Estrogens are secreted throughout the menstrual cycle. Two peaks occur: one before ovulation and the other in the middle of the luteal phase. Estrogens are transported in the blood bound to specific plasma globulins, inactivated and conjugated in the liver, and then excreted in the bile.

Estrogens are necessary for the normal female physical maturation. In concert with other hormones, estrogens provide for the reproductive processes of ovulation, implantation of the products of conception, pregnancy, parturition and lactation by stimulating the development and maintaining the growth of the accessory organs.

Estrogens stimulate the intrauterine development of vagina, uterus, and uterine tubes in the embryo and external genitalia as well as the maintenance of these organs in adulthood.

They also *stimulate the stromal development and ductal growth of the breasts* at puberty, are *responsible for the accelerated pubertal skeletal growth phase* and for closure of the epiphyses of the long bones, contribute to the growth of axillary and pubic hair, and alter the distribution of body fat to produce the typical female body contours. These effects illustrate *estrogens protein anabolic effect.*

Estrogens have a number of important extragenital metabolic effects. They are *responsible for maintaining the normal structure of skin and blood vessels* in women.

Estrogens decrease the rate of bone resorption by antagonizing the effects of calcitonin on bone (that's why osteoporosis is a common problem in estrogen-deficient postmenopausal women).

In the liver estrogens increase the synthesis of transport proteins for thyroxine, estrogen, testosterone, and other hormones.

Estrogens also affect the composition of the plasma lipoproteins: they produce an increase in high-density lipoproteins, a slight reduction in low-density lipoproteins and a reduction in cholesterol levels.

Estrogens also increase plasma triglyceride levels, and *they enhance the coagulability of blood* by effecting increased circulating levels of plasminogen and factors II, VII, IX and X.

The estrogens cause moderate retention of sodium and water. Most women retain sodium and water and gain weight just before menstruation. This occurs because the estrogens facilitate the movement of intravascular fluids into the

extracellular spaces, producing edema and increased sodium and water retention by the kidneys because of the decreased plasma volume. The actions of estrogens are summarized in Table 12–4.

Table 12–4. Physiologic effects of estrogens

General effects	Specific actions
Growth and development	
Reproductive organs	Stimulate development of vagina, uterus, and fallopian tubes in embryo and of secondary sex characteristics during puberty
Skeleton	Accelerate growth of long bones and closure of epiphyses at puberty
Reproductive processes	
Ovulation	Promote growth of ovarian follicles
Fertilization	Alter the cervical secretions to favour survival and transport of sperm Promote motility of sperm within the fallopian tubes by decreasing mucus viscosity
Implantation	Promote development of endometrial lining in the event of pregnancy
Vagina	Proliferate and stimulate vaginal mucosa secretion
Cervix	Increase mucus consistency
Breasts	Stimulate stromal development and ductal growth
General metabolic effects	
Bone resorption	Decrease rate of bone resorption
Plasma proteins	Increase production of thyroid and other binding globulins
Lipoproteins	Increase high-density and slightly decrease low-density lipoproteins

Progesterone is secreted as part of the normal menstrual cycle. The corpus luteum of the ovary secretes large amounts of progesterone after ovulation, and the adrenal cortex secretes small amounts. The hormone circulates in the blood attached to a specific plasma protein. It is metabolized in the liver and conjugated for excretion in the bile.

The local effects of progesterone on reproductive organs include the *glandular development* of the lobular and alveolar tissue of the *breasts* and the *cyclic glandular development of the endometrium.*

Progesterone cause a *decrease in sodium reabsorption*, with a resultant increase in secretion of aldosterone by the adrenal cortex (as occurs in pregnancy).

Progesterone increases basal body temperature and is responsible for the increase in body temperature that occurs with ovulation.

Smooth muscle relaxation under the influence of progesterone plays an important role in maintaining pregnancy by decreasing uterine contractions and is responsible for many of the common discomforts of pregnancy, such as edema, nausea, constipation, flatulence, and headaches.

The increased progesterone present during pregnancy and the luteal phase of the menstrual cycle enhances the ventilatory response to carbon dioxide, leading to a measurable change in arterial and alveolar carbon dioxide levels.

The normal female also produces *androgens*. Approximately 25% of these androgens are secreted from the ovaries, 25% from the adrenal cortex, and 50 % from ovarian or adrenal precursors. In the female, androgens contribute to *normal hair growth at puberty* and may have other important *metabolic effects*.

Unit VI BLOOD PHYSIOLOGY

Chapter 13. General characteristics of blood system. Physiological constants of blood. Physiology of erythrocytes. Hemoglobin. Blood groups

Blood is a highly differentiated, complex living tissue that pulsates through the arteries to every part of the body, interacts with individual cells via an extensive capillary network, and returns to the heart through the venous system. Many of the functions of blood are undertaken in the capillaries, where the blood flow slows dramatically, allowing the efficient diffusion and transport of oxygen, glucose, and other molecules across the monolayer of endothelial cells that form the thin capillary walls. In addition to transport, blood and the cells within it mediate other essential aspects of immunity and hemostasis.

I. BLOOD COMPOSITION

Blood is an opaque, red liquid that contains plasma and formed elements. If blood is collected from an individual and its coagulation prevented by adding anticoagulants to it and then centrifuged sufficiently, the blood is separated into two layers, an upper clear (i.e. free from formed elements) layer, called the *plasma* and a lower layer, where the *formed elements* are packed up.

Formed elements are so called because under microscope they have definite forms. Formed elements are *red blood corpuscles* (RBC) or erythrocytes, *white blood corpuscles* (WBC) or leukocytes, *platelets* or thrombocytes (Fig. 13–1).

On the other hand, if blood is collected without anticoagulant and is allowed to clot and some further time be allowed so that the clot is allowed to shrink, a fluid

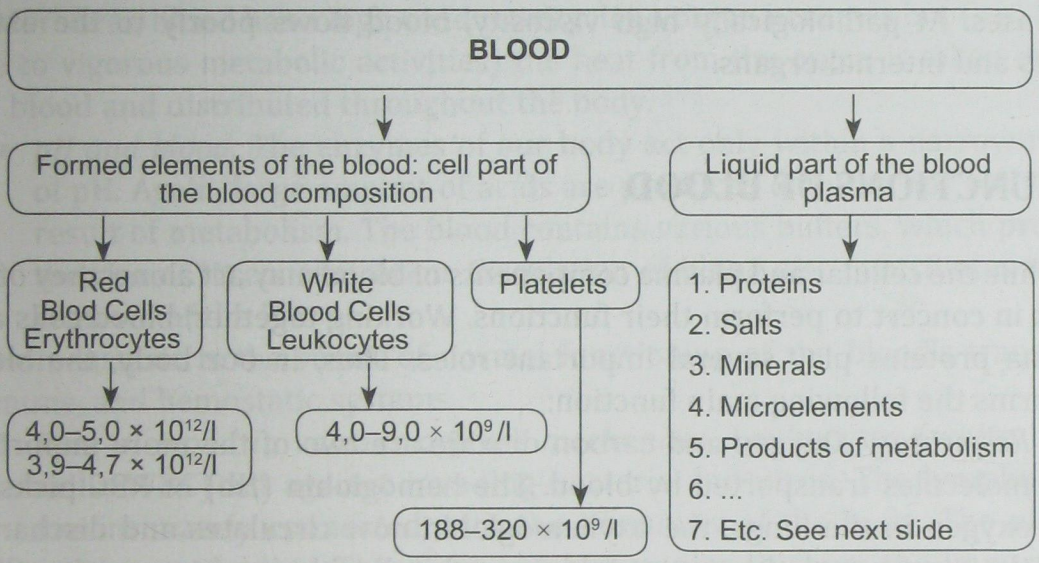


Fig. 13–1. The scheme of the blood composition.

separates out. The fluid can be centrifuged further and the clear supernatant part can be collected. This is *serum*. Basically, serum is plasma minus fibrin.

Plasma is composed mostly of water (93 %) with various dissolved solutes, including proteins, lipids (fats), carbohydrates, amino acids, vitamins, minerals, hormones, wastes, cofactors, gases and electrolytes. The solutes in plasma play crucial roles in homeostasis, such as maintaining normal plasma pH and osmolality.

Blood is normally confined to the circulation, including the heart and the pulmonary and systemic blood vessels. Blood accounts for 6 to 8 % of the body weight of a healthy adult.

The blood volume is normally 5.0 to 6.0 L in men and 4.5 to 5.5 L in women. The *density* (or *specific gravity*) of blood is approximately 1.050 g/ml. Density depends on the number of blood cells present and the composition of the plasma. The density of individual blood cells varies according to cell type and ranges from 11.15 g/L for erythrocytes to 10.70 g/L for certain leukocytes.

While blood is only slightly heavier than water, it is certainly much thicker. The *viscosity* of blood, a measure of resistance to flow, is 3.5 to 5.5 times that of water.

Blood's viscosity increases as the total number of cells present increases and when the concentration of large molecules (macromolecules) in plasma

increases. At pathologically high viscosity, blood flows poorly to the extremities and internal organs.

II. FUNCTIONS OF BLOOD

While the cellular and plasma components of blood may act alone, they often work in concert to perform their functions. Working together, blood cells and plasma proteins play several important roles. Thus, in our body, the blood performs the following main function:

1. *Respiratory.* Oxygen and carbon dioxide are two of the more important molecules transported by blood. The hemoglobin (Hb) of RBC picks up oxygen in the lungs, the oxyhemoglobin now circulates and discharges the oxygen into the tissues which need it. For this purpose, as it will be shown, the RBC and Hb are ideally designed. Further, the CO_2 produced in the tissues are discharged into the blood, it is partly carried by the plasma and partly by the Hb; the CO_2 is intimately disgorged into the lungs from where it is breathed out.
2. *Excretory.* Various waste products of the body are carried by the blood and intimately removed from the body via the kidneys.
3. *Nutritional.* End products of digestion are absorbed from the gastrointestinal tract and transported by the blood to the liver for further processing and then to the tissues who need them.
4. *Role of blood in various homeostatic processes.* Homeostasis is a steady state that provides an optimal internal environment for cell function. Briefly speaking, this is a mechanism, by which the various parameters of our body, like the internal temperature, pH, the concentrations of various substances (e.g., glucose, sodium, potassium etc.) in our body fluids remain within a narrow range.

- *Body temperature and blood.* The volume of blood is large (around 5 L in an adult man) and the "specific head" of blood is high. Therefore, a good deal amount of heat (calories) can be absorbed or lost by the blood without a great rise or fall in the temperature. It is to be noted that the enzymes in our body operate satisfactory only in a narrow range of temperature. So either an abnormal rise or an abnormal fall of temperature, damages the enzymes and ultimately stops all biological activities. In this regard blood acts as a part of our body water content.

Moreover, blood has a high conductivity. If an organ becomes hot (e.g., liver, due to vigorous metabolic activities) the heat from the organ is taken out by the blood and distributed throughout the body.

- *pH and blood.* The enzymes of our body act only within a narrow range of pH. Again, large amount of acids are produced daily, by our body, as a result of metabolism. The blood contains various buffers, which prevent the rise of H^+ concentration. Circulation of blood also contributes to the removal of the H^+.

Homeostasis is the result of normal functioning of the blood's transport, immune, and hemostatic systems.

5. *Role of blood in various defense mechanism.* Leukocytes (white blood cells; WBC) act against invading bacteria and virus. The lymphocytes are intimately connected with the immunity of the body. The plasma carries antibodies. These are some examples to show the importance of blood in connection with the defense. While the skin and mucous membranes physically restrict the entry of infectious agents, microbes constantly penetrate these barriers and continuously threaten internal infection. Blood leukocytes, working in conjunction with plasma proteins, continuously patrol for microbial pathogens in the tissues and in the blood. In most cases, penetrating microbes are efficiently eliminated by the sophisticated and elaborate antimicrobial systems of the blood.
6. *Hemostasis.* Bleeding is controlled by the process of hemostasis. Complex and efficient hemostatic mechanisms have evolved to stop hemorrhage after injury, and their failure can quickly lead to fatal blood loss. Both physical and cellular mechanisms participate in hemostasis. These mechanisms, like those of the immune system, are complex, interrelated, and essential for survival.
7. *Transport* of other substances. Besides oxygen and carbon dioxide, blood carries several important substances from one area of the body to another, including antibodies, acids and bases, ions, vitamins, cofactors, hormones, nutrients, lipids, gases, pigments, minerals, and water. Transport is one of the primary and most important functions of blood, and blood is the primary means of long-distance transport in the body. Substances can be transported free in plasma, bound to plasma proteins, or within blood cells.

Some authors advice to use the following classification of blood functions that is better according to our opinion because it's smaller and nevertheless includes all possible blood functions:

Transport of substances from one area of the body to another;
Immunity, the body's defense against disease;
Homeostasis, the maintenance of a stable internal environment;
Hemostasis, the arrest of bleeding.

III. RED BLOOD CORPUSCLES

Human RBC (erythrocyte) is a circular, biconcave cell without a nucleus and with a diameter of about 7.5 μm (Fig. 13–2). It may be considered as a kind of a living bag containing hemoglobin. This shape increases the surface area available for diffusion of oxygen and allows the cell to change in volume and shape without ruptu-ring its membrane. A cytoskeleton of pro-teins attached to the lipid bilayer provides this unique shape and flexibility. The biconcave form presents the plasma with a surface 20 to 30 times greater than if the RBC were an absolute sphere. The 500 to 1000 times more numerous than other blood cells, are the most common type of blood cell.

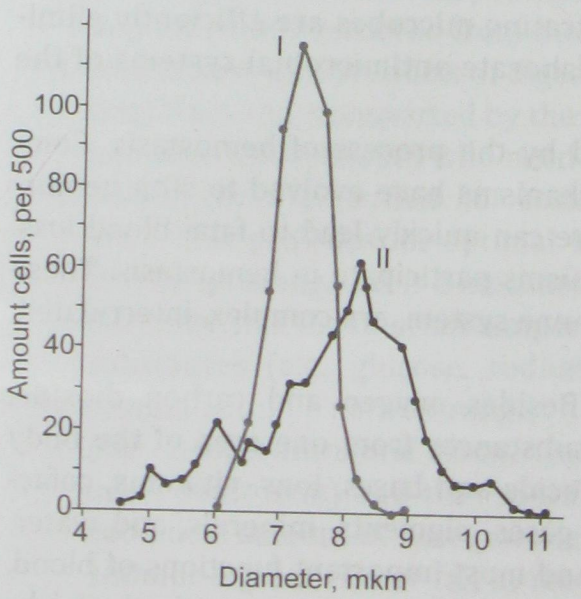


Fig. 13–2. The curve of Price-Johnce. The values of erythrocytes diameter in healthy persons [I] and patients with pernicious anemia [II].

A healthy RBC is very plastic; it can squeeze itself to pass through capillaries whose diameters are very narrow and after coming out of such a capillary it again regains its shape. Compared to its volume (about 80 m^3) it has a very big surface area (135 m^3).

Changes in RBC appearance occur in a variety of pathological conditions (Fig. 13–3). Here, it should be reasonable to stress that erythrocytes diameter dependence upon their amount in the blood depicts on the so called Price-Johnce curve that describes

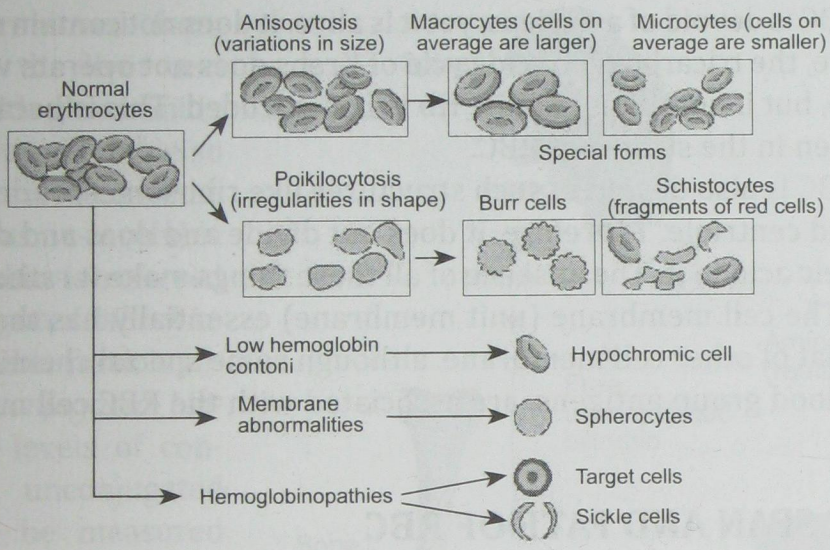


Fig. 13–3. Different pathological forms of erythrocytes' morphology.

the greater erythrocytes appearance in the blood with their average diameter equal to 7.0–7.5 μm. Larger-than-normal erythrocytes are termed *macrocytes* (such condition could be defined as right shift on the Price-Johnce curve); smaller-than normal erythrocytes are referred to as *microcytes* (left shift on the Price-Johnce curve).

Excessive variation in the size of cells is referred to as *anisocytosis. Poikilocytosis* is the presence of irregularly shaped erythrocytes. *Burr cells* are spiked erythrocytes generated by alterations in the plasma environment. *Schistocytes* are fragments of red cells damaged during blood flow through abnormal blood vessels or cardiac prostheses. The hemoglobin content of erythrocytes is also reflected in the staining pattern of cells on dried films. Normal cells appear red-orange throughout, with a very slight central pallor as a result of the cell shape.

Hypochromic cells appear pale with only a ring of deeply colored hemoglobin on the periphery. Other pathological variations in red cell appearance include *spherocytes* -small, densely staining red cells with loss of biconcavity as a result of congenital or acquired cell membrane abnormalities; and *target cells* — which have a densely staining central area with a pale surrounding area. Target cells are thin but bulge in the middle, unlike normal erythrocytes. This alteration is a consequence of *hemoglobinopathies*, mutations in the structure of hemoglobin. Target cells are observed in liver disease and after splenectomy.

The RBC is devoid of a nucleus, yet it is alive. It does not contain mitochondria (therefore, the tricarboxylic acid cycle of Krebs does not operate within it). It is full of Hb, but if an RBC is cut, the Hb is not extruded. This is because the Hb is interwoven in the stroma of RBC.

The RBC is also devoid of such structures like ribosomes, endoplasmic reticulum and centriole. Therefore, it does not divide and does and does not have any nucleic acid in it. The absence of all these things make it rather a very simple cell. The cell membrane (unit membrane) essentially has the same structure as that of other cell membrane, although some special chemical substances, like blood group antigens, are associated with the RBC cell membrane.

IV. LIFE SPAN AND FATE OF RBC

Normally, the life span of RBC is about 120 days. The process of ageing and death of an erythrocyte is as follows.

As age of the RBC increases, the enzymes which protect the erythrocyte from the damaging effects of oxygen begin to lose their efficiently, therefore oxidative damages begin to appear and the RBC becomes rather rigid, spheroidal and fragile.

Suring circulation, the erythrocytes pass through the spleen where the anatomical structure is such that RBCs have to pass through vessels whose diameters are very narrow and the fragile RBCs, as a result, are ruptured. Only the comparatively young and healthy RBCs with a sufficient degree of biconcavity ("discocyte") can survive. In diseases like spherocytosis and G-6-PD deficiency states, etc. the erythrocytes become prematurely senile and their life span is reduced.

Normally, spleen is the important slaughter house for the RBCs. But when the erythrocytes are diseased so that their fragility increases, they break down when passing through other structures also, particularly the liver.

The destruction of red blood cells is facilitated by a group of large phagocytic cells found in the spleen, liver, bone marrow, and lymph nodes. These phagocytic cells ingest the hemoglobin from the ruptured cells and break it down in a series of enzymatic reactions. During these reactions, the aminoacids from the globulin chains and iron from the heme units are salvaged and reused (Fig. 13–4).

The bulk of the heme unit is converted to bilirubin, the pigment of bile, which is insoluble in plasma and attaches to the plasma proteins for transport.

Bilirubin is removed from the blood by the liver and conjugated with glucuronide to render it water soluble so that it can be excreted in the bile. The plasma-insoluble form of bilirubin is referred to as *unconjugated bilirubin;* the water-soluble form is referred to as *conjugated bilirubin.* Serum levels of conjugated and unconjugated bilirubin can be measured in the laboratory and are reported as direct and indirect, respectively. If red cell destruction and consequent bilirubin production are excessive, unconjugated bilirubin accumulates in the blood.

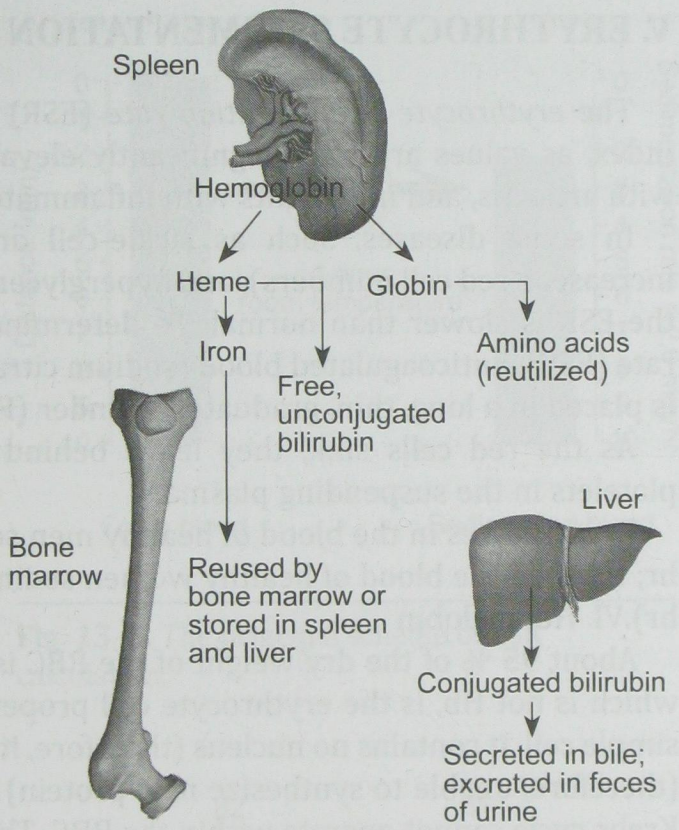


Fig. 13–4. Erythrocytes destruction and the hemoglobin fate.

This results in a yellow discoloration of the skin, called *jaundice.* When red blood cell destruction takes place in the circulation as in hemolytic anemia, the hemoglobin remains in the plasma. The plasma contains a hemoglobin-binding protein called *haptoglobin.* Other plasma proteins, such as albumin, can also bind hemoglobin.

With extensive intravascular destruction of RBC, hemoglobin levels may exceed the hemoglobin-binding capacity of haptoglobin. When this happens, free hemoglobin appears in the blood (*i.e.,* hemoglobinemia) and is excreted in the urine (*i.e.,* hemoglobinuria).

Because excessive red blood cell destruction can occur in hemolytic transfusion reactions, urine samples are tested for free hemoglobin after a transfusion reaction.

If the life span of RBCs are reduced, the erythropoiesis is stepped up in the bone marrow, which therefore compensates the increased loss of RBCs to some extent.

V. ERYTHROCYTE SEDIMENTATION RATE

The *erythrocyte sedimentation rate* (ESR) can be an important diagnostic index, as values are often significantly elevated during infection, in patients with arthritis, and in patients with inflammatory diseases.

In some diseases, such as sickle-cell anemia, polycythemia (abnormal increase in red cell numbers), and hyperglycemia (elevated blood sugar levels), the ESR is slower than normal. To determine the erythrocyte sedimentation rate (ESR), anticoagulated blood (sodium citrate is added to the blood sample) is placed in a long, thin, graduated cylinder (Fig. 13–5).

As the red cells sink, they leave behind the less dense leukocytes and platelets in the suspending plasma.

Erythrocytes in the blood of healthy men sediment at a rate of 2 to 10 mm/hr; those in the blood of healthy women sediment slightly faster (2 to 15 mm/hr).VI. Hemoglobin

About 95 % of the dry weight of the RBC is due to Hb. The part of the RBC, which is not Hb, is the erythrocyte cell proper and as already stated it is very simple cell. It contains no nucleus (therefore, it is unable to divide), no ribosome (therefore, unable to synthesize new protein) and no mitochondria (therefore, Krebs cycle cannot operate within the RBC. The ATP generating machine in the erythrocyte, therefore, is weak).

The function of the red blood cell, facilitated by the hemoglobin molecule, is to transport oxygen to the tissues. Hemoglobin also binds some carbon dioxide and carries it from the tissues to the lungs. The hemoglobin molecule is composed of two pairs of structurally different polypeptide chains determined by genes (Fig. 13–6).

Alterations in these genes can result in abnormal hemoglobins. Each of the four polypeptide chains is attached to a *heme* unit, which surrounds an atom of iron that binds oxygen. Thus, one molecule of hemoglobin can carry four molecules of oxygen. Four types of hemoglobin molecules can be found in human RBC: embryonic, fetal, and two different types found in adults (HbA, HbA_2). HbA consists of a pair of α chains and a pair of β-chains.

Therefore, all hemoglobin variations we can classify on two big groups — physiological and pathological — including the fetal one (Fig. 13–6). The peculiarities of the physiological forms are such that these hemoglobins have iron (Fe) valence +2 and can additionally attach molecule of oxygen. In case of the intoxications Fe valence becomes +3 that makes impossible further hemoglobin transport function performance — oxygen can't be attach to such hemoglobin molecule.

HbF is the predominant hemoglobin in the fetus from the 3rd through the 9th month's of gestation. Embryonic hemoglobin is found earlier in development. It consists of two α-chains and two ε-chains ($\alpha_2\varepsilon_2$). The production of ε-chains ceases at about the third month of fetal development.

The production of each type of globin chain is controlled by an individual structural gene with five different loci. Mutations, which can occur anywhere in these five loci, have resulted in the production of over 550 types of abnormal hemoglobin molecules, most of which have no known clinical significance. Mutations can arise from a single substitution within the nucleic acid of the gene coding for the globin chain, a deletion of the codons, or gene rearrangement as a result of unequal crossing over between homologous chromosomes.

The cell membrane contains some protein materials. Some such proteins, e.g. the glycophorin (which extends the through the membrane) contain the blood group an-

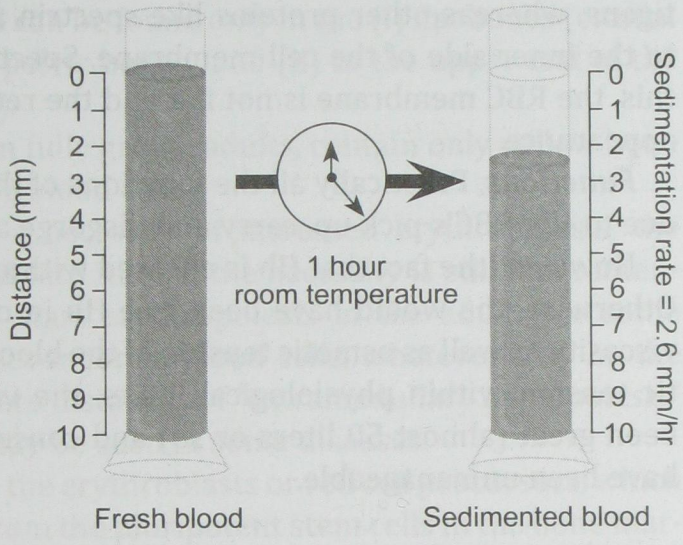


Fig. 13–5. The principal scheme of ESR calculation.

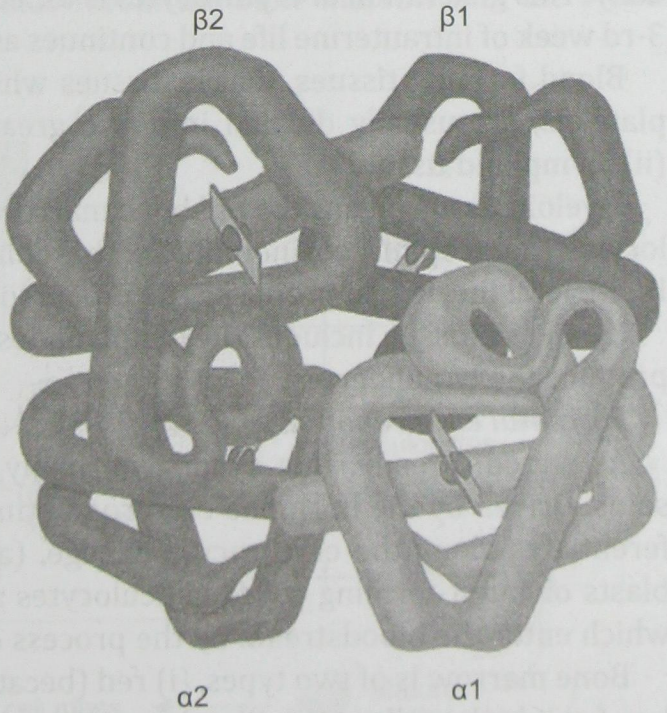


Fig. 13–6. Structure of hemoglobin molecule.

tigens, whereas other proteins like spectrin as well as actin are applied only to the inner side of the cell membrane. Spectrin is contractile and because of this, the RBC membrane is not flat and the red cell as a whole has a biconcave appearance.

Functions. Practically all the functions of the RBCs are due to its content i.e. due to Hb. RBC's pick up, carry and disgorge as well as CO_2.

However, the fact that Hb is encased within the RBC, is of great importance. Otherwise, the would have been free Hb in our blood causing a great rise in viscosity as well as osmotic tension of the blood. To keep the viscosity or osmotic tension within physiological limits, the volume of the blood would have been great (almost 50 liters or so) and consequently, size of our body would have been unmanageable.

VII. ERYTHROPOIESIS

As in our body, normally, loss of huge quantity of RBC occurs daily, to keep the RBC count and Hb % normal, sufficient number of RBCs must be produced daily. This generation of erythrocytes is called "erythropoiesis", and start in the 3-rd week of intrauterine life and continues as long as the person remains alive.

Blood forming tissues, that is, tissues which produce the RBCs, WBCs and platelets, are usually divided into two great classes, (i) Myeloid tissue, and (ii) Lymphoid tissue.

Myeloid tissue means the red bone marrow. It produces the RBCs, the granulocytes (neutrophils, eosinophils and basophils), monocytes and the platelets. In the fetal life, it also produces the precursors of the lymphocytes.

Lymphoid tissue includes the lymph nodes, the thymus and the spleen. They produce the lymphocytes.

Postnatal erythropoiesis:

As stated already, this occurs, normally, in the red bone marrow sinusoids. Driven by the hormone erythropoietin, the uncommitted stem cells differentiates along the erythrocyte lineage, (also referred to as to as erythroblasts or burst-forming cells), reticulocytes and, finally, mature erythrocytes, which enter the bloodstream by the process of diapedesis.

Bone marrow is of two types, (i) red (because it has a red color), and (ii) yellow (as it looks yellowish). Blood formation, including erythropoiesis occurs in the red bone marrow (RBM). At birth, all marrows are RBM. As age advances, some of the RBM are changed into yellow bone marrow.

At around 18 to 20 years, RBM can be found only in the (i) flat bones (cranial bones, ribs, sternum, vertebrae, pelvic bones) and (ii) in the upper end of the long bones (humerus and femur).

The shafts of the long bones, in fully grown adults, contain only yellow bone marrow (which does not produce blood cells).

Nevertheless, when there is a necessity of increased erythropoiesis, the yellow bone marrow is converted into RBM. If the necessity is still more intensive, even the liver and spleen start erythropoiesis in the adult (for some unknown reasons, the thymus never forms blood cells, whatever may be the intensity of the demand). These are instances of "extramedullary hemopoiesis" (medulla = bone marrow) and may be seen in some diseases.

The red cells are derived from the erythroblasts or red cell precursors, which are continuously being formed from the pluripotent stem cells in the bone marrow (Fig. 13–7). As they develop into mature red cells, the red cell precursors move through a series of divisions, each producing a smaller cell. Hemoglobin synthesis begins at an early stage and continues until the cell becomes a mature erythrocyte.

Nucleated red cells are normally not seen in peripheral blood because their nuclei are lost before they move from the bone marrow into the blood.

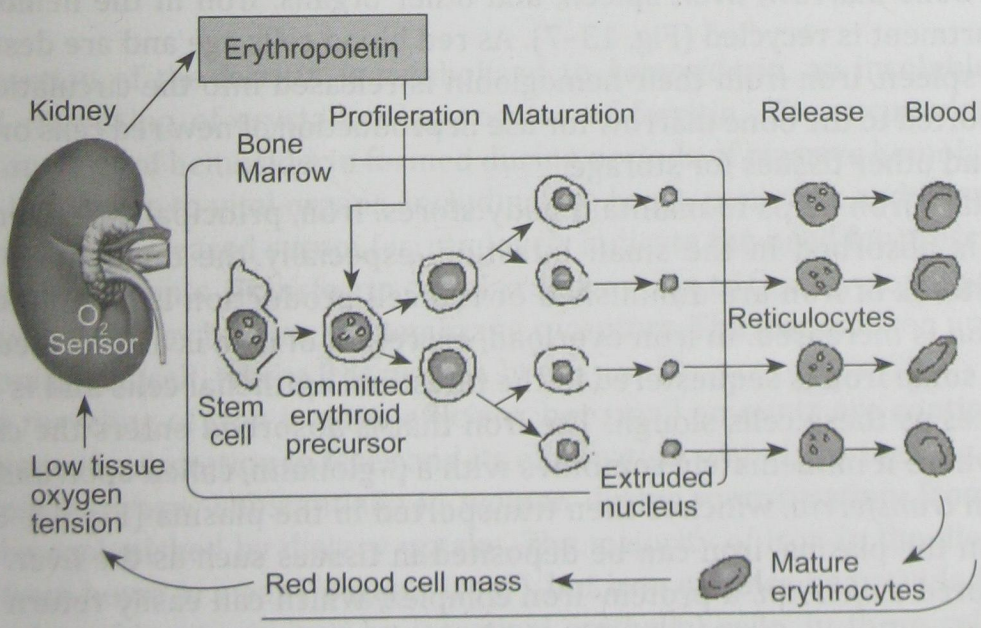


Fig. 13–7. The scheme of erythrocytes development.

However, they appear in many blood and marrow disorders, and their presence can be of diagnostic significance. One type of nucleated red cell, the *normoblast* (Fig. 13–10), is seen in several types of anemias, especially when the marrow is actively responding to demand for new erythrocytes. In seriously ill patients, the appearance of normoblasts in peripheral blood is a grave prognostic sign preceding death, often by several hours. Another nucleated erythrocyte, the megaloblast, is seen in peripheral blood in pernicious anemia and folic acid deficiency.

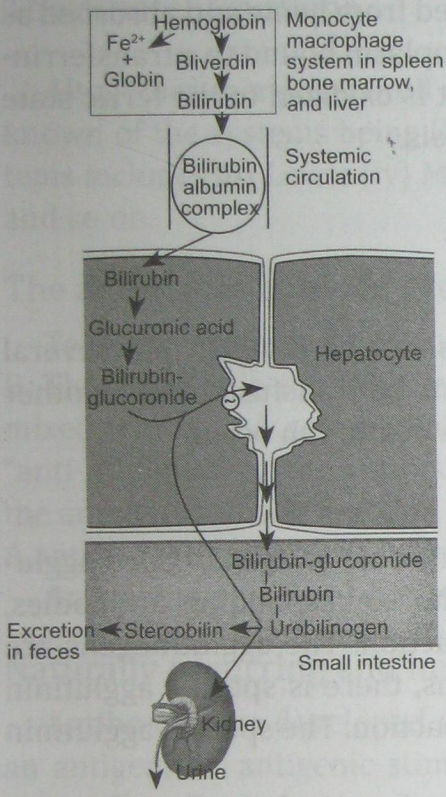
VIII. IRON FULL CYCLING IN RBC METABOLISM

The rate at which hemoglobin is synthesized depends on the availability of iron for heme synthesis. A lack of iron results in relatively small amounts of hemoglobin in the red blood cells. Most of the iron needed for new hemoglobin synthesis is obtained from the heme of senescent red cells.

The amount of iron in the body is approximately 35 to 50 mg/kg of body weight for males and less for females. Body iron is found in several compartments. Most of the functional iron (80 %) is found in hemoglobin, with small amounts being found in the myoglobin of muscle, the cytochromes, and iron-containing enzymes such as catalase. Approximately 15 % to 20 % is stored in the bone marrow, liver, spleen, and other organs. Iron in the hemoglobin compartment is recycled (Fig. 13–7). As red blood cells age and are destroyed in the spleen, iron from their hemoglobin is released into the circulation and transported to the bone marrow for use in production of new red cells or to the liver and other tissues for storage.

Dietary iron helps to maintain body stores. Iron, principally derived from meat, is absorbed in the small intestine, especially the duodenum. When body stores of iron are diminished or red cell production is stimulated, absorption is increased. In iron overload, excretion of iron is accelerated. Normally, some iron is sequestered in the intestinal epithelial cells and is lost in the feces as these cells slough. The iron that is absorbed enters the circulation, where it immediately combines with a β-globulin, called *apotransferrin,* to form *transferrin*, which is then transported in the plasma (Fig. 13–8).

From the plasma, iron can be deposited in tissues such as the liver. There it is stored as *ferritin,* a protein–iron complex, which can easily return to the circulation. Serum ferritin levels, which can be measured in the laboratory, provide an index of body iron stores.



IRON RECYCLING

- Most of the iron needed for new hemoglobin synthesis is obtained from the heme of senescent red cells.
- Iron released by macrophages is transported in the ferric state in plasma bound to the iron transporting protein *transferrin*
- Cells that need iron (e.g.. for heme synthesis) possess membrane receptors to which transferrin binds. The receptor-bound transferrin is then internalized.
- The iron is released, reduced intracellular^ to the ferrous state, and either incorporated into heme or stored as *ferritin*
- Iron is also stored as ferritin by macrophages in the liver
- A portion of the ferritin is catabolized to hemosiderin, an insoluble compound consisting of crystalline aggregates of ferritin.
- The accumulation of large amounts of *hemosiderin* formed during periods of massive hemolysis can result in damage to vital organs, including the heart, pancreas, and liver.
- The recycling of iron is quite efficient, but small amounts are continuously lost. Iron loss increases substantially in women during menstruation. *Iron stores must be replenished by dietary uptake.*

Fig. 13–8. The scheme of iron recycling in the organism. Explanation see in the text.

A portion of the ferritin is catabolized to *hemosiderin*, an insoluble compound consisting of crystalline aggregates of ferritin. The accumulation of large amounts of hemosiderin formed during periods of massive hemolysis can result in damage to vital organs, including the heart, pancreas, and liver.

Clinically, decreased serum ferritin levels indicate the need for prescription of iron supplements. Transferrin can also deliver iron to the developing red cell in bone marrow by binding to membrane receptors. The iron is taken up by the developing red cell, where it is used in heme synthesis.

The recycling of iron is quite efficient, but small amounts are continuously lost due to the excretion in feces and its elimination with the urine (Fig. 13–7). Iron loss increases substantially in women during menstruation. Iron stores must be replenished by dietary uptake. The majority of iron in the diet is derived from heme in meat ("organic iron"), but iron can also be provided by the absorption of inorganic iron by intestinal epithelial cells. In these cells, iron attached to heme is released and reduced to the ferrous form (Fe^{2+}) by intracel-

lular flavoprotein. The reduced iron (both released from heme and absorbed as the inorganic ion) is transported through the cytoplasm bound to a transferrin-like protein. When it is released to the plasma, it is oxidized to the ferric state and bound to transferrin for use in heme synthesis.

IX. BLOOD GROUPS

On the basis of blood groups, human beings can be divided into several categories. If blood from one category of persons be transfused into another category, mismatching, which may even result in death, can occur.

Agglutinogens and agglutinins

Normally, the erythrocytes of a person contain some antigens, called agglutinogens, which when allowed to react with the corresponding antibodies, called agglutinins, produce antigen-antibody reaction. The agglutinogens are of different kinds. For each type of agglutinogens, there is specific agglutinin which alone will produce the antigen-antibody reaction. The specific agglutinin is often called the equivalent antibody.

The antigen-antibody reaction consists of clumping or agglutination of the erythrocytes. A single molecule of agglutinin simultaneously binds with several (either two or more) molecules of agglutinogen present of different erythrocytes and thus produce the clumping of RBCs.

The agglutinated erythrocytes may be subsequently broken down in the spleen or liver and intimately removed by the tissue macrophages, or, there is a complement system present in plasma, are lysed within the blood vessels immediately and the hemoglobin thus released, circulates in the plasma producing various hazards. Therefore, after-a-mismatched transfusion, intravascular hemolysis does not necessary always follow.

The agglutinogens are present on the surface of the erythrocytes and are chemically glycoproteins. The agglutinins are plasma globulins and are either IgM or IgA type of immunoglobulins.

Agglutinogens present not only in the RBCs' but also in various secretes (like the gastric and salivary). The antigens obtained from the gastric/salivary secretions are called soluble antigens. They are produced by their corresponding genes. A large number of such antigens have been discovered until now.

The different systems

Altogether there are, until now, 21 systems of blood groups, the most well known of the systems being (i) ABO system and the (ii) Rh system. Other systems include (iii) Lewis (iv) MNS (v) P (vi) Kell (vii) Duffie and (viii) Lutheran, and so on.

The antibodies of blood group

Take the example of ABO system. In ABO system, the antigens are A and B. Therefore, if in the RBCs' of a person, the antigen antibody reaction, when mixed with anti A serum, that is, a serum which contains the antibody called "anti-A". Similarly, the antibody called "anti-B" reacts against B antigen. When the antigen antibody reaction (e.g., between red cells having A antigen and anti A antibody) occurs, the red cells are agglutinated.

Anti-A and anti-B antibodies are called "naturally occurring" antibodies.

Naturally occurring and immune antibodies

Antibodies are developed only when the body is challenged (provoked) by an antigen, i.e. antigenic stimulus leads to antibody production. Thus, when virus of small pox (antigen) are introduced to our body (e. g., the body is 'challenged' by the small pox virus), the body responds by generating small pox antibodies. This is a general statement.

However, there are some exceptions. The anti-A and anti-B (and many other antibodies of the different blood groups, but not the antibody of Rh system) are produced by the A or B (or any other corresponding) antigen. Thus, these examples of naturally occurring antibodies.

On the other hand, the antibodies against the antigen of Rh system develops only after the body has been exposed (=challenged) to the antigen of the Rh system; it is thus an "immune antibody".

The pregnant mother, but not the fetus, can produce antibody. This is because, to become immunologically competent, the precursors of B-lymphocytes (the ultimate produces of immunoglobulins) must be properly processed. This processing continues in most periods of fetal life and becomes complete only when the pregnancy is near term.

The concentration of the immunoglobulins are very low in the fetus or in the newborn. Now, many naturally occurring foods contain A or B antigens; many substances, normally inhaled by the new born, also contain such antigens. It is possible that the new born, through inhalation of food, receives some doses

of such antigens and as a result begin to generate the antibodies. However, as there is no unanimity of opinion about this, the term naturally occurring antibody should be retained.

Blood transfusion

The classical indication of blood transfusion is hemorrhagic shock due to, say, postpartum hemorrhage, hematemesis etc. In severe anemia, transfusion of erythrocytes rather than whole blood is indicated.

For transfusion purposes, blood is collected from health donor (volunteer or professional), grouped and stored in blood bank for future use. Although, today's society is much more conscious of blood banks, the need of blood outstrips the supply; this is perhaps a global phenomenon.

Before transfusion, grouping of blood of the recipient as well as cross matching of bloods between the donor and the recipient should be made. It is essential that hazards of blood transfusion is borne in mind by the attending clinician.

Hazards of transfusion may be divided into two groups (i) those due to mismatching, and (ii) those due to causes other than mismatching.

Mismatched transfusion, as already stated, causes agglutination of donor's erythrocytes. Dangerous hemolytic transfusion reactions occur when blood is transfused into an individual with an incompatible blood type, ie, an individual who has agglutinins against the red cells in the transfusion. The plasma in the transfusion is usually so diluted in the recipient that it rarely causes agglutination even when the titer of agglutinins against the recipient's cells is high. However, when the recipient's plasma has agglutinins against the donor's red cells, the cells agglutinate and hemolyze. Free hemoglobin is liberated into the plasma. The released hemoglobin (as well as the non-hemolysed but clumped RBCs') may block smaller blood vessels in the coronary/pulmonary/renal circulation, or it may appear in the renal tubules. If the amount of hemoglobin in the renal tubules is sufficiently high, kidney failure can develop (in the presence of other contributing factors). Renal failure may be caused due to the .associated shock also. However, in absence of vascular changes, Hb perhaps does not precipitate in sizeable amounts in the renal tubules.

Reactions which are not due to mismatching: Excess volume of transfusion can cause hypervolemy and cardiopulmonary embarrassment. Donor's platelets are often broken down releasing serotonin which may cause pulmonary edema. Sometimes anaphylactic shock like condition may supervene due to release of large amount of histamine from the donor's basophils. Pyrogenic reactions may also develop.

Therefore, the immediate signs and symptoms may be fever, chest pain, anuria, dyspnoe or severe fall of blood pressure. It should be remembered that with modern facilities and consciousness, mismatching of blood has become very rare.

If the patient survives, jaundice appears afterwards. An altogether different kind of hazard is the development of infections from the donor's blood. Thus, malaria, syphilis or viral hepatitis and nowdays, the AIDS, may be transmitted through the donor's blood.

Each bottle of blood transfused, contains enough citrate so that during transfusion, particularly when it is heavy in amount, good quantity of citrate is introduced which removes the calcium ions of the recipient's plasma. Very rarely, particularly, in cases of liver failure, this may produce tetany and incoagulability of blood because of Ca^{2+} deficiency.

X. RHESUS FACTOR

Aside from the antigens of the ABO system, those of the Rh system are of the greatest clinical importance. The "Rh factor", named for the rhesus monkey because it was first studied using the blood of this animal, is a system composed of many antigens. D is by far the most antigenic, and the term "Rh- positive" as it is generally used means that the individual has agglutinogen D. The "Rh- negative" individual has no D antigen and forms the anti-D agglutinin when injected with D-positive ceils. The Rh typing serum used in routine blood typing is anti-d serum. Eighty-five percent of Caucasians are D-positive and 15 % are D-negative; over 99 % of Orientals are D-positive. D-negative individuals who have received a transfusion of D-positive blood (even years previously) can have appreciable anti-D titers and thus may develop transfusion reactions when transfused again with D-positive blood.

Hemolytic Disease of the Hewborn

Another complication due to "Rh incompatibility" arises when an Rh- negative mother carries an Rh-positive fetus. Briefly, small amounts of fetal blood penetrates into the maternal circulation at the time of delivery, and some mothers develop significant titers of anti-Rh agglutinins during the postpartum period. During the next pregnancy, the mother's agglutinins cross the placenta to the fetus. For details see the next paragraph.

Chapter 14. Protective functions of blood. Physiology of leukocytes. Immunity and its types

Among all, blood cells include *leukocytes* (*white blood cells*). Each microliter (a millionth of a liter) of blood contains from 4,500 to 10,000 leukocytes. The leukocytes, constitute only 1 % of the total blood volume. They originate in the bone marrow and circulate throughout the lymphoid tissues of the body. There they function in the inflammatory and immune processes.

There are several subtypes of leukocytes, defined by morphological differences, each with vastly different functional characteristics and capabilities. They include the granulocytes and the agranulocytes (Fig. 14–1).

The granulocytes are divided into three types (neutrophils, eosinophils, and basophils) according to the staining properties of the granules. They are all phagocytic cells and are identifiable because of their cytoplasmic granules.

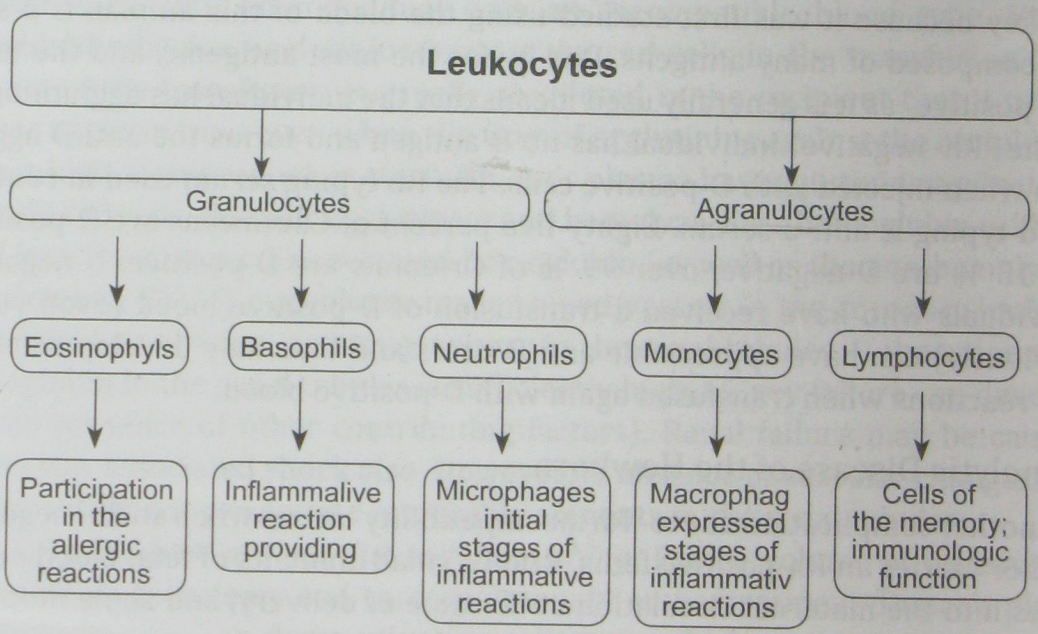


Fig. 14–1. The scheme of the leukocytes classification. The lowest row indicates function of the leukocytes subgroup.

These white blood cells are spherical and have distinctive multilobar nuclei (Fig. 14–2).

Of the total leukocytes, 40 to 75 % are neutrophilic, polymorphonuclear (multi-nucleated) cells, otherwise known as **neutrophils**. These phagocytic cells actively ingest and destroy invading microorganisms. These cells have granules that are neutral and thus do not stain with an acidic or a basic dye. They are usually participated in the initial steps of the inflammative reactions in the organism (Fig. 14–1).

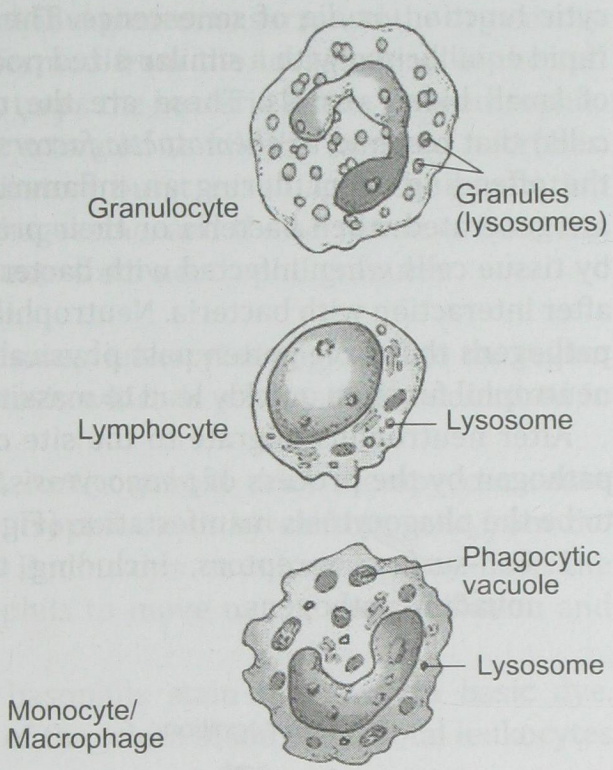


Fig. 14–2. White blood cells — granulocyte, lymphocyte, and monocyte.

The *neutrophils* are primarily responsible for maintaining normal host defenses against invading bacteria, fungi, products of cell destruction, and a variety of foreign substances. The cytoplasm of mature neutrophils contains fine granules. These granules contain degrading enzymes that are used in destroying foreign substances and correspond to lysosomes found in other cells. Enzymes and oxidizing agents associated with these granules are capable of degrading a variety of natural and synthetic substances, including complex polysaccharides, proteins, and lipids. The degradative functions of the neutrophil are important in maintaining normal host defenses and in mediating the inflammatory response.

The *neutrophils* have their origin in the myeloblasts that are found in the bone marrow. The myeloblasts are the committed precursors of the granulocyte pathway and do not normally appear in the peripheral circulation. When they are present, it suggests a disorder of blood cell proliferation and differentiation.

After release from the marrow, the neutrophils spend only about 4 to 8 hrs in the circulation before moving into the tissues. Their survival in the tissues lasts about 4 to 5 days. They die in the tissues while discharging their phago-

cytic function or die of senescence. The pool of circulating neutrophils is in rapid equilibrium with a similar-sized pool of cells marginating along the walls of small blood vessels. These are the neutrophils (amoeba-like phagocytic cells) that respond to *chemotactic factors* and migrate into the tissues toward the offending agent during an inflammatory response. Chemotactic factors are generated when bacteria or their products bind to circulating antibodies, by tissue cells when infected with bacteria, and by lymphocytes and platelets after interaction with bacteria. Neutrophils provide an efficient defense against pathogens that have gotten past physical barriers such as the skin. Defects in neutrophil function quickly lead to massive infection — and, quite often, death.

After neutrophils migrate to the site of infection, they engulf the invading pathogen by the process of *phagocytosis*. The following stages are considered to be the phagocytosis manifestation (Fig. 14–3).

1. Cell-surface receptors, including those for exposed opsonins, sense invading pathogens.

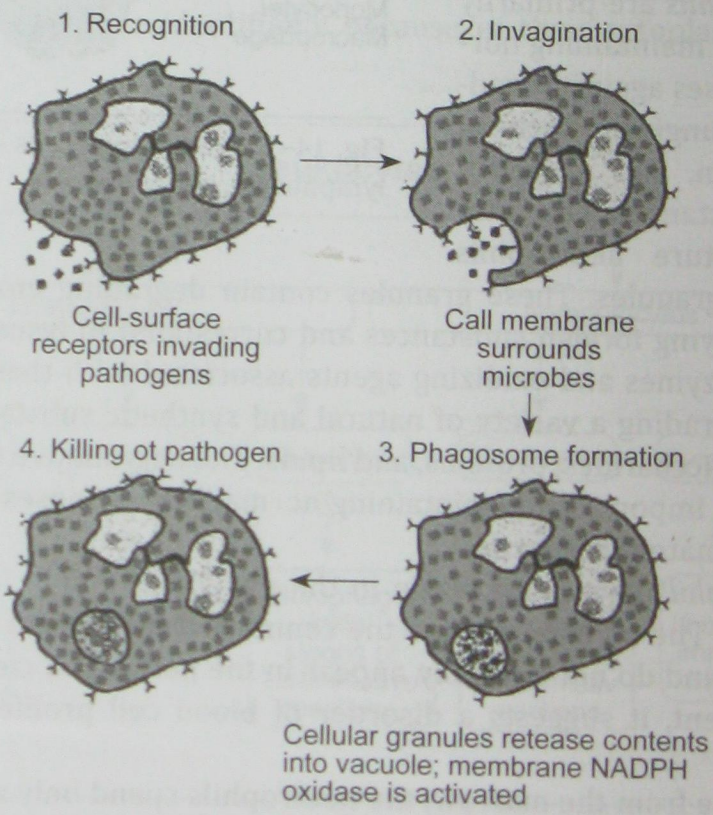


Fig. 14–3. Steps in phagocytosis and intracellular killing by neutrophils.

2. The neutrophil plasma membrane invaginates to surround the organisms.
3. A membrane-bounded vesicle formed from the invagination of the cell membrane, called a phagosome, traps the bacteria inside the neutrophil.
4. Potent metabolic processes are activated to kill the ingested microbes, including activation of the respiratory burst, resulting in the generation of potent oxidants within the phagosome, and the secretion of bacteria-killing enzymes into the phagosome from neutrophil granules.

Briefly, when bacteria invade the tissues, leukocytes migrate into the region and destroy the bacteria by phagocytosis which is schematically described in the Fig. 14–4.

Epinephrine, exercise, stress, and corticosteroid drug therapy can cause rapid increases in the circulating neutrophil count by shifting cells from the marginating to the circulating pool. Endotoxins or microbes may have the opposite effect, by attracting neutrophils to move out of the circulation and into the tissues.

Basophils. The granules of the basophils stain blue with a basic dye. These cells constitute les than 1.0 % of the white blood cells. Total leukocytes

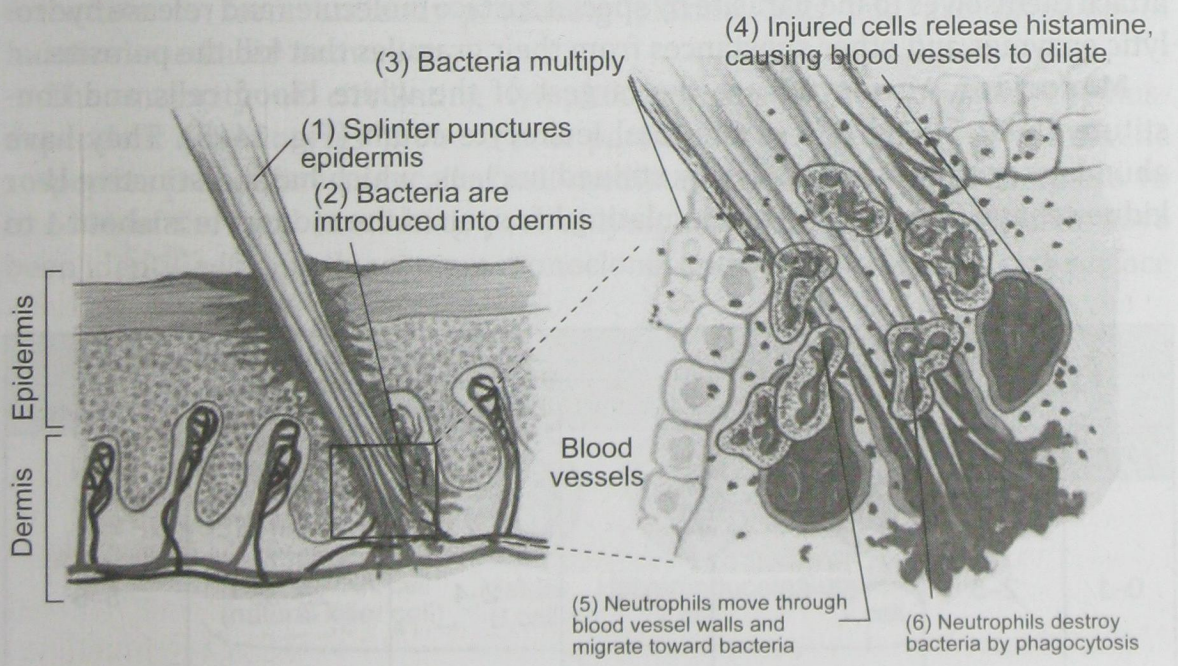


Fig. 14–4. The example of neutrophils participation in phagocytosis in case of bacterial contamination.

percentage including all their subtypes amount defines the leukocytes formula (Fig. 14–5).

The basophils in the circulating blood are similar to the large mast cells located immediately outside the capillaries in body tissues. Both the basophils and mast cells release heparin, an anticoagulant, into the blood. The mast cells and basophils also release histamine, a vasodilator and other inflammatory mediators. The release of these and other mediators by basophils increases regional blood flow, facilitating the transport of other leukocytes to areas of infection and allergic reactivity or other forms of hypersensitivity. The mast cells and basophils play an exceedingly important role in edema formation during the inflammative reactions (Fig. 14–1).

Eosinophils. The cytoplasmic granules of the eosinophils stain red with the acidic dye eosin. These leukocytes constitute 1 % to 5 % (average — 2–4 %) of the total number of white blood cells and increase in number during allergic reactions (Fig. 14–1). Like neutrophils, eosinophils migrate to sites where they are needed and exhibit a metabolic burst when activated. They are thought to release enzymes or chemical mediators that detoxify agents associated with allergic reactions. Eosinophils are also involved in parasitic infections. Although most parasites are too large to be phagocytized by eosinophils, the eosinophils attach themselves to the parasite by special surface molecules and release hydrolytic enzymes and other substances from their granules that kill the parasite.

Monocytes. Monocytes are the largest of the white blood cells and constitute about 5 % to 8 % of the total leukocyte count (Fig. 14–5). They have abundant cytoplasm and a darkly stained nucleus, which has a distinctive U or kidney shape (Fig. 14–2). The circulating life span of the monocyte is about 1 to

Granulocytes					Agranulocytes	
Neutrophils			Basophils	Eosinophils	Lymphocytes	Monocytes
Young	Stick-like	Segmentar-like				
0–1	2–5	55–68	0–1	2–4	23–35	5–8

Fig. 14–5. Leukocytes formula in % pertaining to the totall leukocytes content.

3 days, three to four times longer than that of the granulocytes. These cells survive for months to years in the tissues. The monocytes, which are phagocytic cells, are often referred to as macrophages when they enter the tissues. The monocytes engulf larger and greater quantities of foreign material than do the neutrophils. These leukocytes play an important role in chronic inflammation and are also involved in the immune response by activating lymphocytes and by presenting antigen to T-cells. When the monocyte leaves the vascular system and enters the tissues, it functions as a macrophage with specific activity. The macrophages are known as histiocytes in loose connective tissue, microglial cells in the brain, and Kupffer's cells in the liver.

Lymphocytes. The lymphocytes have their origin in the lymphoid stem cells that are found in the bone marrow. The lymphocytes constitute 23 % to 35 % of the white blood cell count (Fig. 14–5). They have no identifiable granules in the cytoplasm and are referred to as agranulocytes. They move between blood and lymphoid tissues, where they may be stored for hours or years. Their function in the lymph nodes or spleen is to defend against foreign microbes in the immune response.

There are two types of lymphocytes: *B-lymphocytes* and *T-lymphocytes* (Fig. 14–6). The *B-lymphocytes* (20 to 30 % of circulating lymphocytes) differentiate to form antibody-producing plasma cells and are involved in humoral-mediated immunity.

The majority of circulating lymphocytes are *T-lymphocytes* ("thymus-dependent lymphocytes"). These cells participate in certain types of immune responses that define as socalled cell-mediated immunity. T-cells comprise 40 to 60 % of the total circulating pool of lymphocytes. Subtypes of T-cells have been identified using fluorescent monoclonal antibodies to specific cell-surface

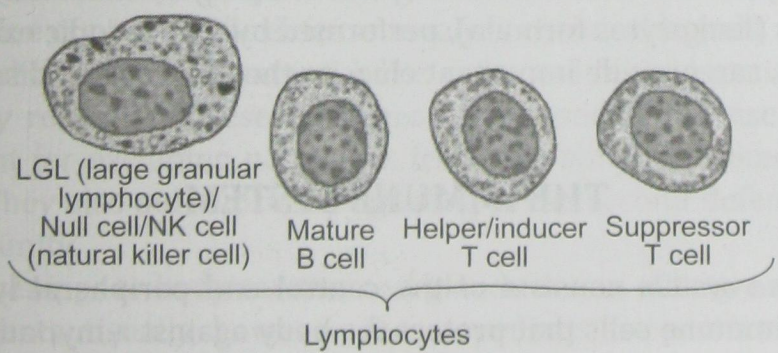


Fig. 14–6. Lymphocytes and their subtypes.

antigens, known as CD antigens. All T-cells possess the common CD3 antigen. So-called *helper T-cells* possess the CD4 antigen cluster, while *suppressor T cells* lack CD4 but possess CD8. Patients with AIDS show decreased circulating levels of CD4-positive cells.

Natural killer (NK) cells are T-lymphocytes that possess the ability to kill tumor cells without prior exposure or priming.

Lymphocytes not characteristic of either T-cells or B-cells are called *null cells.* The entire scope of the function of null cells, which comprise only 1 to 5 % of circulating lymphocytes, is unknown, but it has been established that null cells are capable of destroying tumor cells and virus-infected cells.

While B-cells mediate immune responses by releasing antibody, T-cells often exert their effects by synthesizing and releasing *cytokines*, hormone-like proteins that act by binding specific receptors on their target cells. Recent research has led to the discovery of many cytokines, with activities ranging from tumor destruction, a function of tumor necrosis factor, to the promotion of blood cell production. Cytokines that limit viral replication in cells, known as interferons, suppress or potentiate the function of T-cells, stimulate macrophages, and activate neutrophils. In some cases, cytokines, like other hormones, can exert potent effects when supplied exogenously. The biological diversity and potency of the cytokines has opened the door to the development of a variety of new pharmacological agents that have proved useful in the treatment of cancer, immune disorders, and other diseases.

The number and relative proportion of the leukocyte subtypes can vary widely in different disease states. For example, the absolute neutrophil count often increases during infection, presumably in response to the infection. Eosinophil counts increase when allergic individuals are exposed to allergens. Lymphocyte counts decrease in AIDS and during some other viral infections. For this reason, in addition to a *blood cell count,* a *differential analysis* of leukocyte subtypes (leukocytes formula), performed by microscopic examination of stained slides, can provide important clues to the diagnosis of disease.

THE IMMUNE SYSTEM

The *immune system* consists of the central and peripheral lymphoid tissues and the immune cells that protect the body against a myriad of microbes and foreign substances. The individual components of the substance that the immune system recognizes as foreign are called *antigens.* The interaction of the

collective and coordinated components of the immune system and the antigens of a foreign agent is called the *immune response.*

The body protects again bacteria, viruses, and other foreign substances by many different mechanisms. These include physical barriers, phagocytic cells in the blood and tissues, a class of lymphocytes called natural killer cells, and various blood-borne molecules that protect individuals from a potentially harmful environment. These mechanisms can be divided into two cooperative and tightly interrelated defense systems: the *nonspecific* or *innate immune system* and the *specific* or *acquired [adaptive] immune system.* Elements of the innate or natural immune system include *exterior defenses*, such as skin and mucous membranes; *phagocytic leukocytes*; and serum proteins, which act nonspecifically and quickly against microbial invaders.

Microbes that escape the onslaught of cells and molecules of the innate immune system face destruction by T-cells and B-cells of the adaptive immune system. Activation of the adaptive immune system results in the generation of antibodies and cells that specifically target the inducing organism or foreign molecule. Unlike the innate system, adaptive or acquired immune responses develop gradually but exhibit memory. Therefore, repeat exposure to the same infectious agent results in improved resistance mediated by the specific aspects of the adaptive immune system. Working together, elements of the innate and adaptive immune systems provide a considerable obstacle to the establishment and long-term survival of infectious agents.

Nonspecific Immunity. As a first line of defense system, the nonspecific immune system distinguishes self from non-self but does not distinguish one type of pathogen from another. Nonspecific resistance to microbe invasion results from two general lines of defense. Microorganisms encounter the first line of resistance on exposure to the epithelial layers of our skin and mucous membranes that line our respiratory, gastrointestinal, and urogenital tracts. The second line of nonspecific defense involves chemical signals, antimicrobial substances, phagocytic and natural killer cells, and fever associated with the inflammatory response. These two lines of nonspecific defense mechanisms are important for excluding pathogens from our body and removing them if they enter. They also aid in proper signaling of the second defense system — specific immunity.

Natural openings to body cavities and glands are an effective entry point of infectious agents. Usually, however, these openings are protected from invasion by pathogens in at least two ways. First, they are coated with mucus and other secretions that contain secretory immunoglobulins as well as antibacte-

rial enzymes, such as lysozyme. Second, organisms that invade these openings cannot easily reach the blood but, instead, lodge in an organ that communicates with both the exterior and the interior of the body, such as a lung or the stomach. Many pathogens cannot survive the low pH of stomach acid. In the lungs, organisms face the efficient phagocytic activity of alveolar macrophages. These cells, derived from blood monocytes, are mobile but confined to the pulmonary capillary network. As efficient phagocytic cells, they continuously patrol the pulmonary vasculature to remove inhaled microbes.

Microbes that successfully break through these physical barriers face destruction by the fixed macrophages of the monocyte-macrophage system. These cells line the sinusoids and vasculature of many organs, including the liver, spleen, and bone marrow. The nonmobile, fixed phagocytic macrophages efficiently remove foreign particles, including bacteria, from the circulation.

Specific Immunity. Specific or acquired immunity develops during an individual's lifetime, distinguishes self from non-self, and responds specifically to different pathogens and foreign molecules. White blood cells called *lymphocytes* are key players in the specific or acquired immune response. These cells include the T-lymphocytes, which participate in cell-mediated immunity, and the B-lymphocytes, which participate in humoral immunity. Cell-mediated immunity involves the production of cytotoxic T cells, which have the ability to destroy antigen-bearing cells. Humoral immunity is characterized by the transformation of B cells into plasma cells, which secrete immunoglobulins (antibodies) that have specific activity against the inciting antigen.

The adaptive immune system can be considered at three levels:

1. *The afferent arm*, which gives the system its remarkable ability to recognize specific antigenic determinants of a wide range of infectious agents;
2. *The efferent arm*, which supplies a cellular and molecular assault on the invading pathogens;
3. *Immunological memory*, which specifically accelerates and potentiates subsequent responses to the same activating agent or antigen.

The specificity of the recognition, effector, and memory aspects of the adaptive immune system derives from the specificity of antibody molecules as well as that of receptors on T-cells and B-cells. The lymphocytes of the immune system are capable of recognizing and specifically responding to hundreds of thousands of potential antigens, which may be presented, for example, as

glycoproteins on the surface of bacteria, the coat protein of viruses, microbial toxins, or membranes of infected cells.

Only a few circulating lymphocytes need to recognize an individual antigen initially. This initial recognition induces proliferation of the responsive cell, a process known as *clonal selection* (Fig. 14–7). Clonal selection amplifies the number of specific T-cells or B-cells.

While all of the cells generated after a single clone has expanded are specific for the inducing antigen, they may not all possess the same functional characteristics. Some of the daughter lymphocytes may be effector cells. For example, when B-cells are activated, their progeny *plasma cells* are capable of generating antibodies. Other progeny in the expanded clone may play an afferent recognition role and, thereby, function as *memory cells*.

The increased number of these cells, which mimic the reactive specificity of the original lymphocytes that responded to the antigen, accelerate responsiveness when the antigen is encountered again. Memory cells thus account for one of the primary tenets of immunity: Resistance is increased after initial exposure to the infectious agent. Long-term immunity to many viruses — such as influenza, measles, smallpox, and polio — can be induced by *vaccination* with a killed or mutant form of the pathogen.

It should be mentioned, the innate and adaptive immune systems work together in ways that obscure their differences. Indeed, consideration of these two systems as distinct, individual entities

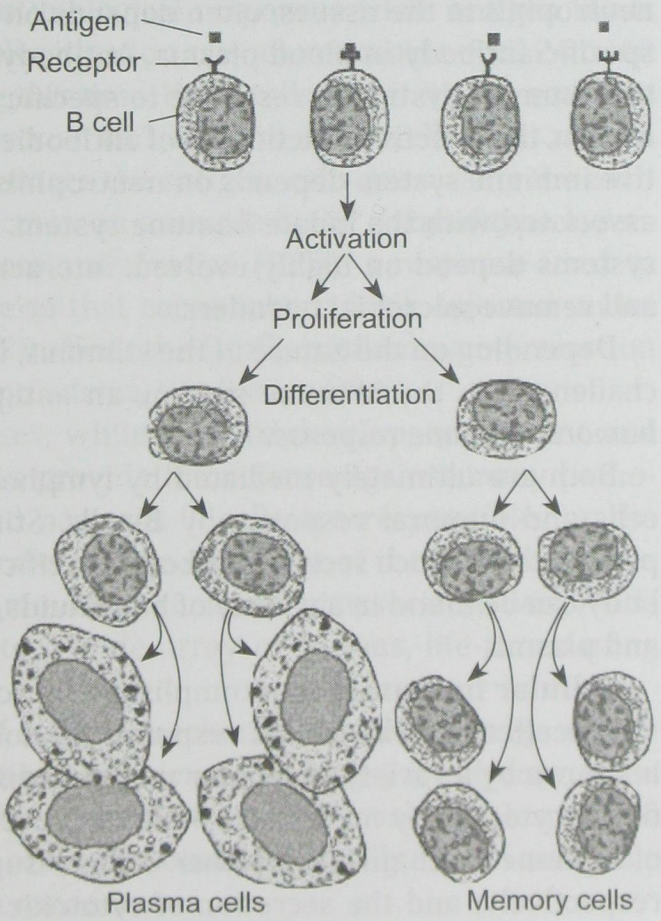


Fig. 14–7. Clonal selection of the committed lymphocytes.

is neither justified nor correct, owing to their extensive interdependence. They are described individually only as an aid to their presentation. In this respect, it is important to define the characteristics that differentiate each system.

In general, responses of the innate immune system are neither *specific* nor *inducible*; that is, the response is not programmed by or directed against a specific pathogen and is not amplified as a result of previous encounters with the pathogen. The adaptive response, in contrast, is both specific and inducible; the response is set in motion by a particular pathogen and develops against that specific pathogen.

While characteristics of the innate and adaptive immune system differ, each system depends on elements of the other for optimal functioning. The initiation of responses by the innate system, as well as efficient phagocytosis by neutrophils in the tissues, often depends on the presence of a small amount of specific antibody in blood plasma. Antibody is generated by cells of the adaptive immune system in response to specific foreign molecules called *antigens.* In turn, the effective functioning of antibodies and other mediators of the adaptive immune system depends on neutrophils and other effector agents usually associated with the innate immune system. Thus, the innate and the adaptive systems depend on highly evolved, interactive, defensive mechanisms to kill and remove microbial intruders.

Depending on the nature of the stimulus, its mode of presentation, and prior challenges to the immune system, an antigen may elicit either a *cellular* or *humoral immune response.*

Both are ultimately mediated by lymphocytes, the cellular response by T-cells and humoral response by B-cells. Stimulated B-cells differentiate into plasma cells, which secrete antibody specific for the inciting stimulus. The antibody can be found in a variety of body fluids, including saliva, other secretions, and plasma.

Cellular immunity is accomplished by activated T-cells.

The effector cells of this response do not secrete antibody but exert their influence by a variety of cellular mechanisms. These effector processes include direct cytotoxicity mediated by *cytotoxic T-cells*; the suppression or activation of immune mechanisms in other cells — suppressor T-cells or helper T-cells, respectively; and the secretion of cytotoxic or immunomodulating cytokines, such as tumor necrosis factor and interleukin-2. T-cells and their products may act directly or exert their effects in concert with other effector cells, such as neutrophils and macrophages.

The immune responses mediated by antibodies and T-lymphocytes differ in several important respects. In general, antibodies are known to induce immediate responses to antigens and, thereby, provoke *immediate hypersensitivity reactions*. For example, *allergy* or *anaphylactic hypersensitivity* results when a certain type of antibody on the surface of fixed *mast cells* binds to its specific antigen. Antibody binding leads to the release of histamine and other mediators of the allergic response from intracellular granules.

Immediate hypersensitivity reactions also occur when circulating antibodies bind antigen in the tissues, thereby forming immune complexes that activate the complement system, a group of at least nine distinct proteins that circulate in plasma. A cascade of events occurs when the first protein recognizes preformed immune complexes, a large cross-linked mesh of antigen molecules bound to antibodies.

In contrast to the rapid onset of biological responses when antigen binds antibody, the consequences of T cell activation are not noticeable until 24 to 48 hrs after antigen challenge. During this time, the T cells that initially recognize the antigen secrete factors that recruit and activate other cells (e.g., macrophages) and release factors that damage the antigen, cells possessing the antigen, or the surrounding tissue. A common example is the *delayed-type hypersensitivity reaction* to purified protein derivative (PPD), a response used to assess prior exposure to the bacteria that cause tuberculosis. Injected under the skin of sensitive individuals, PPD elicits the familiar inflammatory reaction characterized by local erythema and edema 1 to 2 days later.

Cell-mediated immune responses, while slow to develop, are potent and versatile. These delayed responses provide for defense against many pathogens, including viruses, fungi, and bacteria. T cells are responsible for the rejection of transplanted tissue grafts and containment of the growth of neoplastic cells. A deficiency in T cell immunity, such as that associated with AIDS, predisposes the affected patient to a wide array of serious, life-threatening infections.

Humoral immunity consists of defense mechanisms carried out by soluble mediators in the blood plasma. *Antibodies* (*immunoglobulins*) are glycoproteins secreted by plasma cells. Antibodies are found in high levels in plasma and other body fluids. They have the ability to bind specifically to the antigenic determinant that induced their secretion.

There are five major classes of antibodies; these classes are grouped based on differences in the amino acid composition of the constant region of the heavy chains.

IgG is the most prevalent antibody in serum and is responsible for adaptive immunity to bacteria and other microorganisms. Bound to antigen, IgG can activate serum complement, which releases several inflammatory and bactericidal mediators. At the surface of bacteria, portions of IgG molecules facilitate the phagocytosis of bacteria by blood phagocytes, a process called *opsonization*. IgG exists in serum as a monomer. It can cross the placenta and is secreted into colostrum, protecting the fetus as well as the newborn from infection.

Unlike IgG, both *IgM* and *IgA* usually exist as polymers of the fundamental Y-shaped antibody unit. In most IgA molecules, two antibody units are held together by a *secretory piece* (*J-chain*), a protein synthesized by epithelial cells. In this conformation, IgA is actively secreted into saliva, tears, colostrums and mucus. IgA is thus known as *secretory immunoglobulin*. IgM is the first antibody secreted after an initial immune challenge and provides resistance early in the course of infection. IgM consists of five Y units. Its size and large number of antigen-binding sites provide the molecule with an excellent capacity for *agglutination*, the ability to clump particulate antigens, such as bacteria and blood cells. Clumped antigens are efficiently and quickly removed by fixed phagocytic cells of the monocyte- macrophage system.

IgE, a monomeric antibody slightly larger than IgG, avidly binds cells that store and release mediators of allergy and anaphylaxis, including mast cells and basophils. These cells are heavily granulated. The granules contain histamine, leukotrienes, and other biologically active agents that increase vascular permeability, dilate blood vessels (and, thereby, reduce blood pressure), and contract smooth muscle cells in lung airways.

IgD found in plasma and on the surface of some immature B cells, has no known function.

Chapter 15. Physiology of blood coagulation

The term *hemostasis* refers to the blood flow termination. The normal process of hemostasis is regulated by a complex array of activators and inhibitors that maintain blood fluidity and prevent blood from leaving the vascular compartment. Hemostasis is normal when it seals a blood vessel to prevent blood loss and hemorrhage. It is abnormal when it causes inappropriate blood clotting or when clotting is insufficient to stop the flow of blood from the vascular compartment.

Circulating in a high-pressure, closed system that communicates with all tissues and cells in the body, blood exchanges oxygen, nutrients, and wastes and provides necessary components for host defense. This communication takes place largely in the complex and dynamic networks of capillary beds that provide oxygen to almost every cell in the body (only the cornea and intervertebral disks are avascular; these tissues receive oxygen by diffusion). Disruption of the integrity of the fragile capillaries may result from minor tissue injury associated with normal physical activity or from massive tissue trauma as a result of serious injury or infection, and may quickly lead to death. Any opening in the vascular network may lead to massive bruising or blood loss if left unrepaired.

To minimize bleeding and prevent blood loss after tissue injury, components of the hemostatic system are activated. The components of this dynamic, integrated system include blood platelets, endothelial cells, and plasma coagulation factors. They may be activated on exposure to foreign surfaces during bleeding, or by torn tissue at the site of injury, or by products released from the interior of damaged cells.

1. HEMOSTASIS DEFINITIONS

Ano older explanation of hemostasis explains this process as three consecutive stages. Such definition was introduced by Dr. Moravitz in 1906. According to it, hemostasis has the following stages:

Initiation of Blood Coagulation that starts immediately after tissue and blood vessel injury and terminates when active thrombin is formated.

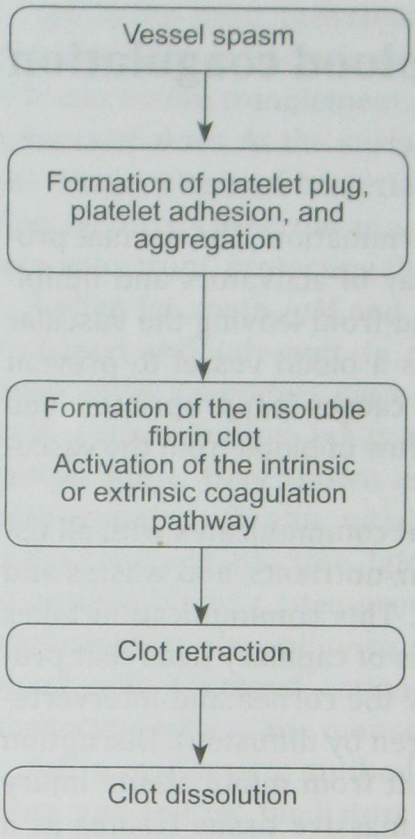


Fig. 15–1. The steps in hemostasis.

Proper Stage of Blood Coagulation that starts when active fibrin formated and terminates in the moment of final plug formation.

Fibrynolysis that in other words means clot retraction.

According to the temporal point of view, hemostasis can be viewed as five separate but interrelated events:

- compression and vasoconstriction, which act immediately to stop the flow of blood;
- formation of a platelet plug;
- blood coagulation or development of an insoluble fibrin clot;
- clot retraction;
- clot dissolution, the second definition — fibrinolysis (Fig. 15–1).

2. VESSEL SPASM

Immediately after tissue injury, blood flow through the disrupted vessel is slowed by the interplay of several important physical factors, including compression or back-pressure exerted by the tissue around the injured area, and vasoconstriction. Vessel spasm is initiated by endothelial injury and caused by local and humoral mechanisms. A spasm constricts the vessel and reduces blood flow. It is a transient event that usually lasts less than 1 minute. Thromboxane A_2 (TXA_2), a prostaglandin released from the platelets, contributes to the vasoconstriction. A second prostaglandin, prostacyclin, released from the vessel endothelium, produces vasodilation and inhibits platelet aggregation.

The degree of compression varies in different tissues; for example, bleeding below the eye is not readily deterred because the skin in this area is easily distensible. Back-pressure increases as blood which leaks out of the disrupted capillaries accumulates. In some tissues, notably the uterus after childbirth, contraction of underlying muscles compresses blood vessels supplying the tissue and minimizes blood loss.

Damaged cells at the site of tissue injury release potent substances, except of TXA_2 and prostacyclin, that directly cause blood vessels to constrict, including serotonin, epinephrine and fibrinopeptide B.

3. PLATELETS' PHYSIOLOGY

The platelet plug, the second line of defense, is initiated as platelets come in contact with the vessel wall. Small breaks in the vessel wall are often sealed with a platelet plug and do not require the development of a blood clot.

Platelets, also called thrombocytes, are large fragments from the cytoplasm of bone marrow cells called megakaryocytes. They are enclosed in a membrane but have no nucleus and cannot reproduce. Their cytoplasmic granules release mediators for hemostasis. Although they lack a nucleus, they have many of the characteristics of a whole cell. They have mitochondria and enzyme systems for producing adenosine triphosphate (ATP) and adenosine diphosphate (ADP), and they have the enzymes needed for synthesis of prostaglandins, which are required for their function in hemostasis. Platelets also produce a growth factor that causes vascular endothelial cells, smooth muscle cells, and fibroblasts to proliferate and grow.

The life span of a platelet is only 8 to 9 days. A protein called *thrombopoietin* causes proliferation and maturation of megakaryocytes. The sources of thrombopoietin include the liver, kidney, smooth muscle, and bone marrow, which controls platelet production. Its production and release are regulated by the number of platelets in the circulation. The newly formed platelets that are released from the bone marrow spend as long as 8 hrs in the spleen before they are released into the blood. The cell membrane of the platelet is important to its function. The outside of the platelet membrane is coated with glycoproteins that repulse adherence to the normal vessel endothelium, while causing adherence to injured areas of the vessel wall, particularly the subendothelial layer. The platelet membrane also has glycoprotein receptors that bind fibrinogen and link platelets together. Glycoprotein receptor antagonists have been developed and are selectively used in the treatment of acute coronary myocardial infarction. In addition, the platelet membrane contains large amounts of phospholipids that play a role in activating several points in the blood-clotting process.

4. FORMATION OF THE PLATELET PLUG

Platelets regulate bleeding in three stages that defines as the *primary* or *platelets' coagulation. First*, they form multicellular aggregates linked by protein strands at sites of openings in blood vessels. The aggregates form a physical barrier that begins to limit blood loss soon after the opening occurs.

Second, phospholipids on the platelet plasma membrane activate the enzyme *thrombin*, which initiates a cascade of events ending in clot formation.

Finally, platelets possess multiple storage granules, which they discharge (secrete) to enhance coagulation.

Platelet activation results in the sequential responses of *adherence, aggregation* and *secretion.* Adherence is initiated when one or more substances, released from cells or activated in plasma at the site of a hemorrhage, bind to receptors in the platelet plasma membrane. Receptor binding results, via second messengers, in adherence (to other platelets and the inner, endothelial surface of blood vessels) and secretion. Disruption of the endothelium at sites of tissue injury exposes a variety of proteins in the subendothelial matrix, such as *collagen* and *laminin*, which either induce or support platelet adherence (Fig. 15–2, A). Endothelial cells also rapidly deploy cellular adherence antigens known as *integrins* on the outer surface of their plasma membranes during wound healing. These adherence antigens are deployed to the cell membrane by cellular processes set in motion by factors generated during coagulation or by factors released from platelets during clotting. In turn, activated endothelial cells release substances that participate in hemostasis. The process of adhesion is controlled by local hormones and substances released by platelet granules.

von Willebrand factor (vWF), a protein synthesized by endothelial cells of blood vessels and megakaryocytes, enhances platelet adherence by forming a bridge between cell surface receptors and collagen in the subendothelial matrix. Therefore, it performs two important functions: it aids in platelet adhesion, and it circulates in the blood as a carrier protein for coagulation factor VIII.

As the platelets adhere to the collagen fibers on the damaged vessel wall, they begin to release large amounts of adenosine diphosphate (ADP) and TXA_2. Platelet aggregation and formation of a loosely organized platelet plug occur as the ADP and TXA_2 cause nearby platelets to become sticky and adhere to the original platelets. Stabilization of the platelet plug occurs as the coagulation pathway is activated on the platelet surface and fibrinogen is converted to fibrin, thereby creating a fibrin meshwork that cements together the platelets and other blood components (Fig. 15–2, B). Thrombin, which is generated by

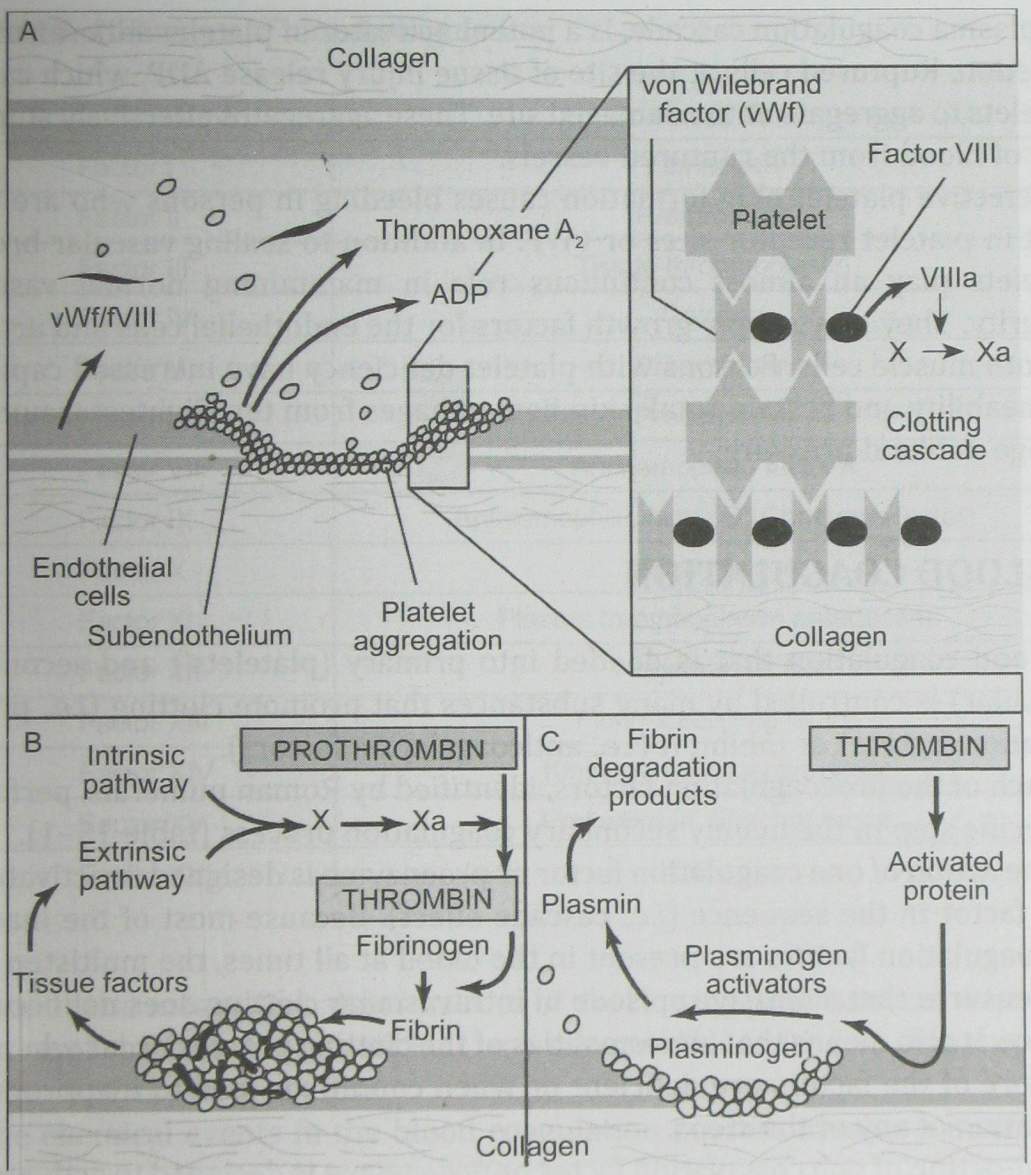


Fig. 15–2. (A) The platelet plug occurs seconds after vessel injury. Von Willebrand's factor, released from the endothelial cells, binds to platelet receptors, causing adhesion of platelets to the exposed collagen. Platelet aggregation is induced by release of TXA2 and ADP.
(B) Coagulation factors, activated on the platelet surface, lead to the formation of thrombin and fibrin, which stabilize the platelet plug.
(C) Control of the coagulation process and clot dissolution are governed by thrombin and plasminogen activators. Thrombin activates protein C, which stimulates the release of plasminogen activators. The plasminogen activators in turn promote the formation of plasmin, which digests the fibrin strands.

the plasma coagulation cascade, is a potent activator of platelet adherence and secretion. Ruptured cells at the site of tissue injury release ADP, which causes platelets to aggregate at the damaged site. These aggregates effectively stop the flow of blood from the ruptured vessels.

Defective platelet plug formation causes bleeding in persons who are deficient in platelet receptor sites or vWF. In addition to sealing vascular breaks, platelets play an almost continuous role in maintaining normal vascular integrity. They may supply growth factors for the endothelial cells and arterial smooth muscle cells. Persons with platelet deficiency have increased capillary permeability and sustain small skin hemorrhages from the slightest trauma or change in blood pressure.

5. BLOOD COAGULATION

Blood coagulation that is divided into primary (platelets') and secondary (vascular) is controlled by many substances that promote clotting (*i.e.,* procoagulation factors) or inhibit it (*i.e.,* anticoagulation factors).

Each of the procoagulation factors, identified by Roman numerals, performs a specific step in the mainly secondary coagulation process (table 15–1).

The action of one coagulation factor or proenzyme is designed to activate the next factor in the sequence (*i.e.,* cascade effect). Because most of the inactive procoagulation factors are present in the blood at all times, the multistep process ensures that a massive episode of intravascular clotting does not occur by chance. It also means that abnormalities of the clotting process occur when one or more of the factors are deficient or when conditions lead to inappropriate activation of any of the steps.

The action of one coagulation factor or proenzyme is designed to activate the next factor in the sequence (*i.e.,* cascade effect). Because most of the inactive procoagulation factors are present in the blood at all times, the multistep process ensures that a massive episode of intravascular clotting does not occur by chance. It also means that abnormalities of the clotting process occur when one or more of the factors are deficient or when conditions lead to inappropriate activation of any of the steps.

The action of one coagulation factor or proenzyme is designed to activate the next factor in the sequence (*i.e.,* cascade effect). Because most of the inactive procoagulation factors are present in the blood at all times, the multistep process ensures that a massive episode of intravascular clotting does not occur by

Table 15–1. Factors of blood coagulation

Scientific Name	Common Name
Factor I	Fibrinogen
Factor II	Prothrombin
Factor III	Tissue thromboplastin
Factor IV	Calcium
Factor V	Proaccelerin
Factor VII	Proconvertin
Factor VIII	Antihemophilic factor A
Factor IX	Antihemophilic factor B (Christmas factor)
Factor X	Stuart factor
Factor XI	Plasma thromboplastin antedecent
Factor XII	Hageman factor
Factor XIII	Fibrin stabilizing factor
Factor XIV	Kininogen, Fitzgerald factor
Factor XV	Prekallikrein, Fletcher factor
Factor XVI	Kallikrein

chance. It also means that abnormalities of the clotting process occur when one or more of the factors are deficient or when conditions lead to inappropriate activation of any of the steps.

The chemical events in the blood coagulation process involve a number of essential steps that result in the conversion of fibrinogen, a circulating plasma protein, to the fibrin strands that enmesh platelets, blood cells, and plasma to form the clot. Briefly, an overview of what happens during blood coagulation one could see as follows.

Platelet aggregates are trapped in a highly organized, firm, and degradable network of *fibrin*, an insoluble protein generated in plasma as a consequence of activation of either the intrinsic or extrinsic clotting cascades, discussed below. The fibrin network traps red cells, leukocytes, platelets, and serum at sites of vascular damage, thereby forming a blood clot. The stable, fibrin-based blood clot eventually replaces the unstable platelet aggregate formed immediately after tissue injury. Fibrin is an insoluble polymer of proteolytic products of the plasma

protein *fibrinogen*. Fibrin molecules are cleaved from fibrinogen by thrombin, which is generated in plasma during clotting. In the initial step of fibrin formation, thrombin cleaves four small peptides (fibrinopeptides) from each molecule of fibrinogen. The fibrinogen molecule devoid of these fibrinopeptides is called *fibrin monomer*. The fibrin monomers spontaneously assemble into ordered fibrous arrays of fibrin, resulting in an insoluble matrix of fibrous strands. At this stage, the clot is held together by noncovalent forces. A plasma enzyme, *fibrin stabilizing factor (Factor XIII)*, catalyzes the formation of covalent bonds between strands of polymerized fibrin, stabilizing and tightening the blood clot.

The initiation of the clotting process occurs by way of the intrinsic or the extrinsic coagulation pathways (Fig. 15–3). The sequential activation of a series of inactive molecules resulting in a biological response is called a *metabolic cascade*.

The intrinsic pathway, which is a relatively slow process, begins in the blood itself.

The extrinsic pathway, which is a much faster process, begins with tissue or vessel trauma and the subsequent release of a complex of several factors, called *tissue factor*, or *tissue thromboplastin*.

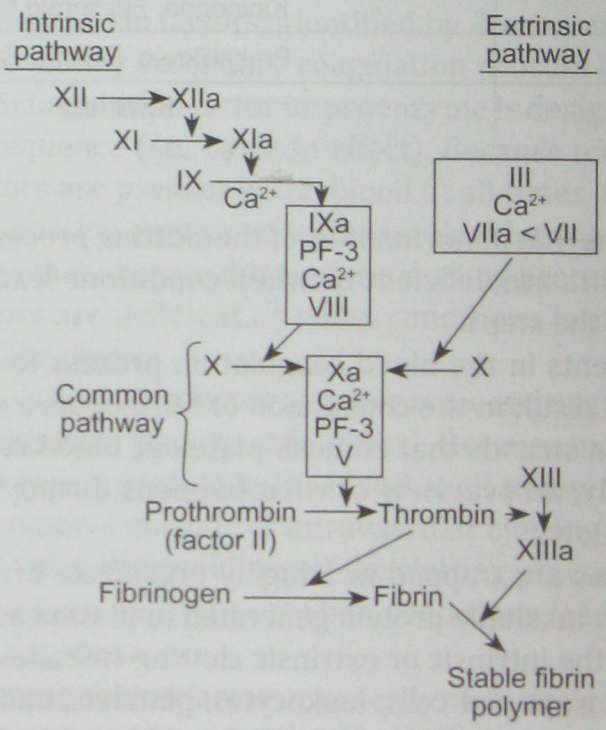


Fig. 15–3. Steps in coagulation cascade. These are the events of the vascular (secondary) hemostasis.

The terminal steps in both pathways are the same: the activation of factor X, the conversion of prothrombin to thrombin, and the conversion of fibrinogen to fibrin. Prothrombin is an unstable plasma protein, which is easily split into smaller parts, one of which is thrombin. Thrombin, in turn, acts as an enzyme to convert fibrinogen to fibrin.

Both the intrinsic and extrinsic pathways are needed for normal hemostasis in the large-diameter (deep) vessels, and many interrelations exist between them. Each system is activated when blood passes out of the vascular system. pathways. In the intrinsic pathway, all the factors required for coagulation are present in the circulation. The intrinsic system is activated as blood comes in contact with collagen in the injured vessel wall and the extrinsic system when blood is exposed to tissue extracts. Bleeding, when it occurs because of defects in the extrinsic system, usually is not as severe as that which results from defects in the intrinsic pathway.

For the initiation of the extrinsic pathway, a factor extrinsic to blood but released from injured tissue, called tissue thromboplastin or tissue factor (factor III), is required. Phospholipids are required for activation of both coagulation pathways. Phospholipids provide a surface for the efficient interaction of several factors. A component of tissue factor provides the necessary phospholipid for the extrinsic pathway. Phospholipids required for the activation of the intrinsic pathway are found on platelet membranes. The final events leading to fibrin formation by both pathways result from the activation of the *common pathway.*

The common pathway is initiated by the conversion of inactive clotting factor X to its active form, factor Xa (Fig. 15–3) and results in the conversion of prothrombin to thrombin, thereby catalyzing the generation of fibrin. Thrombin also enhances the activity of clotting factors V and VIII, accelerating "upstream" events in the coagulation pathway. Finally, thrombin is a potent platelet and endothelial cell stimulus and enhances the participation of these cells in coagulation.

Factor X is activated during both the extrinsic and the intrinsic pathways. In the extrinsic pathway, factor X is activated by a complex consisting of activated factor VII, Ca^{2+}, and factor III (tissue factor). Activation of this complex by tissue factor bypasses the requirement for coagulation factors VIII, IX, XI, and XII used in the intrinsic pathway.

In the intrinsic pathway, clotting is initiated by the activation of factor XII by contact to exposed surfaces, such as collagen in the subendothelial matrix. The activation of factor XII requires several cofactors, including *kallikrein* and *high-*

molecular-weight kininogen. In this pathway, factor X is activated by a complex consisting of factor VIII, factor IXa, platelet factor 3, and Ca^{2+}.

With few exceptions, almost all the blood-clotting factors are synthesized in the liver. Vitamin K is required for the synthesis of prothrombin, factors VII, IX, X, and protein C. Calcium (factor IV) is required in all but the first two steps of the clotting process. The body usually has sufficient amounts of calcium for these reactions. Inactivation of the calcium ion prevents blood from clotting when it is removed from the body.

Coagulation is regulated by several natural anticoagulants. Antithrombin III inactivates coagulation factors and neutralizes thrombin, the last enzyme in the pathway for the conversion of fibrinogen to fibrin. When antithrombin III is complexed with naturally occurring heparin, its action is accelerated and provides protection against uncontrolled thrombus formation on the endothelial surface. Protein C, a plasma protein, acts as an anticoagulant by inactivating factors V and VIII. Protein S, another plasma protein, accelerates the action of protein C. Plasmin breaks down fibrin into fibrin degradation products that act as anticoagulants.

6. CLOT RETRACTION

After the clot has formed, clot retraction, which requires large numbers of platelets, contributes to hemostasis by squeezing serum from the clot and joining the edges of the broken vessel. Clot retraction is a phenomenon that usually occurs within minutes or hours after clot formation. The clot draws together, extruding a very large fraction of the serum. Clot retraction decreases the breakdown of the clot and enhances wound healing.

7. FIBRINOLYSIS (CLOT DISSOLUTION)

The process by which a blood clot dissolves is called *fibrinolysis*. The dissolution of a blood clot begins shortly after its formation; this allows blood flow to be reestablished and permanent tissue repair to take place (Fig. 15–2, C). As with clot formation, clot dissolution requires a sequence of steps controlled by activators and inhibitors (Fig. 15–4).

Platelet function is strongly inhibited, for example, by the endothelial cell metabolite *prostacyclin*, which is generated from arachidonic acid during cellular activation. Activated endothelial cells also release *tissue plasminogen acti-*

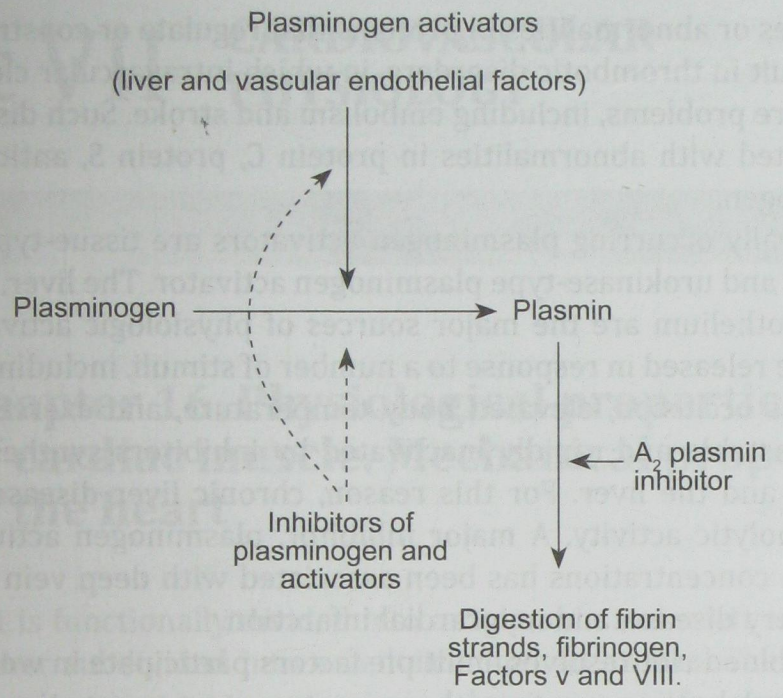


Fig. 15–4. Fibrinolytic system and its modifiers.
The solid lines indicate activation, and the broken lines indicate inactivation.

vator, which converts *plasminogen* to *plasmin*. Plasminogen, the proenzyme for the fibrinolytic process, normally is present in the blood in its inactive form. It is converted to its active form, plasmin, by plasminogen activators formed in the vascular endothelium, liver, and kidneys. The plasmin digests the fibrin strands of the clot and certain clotting factors, such as fibrinogen, factor V, factor VIII, prothrombin, and factor XII. Finally, it hydrolyzes fibrin, resulting in dissolution of the fibrin clot in a process called fibrinolysis. Plasmin is rapidly inactivated by α_2-plasmin inhibitor, which limits fibrinolysis to the local clot and prevents it from occurring in the entire circulation.

Thrombin bound to *thrombomodulin* on the surface of endothelial cells converts *protein C* to an active protease. Activated protein C and its cofactor, *protein S*, restrain further coagulation by proteolysis of factors Va and VIIIa. Furthermore, activated protein C augments fibrinolysis by blocking an inhibitor of TPA. Finally, *antithrombin III* is a potent inhibitor of proteases involved in the coagulation cascade, such as thrombin. The activity of antithrombin III is accelerated by small amounts of heparin, a mucopolysaccharide present in the cells of many tissues.

Deficiencies or abnormalities in proteins that regulate or constrain coagulation may result in thrombotic disorders, in which intravascular clot formation leads to severe problems, including embolism and stroke. Such disorders have been associated with abnormalities in protein C, protein S, antithrombin III, and plasminogen.

Two naturally occurring plasminogen activators are tissue-type plasminogen activator and urokinase-type plasminogen activator. The liver, plasma, and vascular endothelium are the major sources of physiologic activators. These activators are released in response to a number of stimuli, including vasoactive drugs, venous occlusion, elevated body temperature, and exercise. The activators are unstable and rapidly inactivated by inhibitors synthesized by the endothelium and the liver. For this reason, chronic liver disease may cause altered fibrinolytic activity. A major inhibitor, plasminogen activator inhibitor-1, in high concentrations has been associated with deep vein thrombosis, coronary artery disease, and myocardial infarction.

While the blood clot resolves, multiple factors participate in wound healing. Optimal wound healing requires the recruitment or generation of new tissue cells as well as new blood vessels to nourish the repairing tissue. Thus, secreted proteins and lipids that attract cells (chemoattractants), induce cells to proliferate (mitogens), and induce primitive cells to differentiate (growth factors) are called into play. These agents act in concert to induce the formation of new tissue and repair the injured area. The healing area is vascularized by a process known as *angiogenesis*, the formation of new blood vessels from preexisting ones.

Platelets, activated during clotting, play an important role in the angiogenic response because they secrete factors that induce proliferation, migration, and differentiation of two of the major components of blood vessels, endothelial cells, and smooth muscle cells. Of the factors released from platelets involved in the angiogenic response, a novel lipid — *sphingosine 1-phosphate* — plays an important role in wound healing and angiogenesis.

Released during clotting and acting in conjunction with protein growth factors, this lipid induces the proliferation of new tissue cells to replace damaged ones and drives the formation of new blood vessels until the healing process is complete. It does so by inducing the migration, proliferation, and differentiation of fibroblasts, smooth muscle cells, and endothelial cells at the site of tissue repair. Sphingosine 1-phosphate exerts its effects optimally when acting in conjunction with protein growth factors that possess angiogenic capabilities, including *vascular endothelial growth factor* and *fibroblast growth factor*.

Unit VII CARDIOVASCULAR PHYSIOLOGY

Chapter 16. Physiological properties of cardiac muscle. Mechanical properties of the heart

The heart is functionally divided into a right side and a left side. Each side may be further subdivided into a ventricle and an atrium. The primary role of each atrium is to act as a reservoir and *booster pump* for venous return to the heart. Recently, with the discovery of atrial sodium-uretic hormone, other homeostatic roles of the atrium have been proposed. The primary physiologic function of each ventricle is to maintain circulation of blood to the organs of the body. The left heart receives oxygenated blood from the pulmonary circulation, and contraction of the muscles of the left ventricle provide energy to propel that blood through the systemic arterial network. The right ventricle receives blood from the systemic venous system and propels it through the lungs and onward to the left ventricle. The reason that blood flows through the system is because of the pressure gradients set up by the ventricles between the various parts of the circulatory system.

In order to understand how the heart performs its task, one must have an appreciation of the force-generating properties of cardiac muscle, the factors which regulate the transformation of muscle force into intraventricular pressure, the functioning of the cardiac valves and something about the load against which the ventricles contract. This chapter will focus on a description of the pump function of the ventricles with particular attention to a description of those properties as represented on the pressure-volume diagram. Emphasis will be given to the clinically relevant concepts of *contractility, afterload* and *preload.* In addition, we will review how the ventricle and the arterial system interact to determine cardiovascular performance (cardiac output and blood pressure).

1. PROPERTIES OF CARDIAC MUSCLE

According to the traditional teaching, the basic properties of cardiac muscle are the following:

- excitability [that defines as bathmotropy which might be both positive and negative];
- conductivity [dromotropy];
- contractility [inotropy];
- rhythmicity (also called autorhytmicity and automathy) [chronotropy];
- tone of cardiac muscle [tonotropy] (Fig. 16–1).

Of these properties, the ability to contract is mechanical properties. In short, cardiac muscle properties, belong to two groups: (i) mechanical and (ii) electrophysiological.

2. THE MECHANICAL PROPERTIES

The basic mechanical property of the heart muscles is that it can contract. Contraction may or may not cause, shortening. Nevertheless, all contractions generate force and produce squeeze on the blood in ventricular or atrial

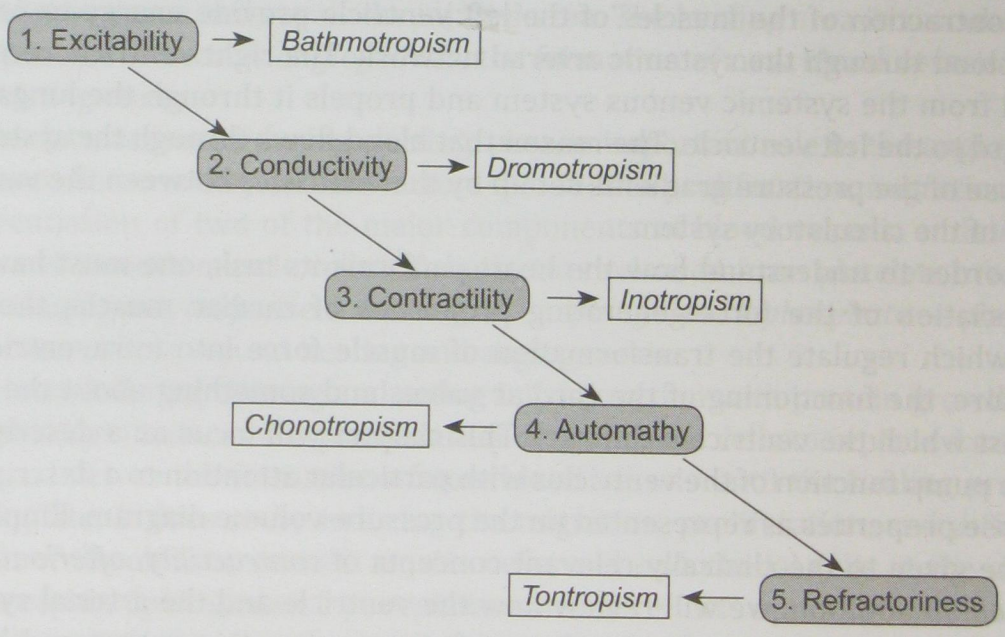


Fig. 16–1. Properties of the cardiac muscle.

cavities → which in turn, ultimately speaking, causes ejection of blood from these cavities.

The major issues, discussed below, under the mechanical properties are:

a) mechanism of shortening
b) excitation contraction coupling
c) the force of contraction

The force of contraction is a big issue and has been subdivided into:

1. Factors influencing the force of contraction, Starling's law and contractility.
2. The concept of $V_{max.}$
3. Effect of Ca^{2+}, noradrenalin.
4. "Staircase" phenomenon.
5. All or none law.
6. Heart lung preparation of Starling.

Before going for details, the student should be familiar with the following two terms: (1) isometric contraction, and (ii) isotonic contraction.

Isometric contraction

In this case, the muscle contracts but does not shorten. For example, when the left ventricle is contracting isometrically, the ventricle shows all other features of contraction but does not shorten. Thus, the ventricle becomes hard, generates heat, generates force, consumes extra energy, shows biochemical changes of contraction within its muscles but does not shorten, i.e., it performs no (external) work and does not eject blood into the aorta. Nevertheless, as it is generating force, the isometrically contracting ventricle exerts great squeezing pressure on the blood in its cavity.

Isotonic contraction

In addition to other features of contraction, the ventricle shortens, does external work and ejects blood.

Mechanism of shortening

Basic anatomical facts need recapitulation, presuming, the student is already aware of the facts, stated under, 'working myocardial cells'. They are the following:

1. Working myocardial cells (WMC) contract but the pacemaker and the junctional tissues, although they belong to myocardium, do not. The gross histology, EM pictures and mechanism of contraction of WMC are remarkably similar to those of skeletal muscle cells (SMC).
2. Each WMC (or SMC) shows under LM (light microscope), alternate dark and light bands called A and I bands respectively.
3. A sarcomere consists of 1/2 I band + one A band + 1/2 I band.
4. In a sarcomere, the thick (myosin) filaments is covered by 6 thin (actin) filaments.
5. Each actin consists of a double helix. Actually a single thin filament may consists of some 300–400 actin molecules (a single thick filament, similarly contains roughly 400 molecules).

Thus, a thin filament is a polymer of many actin molecules called G actin, the polymer being called F actin. Similarly, a thick filament is a polymer of many myosin molecules.

Two other proteins, the regulatory proteins, tropomyosin and troponin are associated with the actin helix.

The actin molecules contain sites for attachment with myosin but the tropomyosin is to attached that most of these sites are covered by the tropomyosin band.

The troponin (actually there are several types of troponin, troponin I, troponin C etc.) molecules are attached with the tropomyosin. Troponin C can bind with ionic Ca^{2+}.

6. The myosin molecule has a globular head and an elongated tail. The head has sites by which it can bind with the actin molecule, provided the actin molecular sites are not covered by tropomyosin. The myosin molecular head has also binding sites for ATP.
7. This head of myosin can split ATP into ADP and inorganic phosphate (Pi). Therefore, myosin head has ATP-ase activity and is called myosin ATPase.

With this background, the following additions can be made:

1. When contraction are about to begin, cross bridges develop between the thick and thin filaments.
2. Each cross bridge swivels (=rotates a little) → as a result of this rotation the actin thin filaments receives a push so that the thin filament proceeds towards the H zone. The cross bridge is next detached from the actin filament, reattaches itself into a new site of the same thin filament and swivels again → further movement of the thin filament towards

the H-zone. These cycles go on and as a result, the H zone is narrowed, Z-line moves towards the H zone and the muscle as a whole shortens. But note, the length of either the thin filament or the thick filament does not reduce. That this is the molecular basis of muscular contraction was discovered by Huxley and is known as 'sliding filament theory' or 'cross bridge theory' or 'ratchet theory' (all self explaining terms).

Now, question arises, at rest there was no cross bridge; why does the cross bridge develop just before the onset of contraction?

Answer is, when free Ca^{2+} ions are available in sufficient concentration within the WMC (or SMC), the Ca^{2+} ions bind with the troponin C. This results in some physiochemical changes in the (resting) actin tropomyosin complex, so that tropomyosin loosens → nextly the tropomyosin moves away from the actin are uncovered. Nextly, the globular head of myosin attaches with the actin, and this constitutes the cross bridge → swiveling action → movement of actin → shortening. That is, at rest, although the myosin heads are present, they cannot join with the actin.

Excitation contraction coupling

When the muscle (WMC/SMC) is stimulated, it develops an action potential (AP). This is excitation.

When the muscle contracts, there is shortening of the muscle by sliding movement of the thin filament, or (in isometric contraction) it shows other mechanical characteristics of contraction.

It is fact that all muscular contractions are preceded by the development of an AP. Question arises what is the 'coupling' (=linking) material between the AP (which is an electrical phenomenon) and the contraction (a mechanical event)? In short, what couples excitation (AP) and contraction (mechanical event)?

The answer is, the Ca^{++} ions. Details are as follows. To understand the mechanisms of contractility change, it is necessary to consider the factors affecting cellular calcium handling.

The processes linking membrane excitation to contraction via calcium ions are illustrated in Fig. 16–2. Since this involves many possible movements and locations of calcium, the processes are considered in the order in which they would be encountered during a single contraction.

The initial event is an action potential (1) traveling along the cell surface. As in skeletal muscle, the action potential enters a T-tubule (2), where it can

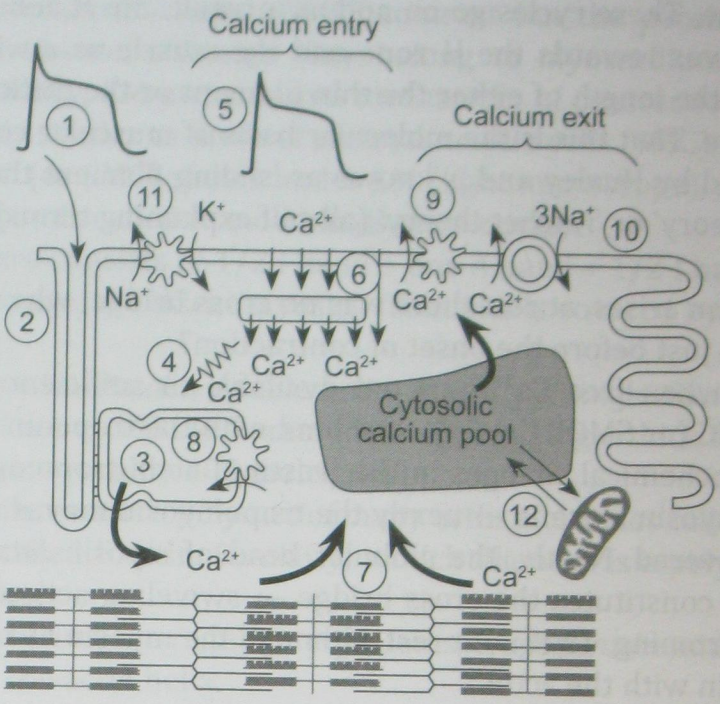


Fig. 16–2. Calcium role in the mechanism of cardiac muscle contraction. Explanation for the events marked with the figs inside the holes are in the text.

communicate with the sarcoplasmatic reticu-lum (3) to cause calcium release. This mechanism for release of calcium in cardiac muscle is much less than in skeletal muscle and is insufficient to cause adequate activation of the contraction. To some extent, activation is aided by a calcium-induced calcium release mechanism (4) triggered by a rise in the cytoplasmatic calcium concentration. The action potential (5) on the cell surface (sarcolemma) also causes the opening of calcium channels, through which strong inward calcium current flows. These calcium ions accumulate just inside the sarcolemma (6), although some probably diffuse rapidly into the cell interior. Calcium induces the rapid release of calcium from the subsarcolemmal SR, and the calcium then diffuses the short distance to the myofilaments (7) and activates them.

The amount of calcium in the cytoplasm, the *cytosolic calcium pool*, determines the magnitude of the myofilament activation and, hence, the level of contractility. During relaxation, the cytoplasmatic calcium concentration is rapidly lowered through several pathways. The sarcoplasmatic reticulum membrane contains a vigorous Ca^{2+}-ATPase (8) that runs continuously and

is further activated, through a protein phosphorylation mechanism, by high levels of cytoplasmic calcium. At the level of the sarcolemma, two additional mechanisms work to rid the cell of the calcium that entered via previous action potentials.

A membrane Ca^{2+}-ATPase (9) actively extrudes calcium, ejecting one calcium ion for each ATP molecule consumed. Additional calcium is removed by a Na^+/Ca^{2+}-exchange mechanism (10), also located in the cell membrane. This mechanism is part of a coupled transport system in which three sodium ions, entering the cell down their electrochemical gradient, are exchanged for the ejection of one calcium ion. Proper function of this exchange mechanism requires a steep sodium concentration gradient, maintained by the membrane Na^+/K^+-ATPase (11) located in the sarcolemma.

Because the Na^+/Ca^{2+} — exchange mechanism derives its energy from the sodium gradient, any reduction in the pumping action of the Na^+/K^+-ATPase leads to reduced calcium extrusion. Under normal conditions, these mechanisms can maintain a 10,000-fold Ca^{2+} concentration difference between the inside and outside of the cell. Since a cardiac cell contracts repeatedly many times per minute with each beat being accompanied by an influx of calcium, the extrusion mechanisms must also work continuously to balance the incoming calcium. The mitochondria of cardiac muscle (Fig. 16–2, 12) are also capable of accumulating and releasing calcium, although this system does not appear to play a role in the normal functioning of the cell.

Force of Contraction

(a) Factors influencing the force of contraction.

When the heart contracts it generates a force (called, the force of contraction) by virtue of which it expels blood from its cavity. Therefore, greater the force of contraction, greater is the stroke output and vice versa.

The force of contraction changes from time to time in the same heart.

There are two fundamental mechanisms by which the force of contraction of the myocardium can be changed:

- by altering the initial length of the heart muscle (the 'end diastolic fiber length', EDFL); this is expressed by Frank Starling's (also called, simply, Starling's) law of the heart;
- by altering the contractility (nowadays, very frequently called the inotropic state; following A. V. Hill's nomenclature, it may also be called, 'active state' of the heart).

These two terms, the Starling's law and the contractility should not be confused; they are different (although not totally divorced from each other) entities.

Contractility

There is another way (besides increasing the EDFL) by which the force of cardiac contraction can be increased and that is, by increasing the contractility of myocardium. What is then contractility? To answer it, it may be said that rise in contractility is characterized by two features: (1) the muscles lifts a greater load (without its initial length) and (2) the muscle contracts more speedily.

Role of noradrenalin and Ca^{2+}

These substances increase the contractility of the myocardium. Sympathetic stimulation, thus powerfully raises the inotropic state. Inotropic agents usually work through changes in the internal calcium content of the cell.

An increase in the heart rate, for instance, allows more separate influxes of calcium per minute, and the amount of releasable calcium in the subsarcolemmal space and sarcoplasmatic reticulum increases. More cross-bridges are activated, and the force of isometric contraction (and other indicators of contractility) increases. This is the basis of the *force-frequency relationship*, one of the principal means of changing myocardial contractility.

Cardiac glycosides are an important class of therapeutic agents used to increase the contractility of failing hearts. The drug *digitalis*, used for centuries for its effects on the circulation, is typical of these agents.

Other factors influencing contractility

Acidosis reduces the contractility of the heart while alkalinity increases it. Rise of temperature and tachycardia cause increase of contractility. These factors are, in addition to those factors described previously, increased Ca^{2+} concentration, adrenalin or noradrenalin addition and sympathetic stimulation.

Staircase phenomenon

This phenomenon, first described by H.P. Bowditch over 100 years ago, can be described as follows:

if the vagal supply to the heart of a frog be strongly stimulated, cardiac standstill develops. Now, withdraw the stimulus. The heart resumes its bead. The first few beats, after resumption, are progressively higher and higher, so that on the kymograph paper, the record looks like a 'staircase'. Hence, this

phenomenon is spoken of as a 'staircase phenomenon' (also called 'treppe', which is the German equivalent for staircase).

Post extrasystolic potentiation

In extrasystole the following happens: suppose the heart is beating regularly. Now, something happens, a premature arrives at the ventricle → the ventricle contracts → following this unusually long interval occurs → after this a bigger than usual beat is due to the post extrasystolic potentiation of force of contraction. Previously its explanation was: the longer pause means greater cardiac inflow and greater EDFL → operation of Frank Starling's Law → increased force of contraction. It is now known, the correct explanation is: during the bigger heart beat more Ca^{2+} are available → increased contractility → increased force of contraction.

3. THE CARDIAC CYCLE AND PRESSURE-VOLUME LOOPS

The *cardiac cycle* refers to the sequence of electrical and mechanical events occurring in the heart during a single beat and the resulting changes in pressure, flow, and volume in the various cardiac chambers. The cardiac cycle (the period of time required for one heart beat, equal to 0.8 s — Fig. 16–3) is divided into two parts: systole and diastole. The importance of these two stages shown of the Fig. 16–4.

Systole (from Greek, meaning "contracting") is the period of time during which the muscle transforms from its totally relaxed state (with cross-bridges uncoupled) to the instant of maximal mechanical activation (point of maxi-

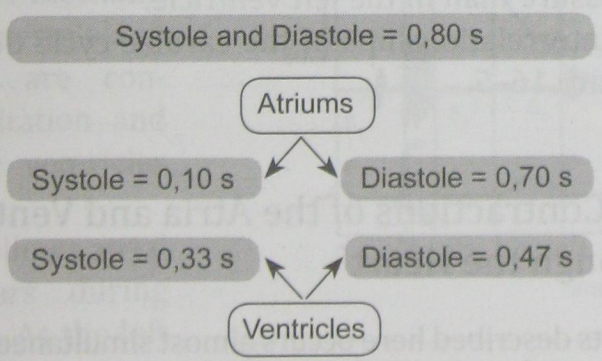


Fig. 16–3. Cardiac cycle duration.

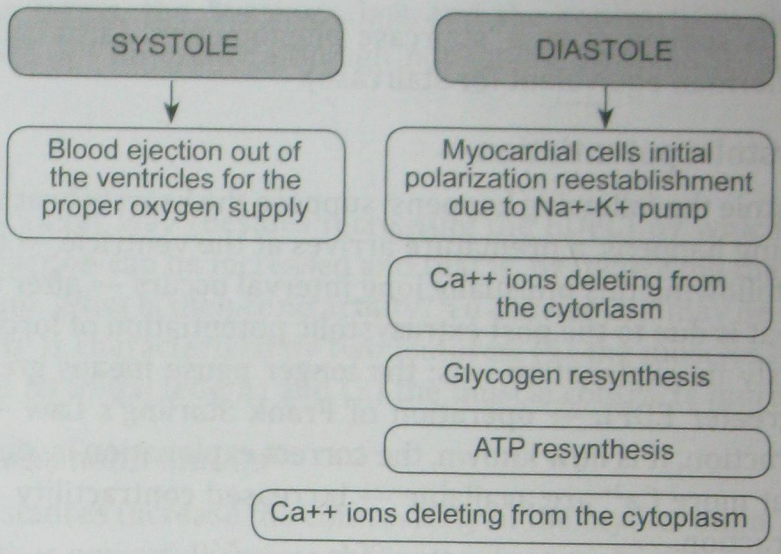


Fig. 16–4. Cardiac cycle stages importance.

mal cross-bridge coupling); this period of time includes the electrical events responsible for initiating the contraction.

Diastole (from Greek, meaning "dilation") is the period of time during which the muscle relaxes from the end-systolic (maximally activated) state back towards its resting state. Systole is considered to start at the onset of electrical activation of the myocardium (onset of the ECG); systole ends and diastole begins as the activation process of the myofilaments passes through a maximum. In the discussion to follow, we will review the hemodynamic events occurring during the cardiac cycle in the *left* ventricle. The events in the *right* ventricle are similar, though occurring at slightly different times and at different levels of pressure than in the left ventricle.

The functional interrelationships of the cardiac cycle described below are represented in Figure 16–5.

3.1. Sequential Contractions of the Atria and Ventricles Pump Blood Through the Heart

The cycle of events described here occurs almost simultaneously in the right and left heart; the main difference is that the pressures are higher on the left side. The focus is on the left side of the heart, beginning with electrical activation of the atria.

Atrial Systole and Diastole. The P wave of the electrocardiogram (ECG) reflects atrial depolarization, which initiates *atrial systole*. Contraction of the atria "tops off" ventricular filling with a final, small volume of blood from the atria, producing the **a wave**. Under resting conditions, atrial systole is not essential for ventricular filling and, in its absence, ventricular filling is only slightly reduced. However, when increased cardiac output is required, as during exercise, the absence of atrial systole can limit ventricular filling and stroke volume. This happens in patients with atrial fibrillation, whose atria do not contract synchronously. The P wave is followed by an electrically quiet period, during which atrioventricular (AV) node transmission occurs (the PR segment). During this electrical pause, the mechanical events of atrial systole and ventricular filling are concluded before excitation and contraction of the ventricles begin.

Atrial diastole follows atrial systole and occurs during ventricular systole. As the left atrium relaxes, blood enters the atrium from the pulmo-

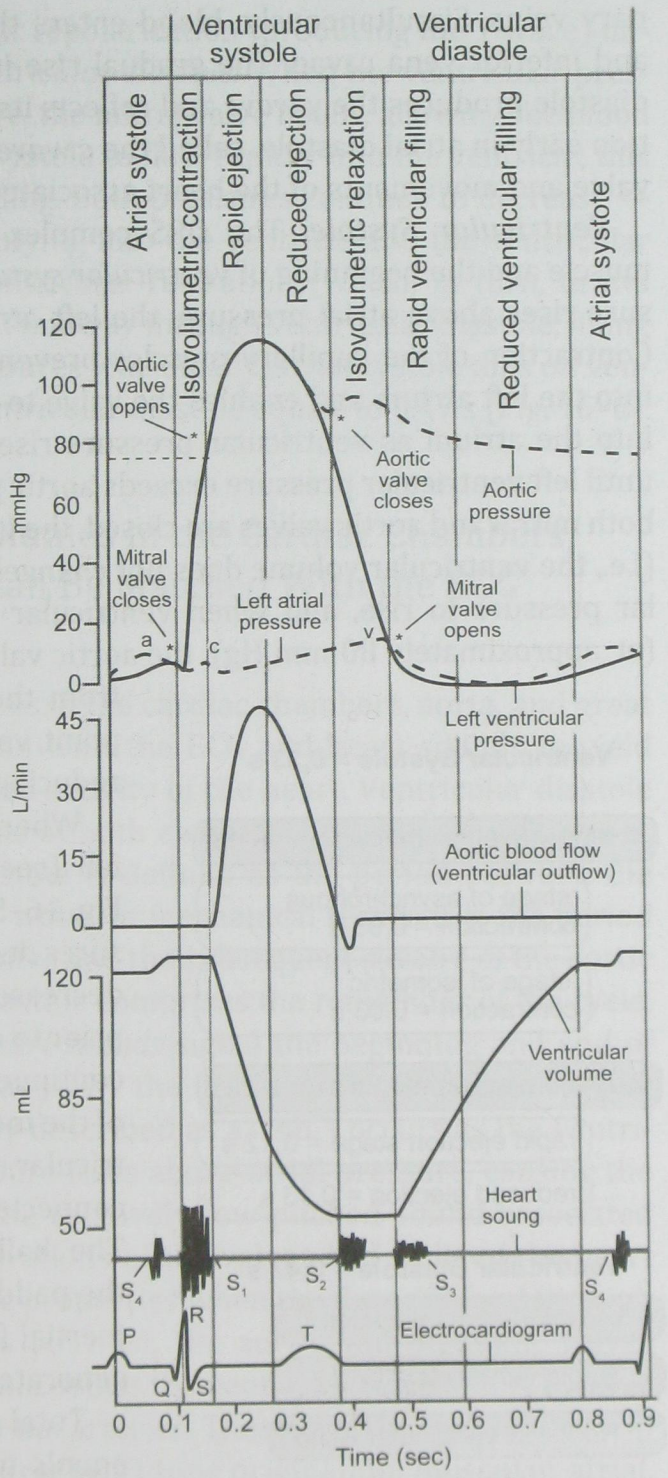


Fig. 16–5. The time-dependence of various events throughout the cardiac cycle.

nary veins. Simultaneously, blood enters the right atrium from the superior and inferior vena cavae. The gradual rise in left atrial pressure during atrial diastole produces the *v wave* and reflects its filling. The small pressure oscillation early in atrial diastole, called the *c wave*, is caused by bulging of the mitral valve and movements of the heart associated with ventricular contraction.

Ventricular Systole. The QRS complex reflects excitation of ventricular muscle and the beginning of *ventricular systole* (Fig. 16–5). As ventricular pressure rises above atrial pressure, the left *atrioventricular (mitral) valve* closes. Contraction of the papillary muscles prevents the mitral valve from reverting into the left atrium and enables the valve to prevent the regurgitation of blood into the atrium as ventricular pressure rises. The aortic valve does not open until left ventricular pressure exceeds aortic pressure. During the interval when both mitral and aortic valves are closed, the ventricle contracts *isovolumetrically* (i.e., the ventricular volume does not change). The contraction causes ventricular pressure to rise, and when ventricular pressure exceeds aortic pressure (at approximately 80 mm Hg), the aortic valve opens and allows blood to flow from the ventricle into the aorta. At this point, ventricular muscle begins to shorten, reducing the volume of the ventricle.

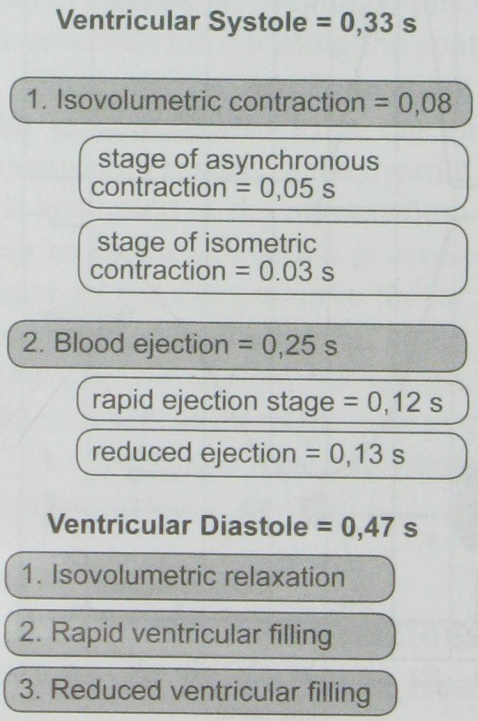


Fig. 16–6. The duration of ventricular systole and diastole periods.

When the rate of ejection begins to fall (see the aortic blood flow record in Fig. 16–5), the aortic and ventricular pressures decline. Ventricular pressure actually decreases slightly below aortic pressure prior to closure of the aortic valve, but flow continues through the aortic valve because of the inertia imparted to the blood by ventricular contraction. (Think of a rubber ball connected to a paddle by a rubber band. The ball continues to travel away from the paddle after you pull back because the inertial force on the ball exceeds the force generated by the rubber band).

Total duration of the ventricular systole equals to 0.33 s with the time of its two stages — isovolumetric contraction and blood ejection — equals to 0.08 s and 0.25 s, correspondently (Fig. 16–6).

Ventricular Diastole. Ventricular repolarization (producing the T wave) initiates ventricular relaxation or *ventricular diastole.* When the ventricular pressure drops below the atrial pressure, the mitral valve opens, allowing the blood accumulated in the atrium during systole to flow rapidly into the ventricle; this is the rapid phase of ventricular filling. Both pressures continue to decrease — the atrial pressure because of emptying into the ventricle and the ventricular pressure because of continued ventricular relaxation (which, in turn, draws more blood from the atrium). About midway through ventricular diastole, filling slows as ventricular and atrial pressures join. Finally, atrial systole tops off ventricular volume. The time of the ventricular diastole equals to 0.47 s (Fig. 16–6).

3.2. Pressures, Flows, and Volumes in the Cardiac Chambers, Aorta, and Great Veins Can Be Matched With the ECG and Heart Sounds

The pressures, flows, and volumes in the cardiac chambers, aorta, and great veins can be studied in conjunction with the ECG and heart sounds to yield an understanding of the coordinated activity of the heart. Ventricular diastole and systole can be defined in terms of both electrical and mechanical events. In electrical terms, ventricular systole is defined as the period between the QRS complex and the end of the T wave. In mechanical terms, it is the period between the closure of the mitral valve and the subsequent closure of the aortic valve. In either case, ventricular diastole comprises the remainder of the cycle.

The *first* (S_1) and *second* (S_2) *heart sounds* signal the beginning and end of mechanical systole (Fig. 16–7). That's why the first heart tone defines as 'systolic'. The first heart sound (usually described as a "lub") occurs as the ventricle contracts and ventricular pressure rises above atrial pressure, causing the atrioventricular valves to close. The relatively low-pitched sound associated with their closure is caused by vibrations of the valves and walls of the heart that occur as a result of their elastic properties when the flow of blood through the valves is suddenly stopped. In contrast, the aortic and pulmonic valves close at the end of ventricular systole, when the ventricles relax and pressures in the ventricles fall below those in the arteries. Therefore, there are four components which participated in the first heart tone origination: muscular, atrial, valvular and vascular components.

The elastic properties of the aortic and pulmonic valves produce the second heart sound, which is relatively high-pitched (typically described as a "dup").

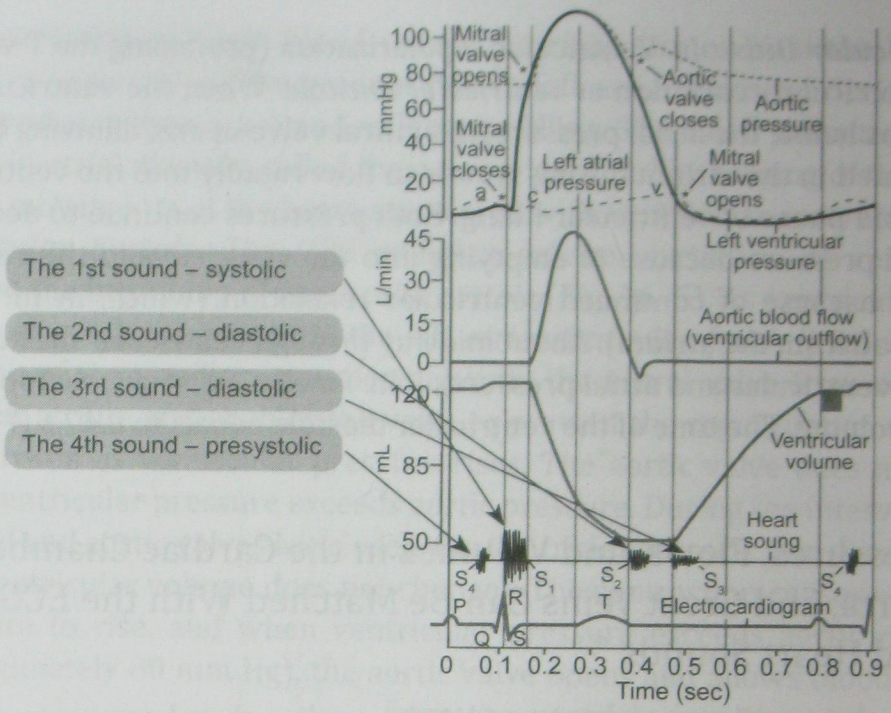


Fig. 16–7. Heart sounds.

The second heart sound defines as diastolic. Mechanical events other than vibrations of the valves and nearby structures contribute to these two sounds, especially S_1; these factors include movement of the great vessels and turbulence of the rapidly moving blood. The second heart sound often has two components — the first corresponds to aortic valve closure and the second to pulmonic valve closure. In normal individuals, splitting widens with inspiration and narrows or disappears with expiration. If compare with the first heart sound, the second has only two components which participated in its: valvular and vascular components.

A *third heart sound* (S_3) results from vibrations during the rapid phase of ventricular filling and is associated with ventricular filling that is too rapid. This sound is diastolic also (Fig. 16–7). Although it may be heard in normal children and adolescents, its appearance in a patient older than age 35 usually signals the presence of a cardiac abnormality.

A *fourth heart sound* (S_4) may be heard during atrial systole. It is caused by blood movement resulting from atrial contraction and, like S_3, is more common in patients with abnormal hearts. This sound, according to origination, is presystolic.

3.3. Pressure-volume dependences throughout the cardiac cycle

The mechanical events occurring during the cardiac cycle consist of changes in pressure in the ventricular chamber which cause blood to move in and out of the ventricle.

Thus, we can characterize the cardiac cycle by tracking changes in pressures and volumes in the ventricles as shown in the Fig. 16–8 where ventricular volume (LVV), ventricular pressure (LVP) and aortic pressure (AoP) are plotted as a function of time.

Shortly prior to time "A" LVP and LVV are constant and AoP is gradually declining. During this time the heart is in its relaxed (diastolic) state; AoP falls as the blood ejected into the arterial system on the previous beat gradually moves from the large arteries to the capillary bed.

At time A contraction begins, and pressure rises inside the chamber. Early after contraction begins, LVP rises to be greater than left atrial pressure (LAP, not shown in the graph) and the mitral valve closes. Since LVP is less than AoP, the aortic valve is closed as well. Since both valves are closed, no blood can enter or leave the ventricle during this time, and therefore the ventricle is contracting *isovolumetrically* (at a constant volume). This period is called *isovolumetric contraction.*

Eventually (at time B), LVP slightly exceeds AoP and the aortic valve opens. During the time when the aortic valve is open there is very little difference between LVP and AoP, provided that AoP is measured just on the distal side of the aortic valve. During this time, blood is ejected from the ventricle into the aorta, as indicated on the LVV tracing as a reduction in ventricular volume.

The exact shape of the aortic pressure wave and LVV wave during this *ejection* phase is determined by the complex

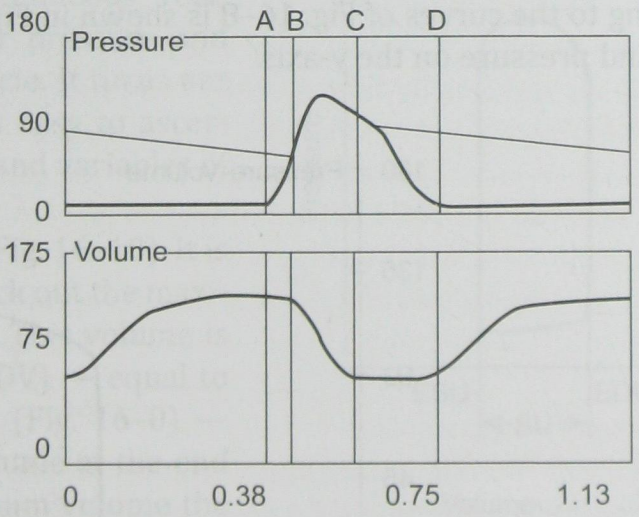


Fig. 16–8. Graph of left ventricular pressure (PRESSURE), blood pressure in aorta (upper fragment) and ventricular volume (VOLUME; lower fragment) dependence upon the time from the systole beginning.

interaction between the ongoing contraction process of the cardiac muscles and the properties of the arterial system, and is beyond the scope of this chapter.

As the contraction process of the cardiac muscle reaches its maximal effort, ejection slows down and ultimately, as the muscles begin to relax, LVP falls below AoP (time C) and the aortic valve closes. At this point ejection has ended and the ventricle is at its lowest volume.

The relaxation process continues as indicated by the continued decline of LVP, but LVV is constant at its low level (Fig. 16–8). This is because, once again, both mitral and aortic valves are closed; this phase is called *isovolumetric relaxation.*

Eventually, LVP falls below the pressure existing in the left atrium and the mitral valve opens (at time D). At this point, blood flows from the LA into the LV as indicated by the rise of LVV; also note the slight rise in LVP as filling proceeds. This phase is called *filling.* In general terms, systole includes isovolumetric contraction and ejection; diastole includes isovolumetric relaxation and filling.

Whereas the four phases of the cardiac cycle are clearly illustrated on the plots of LVV, LVP and AoP as a function of time, it turns out that there are many advantages to displaying LVP as a function of LVV on a "pressure-volume diagram" (these advantages will be made clear by the end of the chapter). This is accomplished simply by plotting the simultaneously measured LVV and LVP on appropriately scaled axes; the resulting pressure-volume diagram corresponding to the curves of Fig. 16–8 is shown in Fig. 16–9, with volume on the x-axis and pressure on the y-axis.

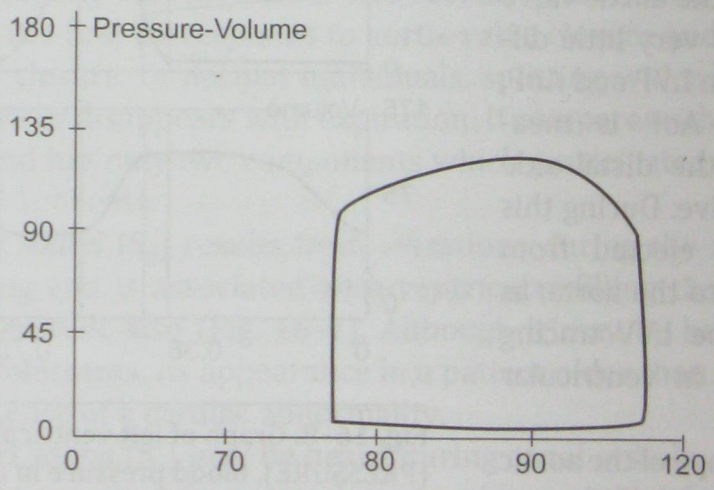


Fig. 16–9. The graph of left ventricle pressure (y-axis) dependence upon the left ventricle volume (x-axis).

As shown, the plot of pressure versus volume for one cardiac cycle forms a loop. This loop is called the *pressure-volume loop* (abbreviated PV loop). As time proceeds, the PV points go around the Figure loop in a counter clockwise direction. The point of maximal volume and minimal pressure (i.e., the bottom right corner of the loop) corresponds to time A on Fig. 16–8, the onset of systole. During the first part of the cycle, pressure rises but volume stays the same (isovolumetric contraction).

Ultimately, LVP rises above AoP, the aortic valve opens, ejection begins and volume starts to go down. (Obviously, with this representation, AoP is not explicitly plotted; however, as will be reviewed below, several features of AoP are readily obtained from the PV loop.) After the ventricle reaches its maximum activated state (upper left corner of PV loop), LVP falls below AoP, the aortic valve closes and isovolumetric relaxation commences. Finally, filling begins with mitral valve opening (bottom left corner).

4. PHYSIOLOGIC MEASUREMENTS RETRIEVABLE FROM THE PRESSURE-VOLUME LOOP

As reviewed above, the ventricular pressure-volume loop displays the instantaneous relationship between intraventricular pressure and volume throughout the cardiac cycle. It turns out that with this representation it is easy to ascertain values of several parameters and variables of physiologic importance.

Consider first the volume axis (Fig. 16–10). It is appreciated that we can readily pick out the maximum volume of the cardiac cycle. This volume is called the *end-diastolic volume* (EDV) — equal to an average 130–140 ml of blood (Fig. 16–8) — because this is the ventricular volume at the end of a cardiac cycle. Also, the minimum volume the heart attains is also retrieved; this volume is known as the *end-systolic volume* (ESV) and is the ventricular volume at the end of the ejection phase. This volume equals to 40–60 ml. The difference between EDV and ESV represents

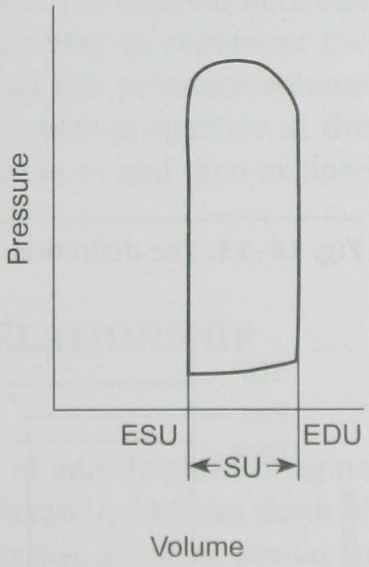


Fig. 16–10. Schematical representation of the ventricles volumes during the cardiac cycle.

the amount of blood ejected during the cardiac cycle and is called the *stroke volume* (SV) (Fig. 16–11).

Now consider the pressure axis (Fig.16–12). Near the top of the loop we can identify the point at which the ventricle begins to eject (that is, the point at which volume starts to decrease) is the point at which ventricular pressure just exceeds aortic pressure; this pressure therefore reflects the pressure existing in the aorta at the onset of ejection and is called the *diastolic blood pressure* (DBF).

During the ejection phase, aortic and ventricular pressures are essentially equal; therefore, the point of greatest pressure on the loop also represents the greatest pressure in the aorta, and this is called the *systolic blood pressure*

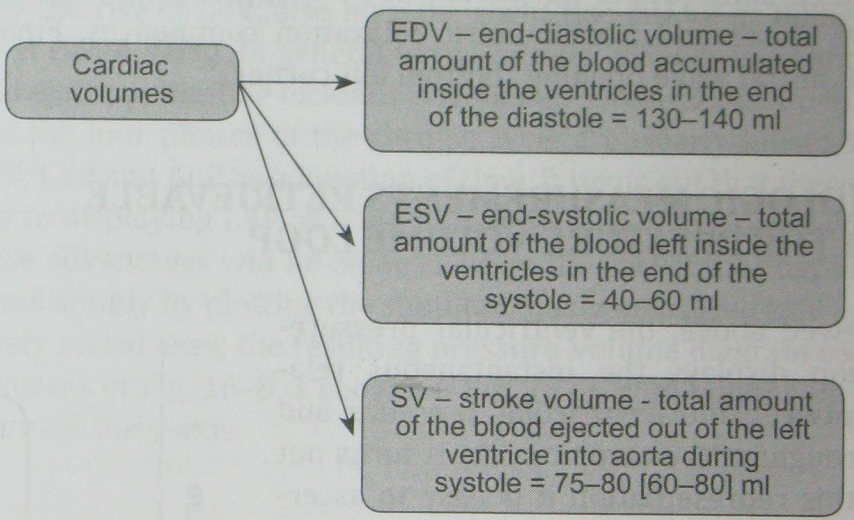


Fig. 16–11. The definitions and range of cardiac volumes.

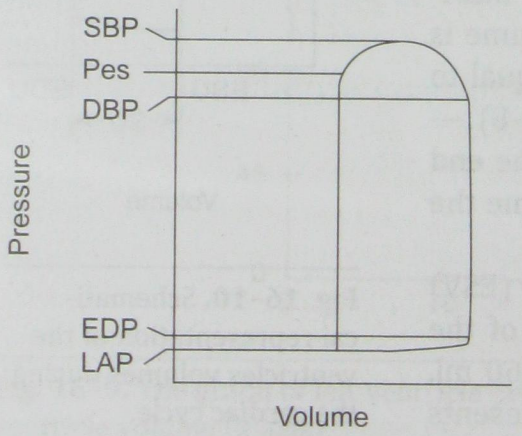


Fig. 16–12 The detailed explanation of the pressure changes within the ventricles during the heart cycle.

(SBP). One additional pressure, the *end-systolic pressure* (Pes) is identified as the pressure of the left upper corner of the loop; the significance of this pressure will be discussed in detail below. Moving to the bottom of the loop, we can reason that the pressure of the left lower corner (the point at which the mitral valve opens and ejection begins) is roughly equal to the pressure existing in the left atrium (LAP) at that instant in time (recall that atrial pressure is not a constant, but varies with atrial contraction and instantaneous atrial volume).

The pressure of the point at the bottom right corner of the loop is the pressure in the ventricle at the end of the cardiac cycle and is called the *end-diastolic pressure* (EDP).

5. PRESSURES-VOLUME RELATIONSHIPS

With each cardiac cycle, the muscles in the ventricular wall contract and relax causing the chamber to stiffen (reaching a maximal *stiffness* at the end of systole) and then to become less stiff during the relaxation phase (reaching its minimal *stiffness* at end-diastole).

Thus, the mechanical properties of the ventricle are time-varying, they vary in a cyclic manner, and the period of the cardiac cycle is the interval between beats. In the following discussion we will explore one way to represent the time-varying mechanical properties of the heart using the pressure-volume diagram. We will start with a consideration of ventricular properties at the extreme states of *stiffness* — end systole and end diastole — and then explore the mechanical properties throughout the cardiac cycle.

6. END-DIASTOLIC PRESSURE-VOLUME RELATIONSHIP (EDPVR)

Let us first examine the properties of the ventricle at end-diastole. Imagine the ventricle frozen in time in a state of complete relaxation. We can think of the properties of this ventricle with weak, relaxed muscles, as being similar to those of a floppy balloon. What would happen to the pressure inside a floppy balloon if we were to vary its volume?

Let's start with no volume inside the balloon; naturally there would be no pressure. As we start blowing air into the balloon there is initially little resistance to our efforts as the balloon wall expands to a certain point. Up to that

point, the volume increases but pressure does not change. We will refer to this volume as Vo, or the maximal volume at which pressure is still zero mmHg; this volume is also frequently referred to as the unstressed volume.

As the volume increases we meet with increasing resistance to our efforts to expand the balloon, indicating that the pressure inside the balloon is becoming higher and higher. The ventricle, *frozen* in its diastolic state, is much like this balloon. A typical relationship between pressure and volume in the ventricle at the end of diastole is shown in Fig. 16–13.

As the volume is increased initially, there is little increase in pressure until a certain point, designated "Vo". After this point, pressure increases with further increases in volume. The uantitative analysis of such curves measured from animal as well as from patient hearts has shown that pressure and volume are related by a nonlinear function —

$$EDP = k\{e^{a}(V-V_0) - 1\}, \qquad [16.1]$$

where EDP is the end-diastolic pressure,
V is the volume inside the ventricle,

Vo is as defined above, and k and a are constants which specify the curvature of the line and are determined by the mechanical properties of the muscle as well as the structural features of the ventricle. This curve is called the "end-diastolic pressure-volume relationship" (EDPVR).

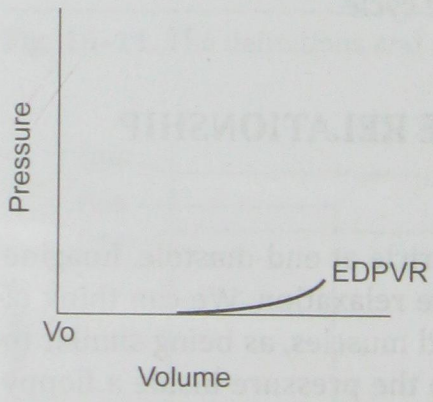


Fig. 16–13. Graphical illustration of end-diastolic pressure-volume relationship.

Under normal conditions, the heart would never exist in such a frozen state as proposed above. However, during each contraction there is a period of time during which the mechanical properties of the heart are characterized by the EDPVR; knowledge of the EDPVR allows one to specify, for the end of diastole, EDP if EDV is known, or visa versa. Furthermore, since the EDPVR provides the pressure-volume relation with the heart in its most relaxed state, the EDPVR provides a boundary on which the PV loop *must* fall at the end of the cardiac cycle.

Under certain circumstances, the EDPVR may change. Physiologically, the EDPVR

changes as the heart grows during childhood. Most other changes in the EDPVR accompany pathologic situations; examples include the changes which occur with hypertrophy, the healing of an infarct, and the evolution I of a dilated cardiomyopathy, to name a few.

There is a term which is frequently used in discussions of the end-diastolic ventricular properties: *compliance.* Technically, *compliance* is the change in volume for a given change in pressure or, expressed in mathematical terms, it is the reciprocal of the derivative of the EDPVR. Since the EDPVR is nonlinear, the compliance varies with volume; compliance is greatest at low volume and smallest at high volumes. In the clinical arena, however, *compliance* is used in two different ways.

First, it is used to express the idea that the diastolic properties are, in a general way, altered compared to normal; that is, that the EDPVR is either elevated or depressed compared to normal.

Second, it is used to express the idea that the heart is working at a point on the EDPVR where its slope is either high or low (this usage is technically more correct). Undoubtedly you will hear this word used in the clinical setting, usually in a casual manner: "The patients heart is noncompliant". Such a statement relays no *specific* information about what is going on with the diastolic properties of the heart. Statements specifying changes in the EDPVR or changes in the working volume range relay much more information.

Understanding of these concepts has been highlighted recently with growing appreciation for the fact that some patients can experience heart failure when the EDPVR becomes elevated (as in hypertrophy) despite that fact that the strength of the heart during contraction is normal. This clinical phenomenon has been referred to as *diastolic dysfunction.*

7. END-SYSTOLIC PRESSURE-VOLUME RELATIONSHIP (ESPVR)

Let us move now to the opposite extreme in the cardiac cycle: end-systole. At that instant of the cardiac cycle, the muscles are in their maximally *activated* state and it is easy to imagine the heart as a much stiffer chamber. As for end diastole, we can construct a pressure-volume relationship at end systole if we imagine the heart frozen in this state of maximal activation.

An example is shown in Fig.16–14. As for the EDPVR, the end-systolic pressure volume relationship (ESPVR) intersects the volume axis at a slightly

positive value (Vo), indicating that a definite amount of volume must fill the ventricle before it can generate any pressure.

For our purposes, we can assume that the Vo of the ESPVR and the Vo of the EDPVR are the same (this is not exactly true, but little error is made in assuming this and it simplifies further discussions). In contrast to the nonlinear EDPVR, the ESPVR has been shown to be reasonably linear over a wide range of conditions, and can therefore be expressed by a simple equation:

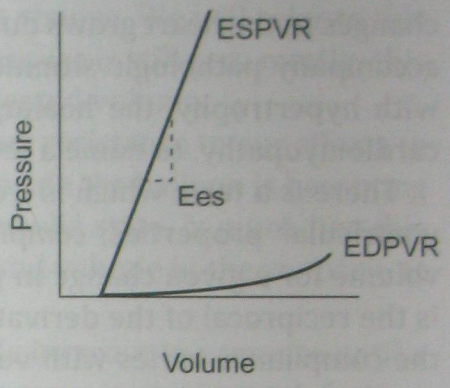


Fig. 16-14. Graphical illustration of end-diastolic pressure-volume relationship and end-systolic pressure-volume relationship interposition with the explanation of elastance phenomenon.

$$\mathbf{Pes = Ees(V—Vo),} \qquad [16.2]$$

where Pa is the end-systolic pressure,
Vo is as defined above,
V is the volume of interest
and E_a is the slope of the linear relation.

There isn't a *priori* reason to expect that this relationship should be linear, it is simply an experimental observation. E_{es} stands for *end systolic elastance. Elastance* means essentially the same thing as *stiffness*, and is defined as the change in pressure for a given change in volume within a chamber; the higher the elastance, the stiffer the wall of the chamber.

As discussed above for the EDPVR, the heart would never exist in *a frozen* state of maximal activation. However, it does pass through this state during each cardiac cycle. The ESPVR provides a line which the PV loop will hit at end-systole, thus providing a boundary for the upper left hand corner of the PV loop. We will discuss further about the properties of the ESPVR after reviewing the mechanical properties of the ventricle throughout the cardiac cycle.

8. TIME VARYING ELASTANCE: E(t)

In the above discussion we have described the pressure-volume relationships at two instances in the cardiac cycle: end diastole and end systole. The idea of considering the pressure-volume relation with the heart frozen in a

given state can be generalized to any point during the cardiac cycle. That is, at each instant of time during the cardiac cycle there exists a pressure-volume relationship. Such relations have been determined in the physiology laboratory. These experiments show, basically, that there is a relatively smooth transition from the EDPVR to the ESPVR and back. For most parts of the cardiac cycle these relations can be considered (1) to be linear and (2) to intersect at a common point, namely Vo. This idea is schematized in Fig. 16–15.

In the left panel, the transition from the EDPVR towards the ESPVR during the contraction phase is illustrated, and the relaxation phase is depicted in the right panel. Since the instantaneous pressure-volume relations (PVR) are reasonably linear and intersect at a common point, it is possible to characterize the time course of change in ventricular mechanical properties by plotting the time course of change in the slope of the instantaneous PVR.

Above, we referred to the slope of the ESPVR as an *elastance.* Similarly, we can refer to the slopes of the instantaneous PVRs as *elastances.* A rough approximation of the instantaneous elastance throughout a cardiac cycle is shown in Fig. 16–16. Note that the maximal value, E_{es}, is the slope of the ESPVR. The minimum slope, E_{min}, is the slope of the EDPVR in the low volume range.

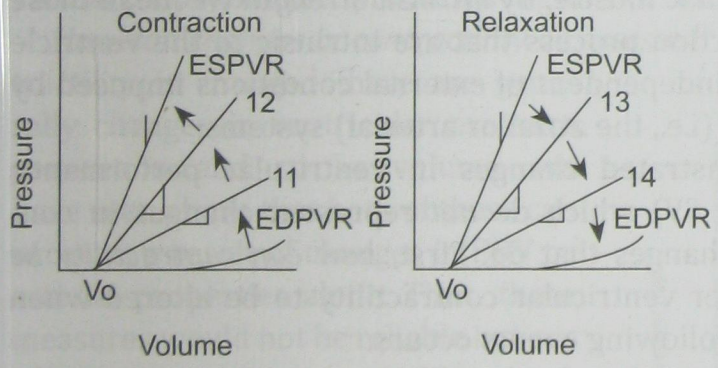


Fig. 16–15. Schematical illustration of contraction-relaxation processes dependence upon the ESPVR and EDPVR values.

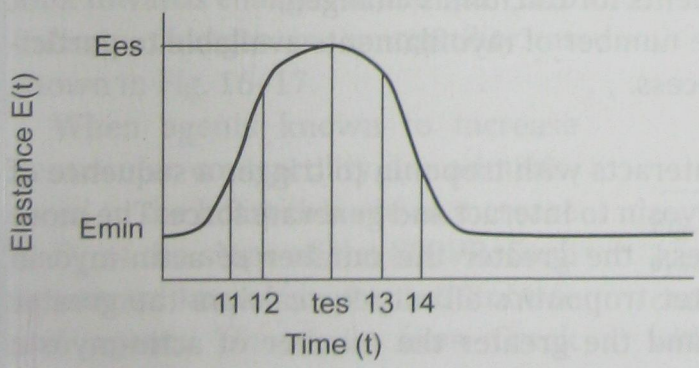


Fig. 16–16. Graphical illustration of the time-elastance dependence throughout the cardiac cycle.

We refer to the function depicted in Fig. 16–16 as the *time varying elastance* and it is referred to as E(t).

With this function it is possible to relate the instantaneous pressure (P) and volume (V) throughout the cardiac cycle:

$$P(V,t) = E(t)\,[V(t) - Vo], \qquad [16.3]$$

where Vo and E(t) are as defined above and V(t) is the time varying volume.

This relationship breaks down near end-diastole and early systole when there are significant nonlinearities in the pressure-volume relations at higher volumes. The implication of this equation is that if one knows the E(t) function and if one knows the time course of volume changes during the cycle, one can predict the time course of pressure changes throughout the cycle.

9. CONTRACTILITY

Contractility is an ill-defined concept used when referring to the intrinsic strength of the ventricle or cardiac muscle. By *intrinsic strength* we mean those features of the cardiac contraction process that are intrinsic to the ventricle (and cardiac muscle) and are independent of external conditions imposed by either the preload or afterload (i.e., the atrial or arterial) systems.

Now that we have demonstrated changes in ventricular performance (i.e., pressure generation and SV) which do *not* represent changes in contractility, let's explore some changes that do. First, *how can contractility be changed?* Basically, we consider ventricular contractility to be altered when anyone or combination of the following events occurs:

1) the amount of calcium released to the myofilaments is changed;
2) the affinity of the myofilaments for calcium is changed;
3) there is an alteration in the number of myofilaments available to participate in the contraction process.

You will recall that calcium interacts with troponin to trigger a sequence of events which allows actin and myosin to interact and generate force. The more calcium available for this process, the greater the number of actin-myosin interactions. Similarly, the greater troponin's affinity for calcium the greater the amount of calcium bound and the greater the number of actin-myosin

interactions. Here we are linking *contractility* to cellular mechanisms which underlie excitation-contraction coupling and thus, changes in ventricular contractility would be the global expression of changes in contractility of the cells that make up the heart. Stated another way, ventricular contractility is reflects *myocardial contractility* (the contractility of individual cardiac cells).

Through the third mechanism, changes in the number of muscle cells, as apposed to the functioning of any given muscle cell, cause changes in the performance of the ventricle as an organ. In acknowledging this as a mechanism through which ventricular contractility can be modified we recognize that ventricular contractility and myocardial contractility are <u>not</u> always linked to each other.

Humoral and pharmacologic agents can modify ventricular contractility by the first two mechanisms. β-adrenergic agonists (e.g. epinephrine) increase the amount of calcium released to the myofilaments and cause an increase in contractility. In contrast, β-adrenergic antagonists (e.g., propranolol) blocks the effects of circulating epinephrine and norepinephrine and reduces contractility. Nifedipine is a drug that blocks entry of calcium into the cell and therefore reduces contractility. One example of how ventricular contractility can be modified by the third mechanism mentioned above is the reduction in ventricular contractility following a myocardial infarction where there is loss of myocardial tissue, but the unaffected regions of the ventricle function normally.

How can changes in contractility be assessed? While it is true that when contractility is changed there are generally changes in ventricular pressure generation and stroke volume, we have seen above that both of these can occur as a result of changes in EDV or arterial properties alone. Thus, these measures would not be reliable indices of contractility. It turns out that we can look towards changes in the ESPVR to indicate changes in contractility, as shown in Fig. 16–17.

When agents known to increase ventricular contractility are administered to the heart there is an increase in E_{es} of the slope of the ESPVR. Such agents are known as positive "inotropic" agents. *(Inotropic:* from Greek

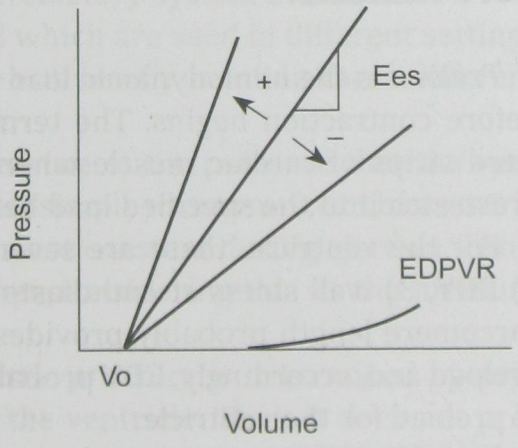


Fig. 16–17 Graphical illustration of end-systolic pressure-volume relationship changes throughout the cardiac cycle with the explanation of properties.

meaning influencing the contractility of muscular tissue). Conversely, agents which are negatively inotropic reduce E_0. It is significant that neither Vo (the volume-axis intercept of the ESPVR) nor the EDPVR are affected significantly by these acute changes in contractility. Thus, because E_0 varies with ventricular contractility but is not affected by changes in the arterial system properties nor changes in EDV, E_a is considered to be an index of contractility.

The major draw back to the use of ESE in the clinical setting is that it is very difficult to measure ventricular volume. Clearly, it is required that volume must be measured in the assessment of E_0. Currently, the most commonly employed index of contractility in the clinical arena is *ejection fraction* (EF). EF is defined as the ratio between EDV and SV:

$$\mathbf{EF = SV/EDV} \qquad [16.4]$$

This number ranges from 0 to 1 (or from 0 % to 100 %) and represents the fraction of the volume present at the start of the contraction that is ejected during the contraction. The normal value of EF ranges between 0.6 and 0.7. EF can be estimated by a number of techniques, including echocardiography and nuclear imagining techniques (MUGA). The main disadvantage of this index is that it is a function of the properties of the arterial system.

10. PRELOAD

Preload is the hemodynamic load on the ventricle at the end of diastole just before contraction begins. The term was originally coined in studies of isolated strips of cardiac muscle where a weight was hung from the muscle to prestretch it to the specified load before *(pre-)* contraction.

For the ventricle, there are several possible measures *of preload:* 1) EDP, 2) EDV, 3) wall stress at end-diastole and 4) end-diastolic sarcomere length. Sarcomere length probably provides the most meaningful measure of muscle preload and, accordingly, EDV probably provides the most meaningful measure of preload for the ventricle.

However, EDP is simple to measure and, if the EDPVR is known serves to localize the PV loop on the PV diagram as well. EDP can be assessed clinically by measuring the pulmonary capillary wedge pressure (PCWP) using a Swan- Ganz catheter that is placed through the right ventricle into the pulmonary artery.

11. AFTERLOAD

Afterload is the hydraulic *load* imposed on the ventricle during ejection, usually by the arterial system. Under pathologic conditions when either the mitral valve is incompetent (i.e., leaky) or the aortic valve is stenotic (i.e., constricted) *afterload* is determined by factors other than the properties of the arterial system.

Each case in Fig. 16–18 (center and left) demonstrates that the active portion of the length-tension curve provides the limit to shortening and, thus, interacts with the particular afterload chosen. With smaller afterloads, the muscle will shorten further than it would with a larger afterload. It is important to realize that during isotonic shortening, the muscle force is limited by the magnitude of the afterload and *not* by the length-tension capability of the muscle. It is the *extent of shortening* at a given afterload that is limited by the length tension property of the muscle; this is a very important consideration when measuring cardiac performance under conditions of changing blood pressure and filling of the heart.

This length- and force-dependent behavior is the key to *autoregulation* (self-regulation) by cardiac muscle and is the functional basis of *Starling's law of the heart;* when the muscle is set to a longer length at rest, the active contraction results in a greater shortening that is also more rapid and is preceded by a more rapid isometric phase. This allows the heart to adjust its pumping to exactly the amount required to keep the circulatory system in balance.

There are several measures of afterload which are used in different settings (clinical *versus* basic science settings). We will briefly mention three different measures of afterload.

1. ***Aortic Pressure.*** This provides a measure of the pressure that the ventricle must overcome to eject blood. It is simple to measure, but has several shortcomings. First, aortic pressure is not constant during ejection. Thus, many people use the mean value when considering this as the measure of afterload.

Second, as will become clear below, aortic pressure is determined by properties of both the arterial system and of the ventricle. Thus, it does not provide a measure which relays information exclusively about the arterial system.

2. ***Total Peripheral Resistance.*** The total peripheral resistance (TPR) is the ratio between the mean pressure drop across the arterial system [which is equal to the mean aortic pressure (MAP) minus the central venous pressure (CVP)] and mean flow into the arterial system [which is equal

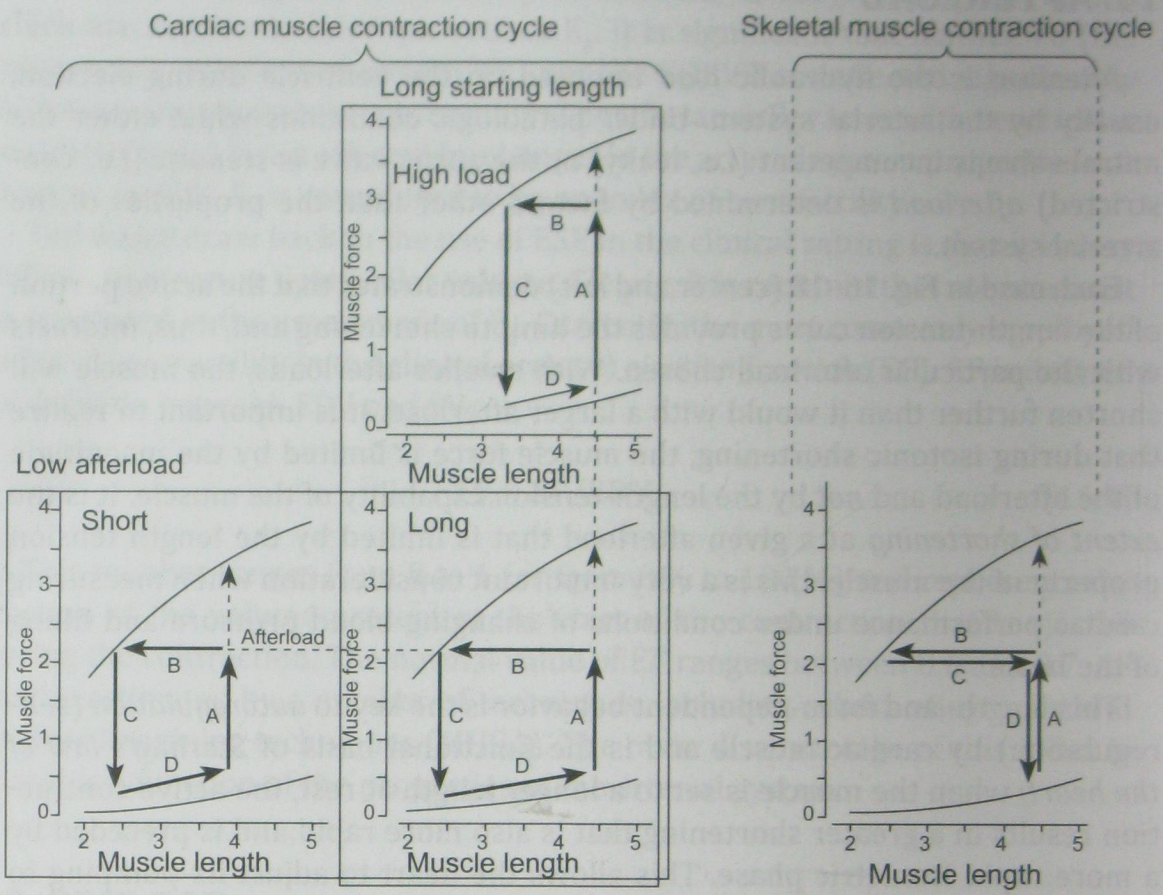


Fig. 16–18. Afterloaded contractions under a variety of conditions.

Left: Cardiac muscle contraction cycles.

The horizontal box shows the effect of starting at two different initial lengths at the same afterload. The vertical box shows the effect of two different afterloads on shortening that begins at the same initial length. Increasing the afterload reduces the amount of shortening possible, as does decreasing the starting length; in both cases, the limit to shortening is determined by the length-tension curve.

Right: The contraction cycle of skeletal muscle.

Contraction and relaxation pathways are the same.

to the cardiac output (CO)]. Unlike aortic pressure by itself, this measure is independent of the functioning of the ventricle. Therefore, it is an index which describes arterial properties. According to its mathematical definition, it can only be used to relate mean flows and pressures through the arterial system.

3. ***Total Peripheral Resistance.*** The total peripheral resistance (TPR) is the ratio between the mean pressure drop across the arterial system [which is equal to the mean aortic pressure (MAP) minus the central venous pressure (CVP)] and mean flow into the arterial system [which is equal to the cardiac output (CO)]. Unlike aortic pressure by itself, this measure is independent of the functioning of the ventricle. Therefore, it is an index which describes arterial properties. According to its mathematical definition, it can only be used to relate mean flows and pressures through the arterial system.
4. ***Arterial Impedance.*** This is an analysis of the relationship between pulsatile flow and pressure waves in the arterial system. It is based on the theories of Fourier analysis in which flow and pressure waves are decomposed into their harmonic components and the ratio between the magnitudes of pressure and flow waves are determined on a harmonic-by-harmonic basis. Thus, in simplistic terms, impedance provides a measure of resistance at different frequencies. Unlike TPR, impedance allows one to relate instantaneous pressure and flow. It is more difficult to understand, most difficult to measure, but the most comprehensive description of the properties of the arterial system as they pertain to understanding the influence of afterload on ventricular performance.

12. UNDERSTANDING THE DETERMINANTS OF VENTRICULAR PERFORMANCE

Two primary measurements of overall cardiovascular performance are the arterial blood pressure and the cardiac output.

These measures are also of primary concern in the clinical settings since both adequate blood pressure and adequate cardiac output are necessary to maintain life. It is important to appreciate, however, that both of these variables are determined by the interaction between the heart the arterial system (afterload) and the venous system (preload); this is a fundamental concept. Furthermore, it is important to develop an appreciation for how the heart and vasculature interact to determine these indices of performance.

This is highlighted by the fact that one major facet of intensive care medicine deals with maintaining adequate blood pressure and cardiac output by manipulating ventricular contractility, arterial resistance and ventricular preload.

Two approaches to understanding how these parameters regulate cardiovascular performance will be reviewed: a classical approach, commonly

referred to as *Frank-Starling Curves,* and a more modern approach based upon pressure-volume analyses.

13. FRANK-STARLING CURVES

Otto Frank (1890) is credited with the empirical observation that peak ventricular pressure increases as the end-diastolic volume is increased (as in Fig. 16–14). This observation was made in an isolated frog heart preparation in which ventricular volume could be measured with relative ease. Though of primary importance, the significance may not have been appreciated to the degree it could have been because it was (and remains) difficult to measure ventricular volume in more intact settings (e.g., experimental animals or patients); thus it was difficult for other investigators to study the relationship between pressure and volume in these more relevant settings.

Around the mid 1910's, Starling and coworkers observed a related phenomenon, which they presented in a manner that was much more useful to physiologists and ultimately to clinicians. They measured the relationship between ventricular filling pressure (related to end-diastolic volume) and cardiac output (CO=SV · HR). They showed that there was a nonlinear relationship between end-diastolic pressure (EDP, also referred to as ventricular *filling pressure)* and CO as shown in Fig. 16–19; as filling pressure was increased in the low range there is a marked increase in CO, whereas the slope of this relationship becomes less steep at higher filling pressures.

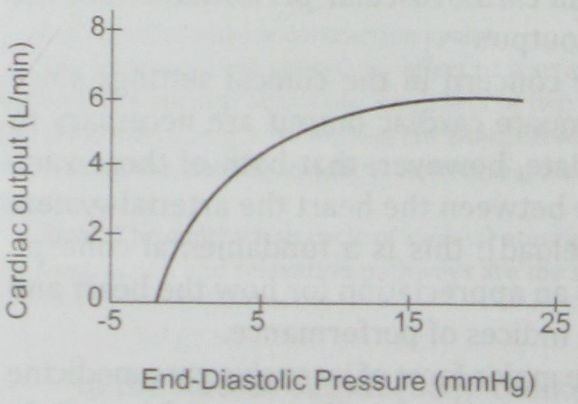


Fig. 16–19. Graphical illustration of end-diastolic pressure dependence upon the value of cardiac volume.

The observations of Frank and of Starling form one of the basic concepts of cardiovascular physiology: *cardiac performance (vs ability to generate pressure or to pump blood) increases with preload.* There are a few caveats, however. Recall from the anatomy of the cardiovascular system that left ventricular filling pressure is approximately equal to pulmonary venous pressure. As pulmonary venous pressure rises, there is an increased tendency

(Starling Forces) for fluid to leak out of the capillaries and into the interstitial space and alveoli.

When this happens, there is impairment of gas exchange across the alveoli and hemoglobin oxygen saturation can be markedly diminished. This phenomenon typically comes into play when pulmonary venous pressure rises above ~20 mmHg and becomes increasingly prominent with further increases. When pulmonary venous pressures increases above 30 mmHg, there is profound transudation of fluid into the alveoli and *pulmonary edema* is usually prominent. Therefore, factors extrinsic to the heart dictate a practical limit to how high filling pressure can be increased.

As noted above, factors other than preload are important for determining cardiac performance: ventricular *contractility* and afterload properties. Both of these factors can influence the Frank-Starling Curves. When ventricular contractile state is increased, CO for a given EDP will increase and when contractile state is depressed, CO will decrease (Fig. 16–20).

When arterial resistance is increased, CO will decrease for a given EDP while CO will increase when arterial resistance is decreased (Fig. 16–21).

Thus, shifts of the Frank-Starling curve are nonspecific in that they may signify either a change in contractility or a change in afterload. It is for this reason that Starling-Curves are *not* used as a means of indexing ventricular contractile strength.

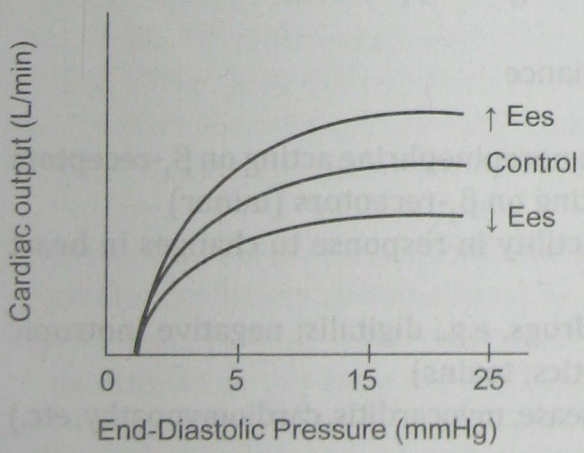


Fig. 16–20. Graphical illustration of end-diastolic pressure dependence upon the value of ventricular contractility.

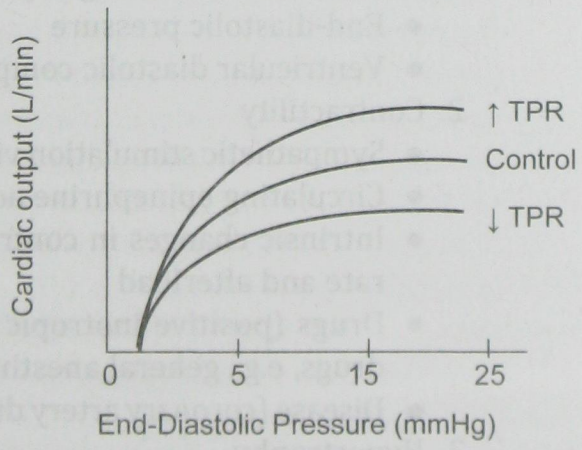


Fig. 16–21. Graphical illustration of end-diastolic pressure dependence upon the value of arterial resistance.

14. CARDIAC OUTPUT

Cardiac output (CO) is defined as the volume of blood ejected from the heart per unit time. The usual resting values for adults are 5 to 6 L/min, or approximately 8 % of body weight per minute. Cardiac output divided by body surface area is called the **cardiac index.** When it is necessary to normalize the value to compare the cardiac output among individuals of different sizes, either cardiac index or cardiac output divided by body weight can be used.

Cardiac output is the product of heart rate (HR) and **stroke volume** (SV), the volume of blood ejected with each beat:

$$\mathbf{CO = SV \cdot HR} \qquad [16.5]$$

Stroke volume is the difference in the volume of blood in the ventricle at the end of diastole — **end-diastolic volume** — and the volume of blood in the ventricle at the end of systole — **end-systolic volume**.

If heart rate remains constant, cardiac output increases in proportion to stroke volume, and stroke volume increases in proportion to cardiac output. The factors that influence cardiac output are outlined lower.

Factors Influencing Cardiac Output

I. Stroke volume
  A. Force of contraction
    1. End-diastolic fiber length (Starling's law, preload)
      - End-diastolic pressure
      - Ventricular diastolic compliance
    2. Contractility
      - Sympathetic stimulation via norepinephrine acting on β_1-receptors
      - Circulating epinephrine acting on β_1-receptors (minor)
      - Intrinsic changes in contractility in response to changes in heart rate and afterload
      - Drugs (positive inotropic drugs, e.g., digitalis; negative inotropic drugs, e.g., general anesthetics; toxins)
      - Disease (coronary artery disease, myocarditis, cardiomyopathy, etc.)
    3. Hypertrophy
  B. Afterload
      - Ventricular radius
      - Ventricular systolic pressure

II. Heart rate (and pattern of electrical excitation)

Ejection fraction (EF) is a commonly used measure of cardiac performance. It is the ratio of stroke volume to end-diastolic volume (EDV), expressed as a percentage:

$$\mathbf{EF = (SV \backslash EDV) \cdot 100} \qquad [16.6]$$

Ejection fraction is normally more than 55 %. It is dependent on heart rate, preload, afterload, and contractility and provides a nonspecific index of ventricular function. Still, it has proved to be valuable in predicting the severity of heart disease in individual patients.

Cardiac output is regulated by changing stroke volume and heart rate. Stroke volume is influenced by the contractile force of the ventricular myocardium and by the force opposing ejection (the aortic pressure or afterload). Myocardial contractile force depends on ventricular end-diastolic fiber length (Starling's law) and myocardial contractility.

Chapter 17. Electrical displays of cardiac activity

The heart beats in the absence of any nervous connections because the electrical (pacemaker) activity that generates the heartbeat resides within the cardiac muscle. After initiation, the electrical activity spreads throughout the heart, reaching every cardiac cell rapidly with the correct timing. This enables coordinated contraction of individual cells.

Movement of a depolarizing electrical wave over the heart is obligatory for contraction of the heart muscle and the pumping of an adequate amount of blood to the body. Abnormalities in the electrical properties of the heart result from cardiac disease and compromise the ability of the heart to pump blood.

1. THE NORMAL ELECTRICAL ACTIVITY OF THE HEART

Cardiac electrical activity is normally initiated and spread in an orderly fashion. The heart is said to be *a functional syncytium* because the excitation of one cardiac cell eventually leads to the excitation of all cells.

1. ***The sinus node initiates electrical activity.*** In the normal heart, an electrical impulse in the form of an action potential, arises in a special region of the superior right atrium called the sinus node (Figs. 17–1, 17–2).
2. ***Atrial activation.*** The impulse that has arisen in the sinus node propagates into the right atrial myocardium. From here, it moves throughout the right and left atrial myocardium in a broad wave front at a velocity of about *70-90 cm/sec*
3. Atrioventricular conduction. In the normal heart, the atria and ventricles are connected by a narrow bridge of muscle which runs from the inferior part of the right atrium to the superior part of the ventricular septum. This narrow bridge begins as a structure called the atrioventricular (AV) node. The electrical impulse from the atria conducts through the AV node at a very slow speed (*5-7 cm/sec*). The slower conduction velocity is partially explained by the small size of the nodal cells. Less current is produced by the depolarization of a small nodal cell (compared with a large atrial or ventricular cell), and the relatively smaller current brings

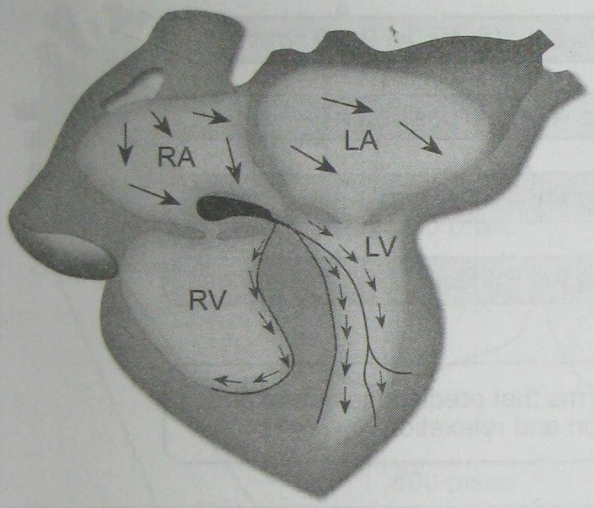


Fig. 17–1. Schematical illustration of the impulses spreading along myocardium.
RA — right atrium;
RV — right ventricle;
LA — left atrium;
LV — left ventricle;
S — sinus node;
BH — bundle of His.

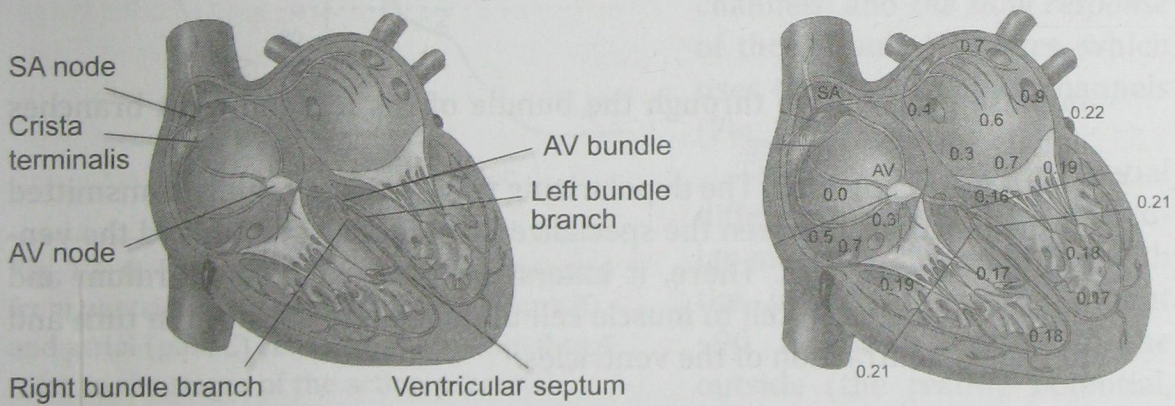


Fig. 17–2. The timing of excitation of different areas of the myocardium (in fractions of a second).

neighboring cells to threshold more slowly, decreasing the rate at which electrical activation spreads. Propagation of the action potential through the AV node takes approximately *10–12 msec* (Fig. 17–3).

Excitation then proceeds through the AV bundle (bundle of His), the left and right bundle branches, and the Purkinje system.

4. The AV node is connected to a structure called the *bundle of His* (Fig. 17–2). The bundle of His is directly connected to the Ventricular Specialized Conducting System. The conducting system is formed by a special kind of cardiac muscle cell, called the Purkinje cell. The impulse from the AV node spreads rapidly throughout this conducting system at a velocity of *100-300 cm/sec* along the bundle of His. The velocity of

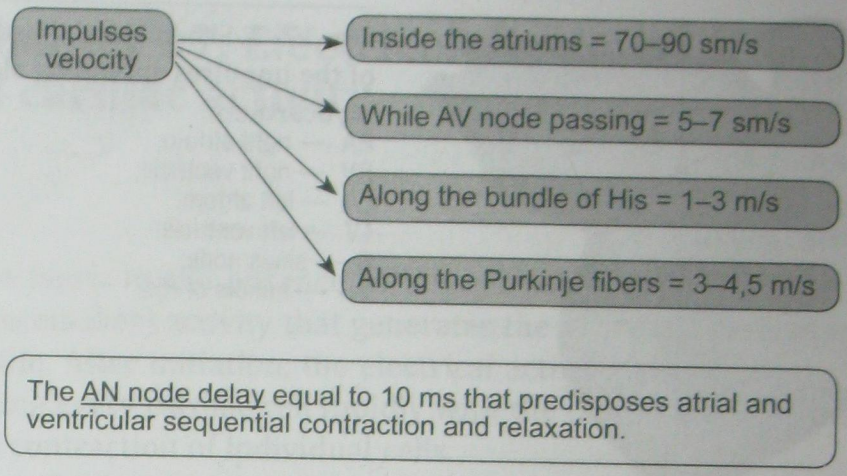


Fig. 17–3. Velocity of the impulses spreading along the heart conductive system.

impulse propagation through the bundle of His left and right branches equals *300–450 cm/sec.*

5. *Ventricular activation.* The depolarizing wave front is rapidly transmitted to the junctions between the specialized conducting system and the ventricular muscle cells. There, it enters the ventricular myocardium and moves from muscle cell to muscle cell to depolarize each cell in turn and to trigger contraction of the ventricles.

This sequence of events occurs once for each heart beat.

The sinus node, atrioventricular node, His bundle and ventricular specialized conducting system constitute the *specialized fibers of the heart.* Atrial and ventricular muscle constitute the *working myocardium.*

2. CELLULAR ORIGIN OF ELECTRICAL ACTIVITY IN THE HEART

1. The cardiac muscle action potential

There are many similarities between the origin of the nerve action potential and the cardiac action potential. Besides, potentials of action registered from different parts of the heart are different also (Fig. 17–4). There are also aspects of the cardiac action potential that are unique.

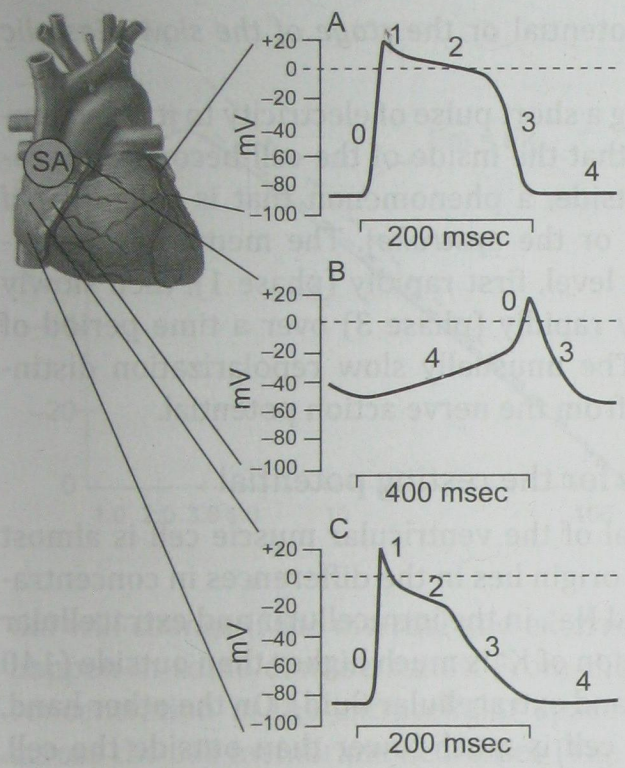


Fig. 17–4. The potentials of action registered from ventricular (part A), sinoatrial (part B) and atrial (part C) cells. Numbers from 0 to 4 refer to the stages of the action potential.

Firstly, it should be mentioned that cardiac action potentials could be generated with the three types of ion channels participation: the fast Na^+-channels, the slow Ca^{2+}-Na^+-channels and the K^+ ones (Fig. 17–5).

Cardiac potentials of action could be of two types: *the fast response*, which occurs in atrial and ventricular muscle cells and the Purkinje conduction system and uses the fast Na^+ channels, and *the slow response* of the SA and AV nodes, which uses the slow calcium channels (Fig. 17–5).

There is a steady potential difference across the ventricular muscle cell membrane with the inside of the cell about –90 mV with respect to the outside (the resting potential,

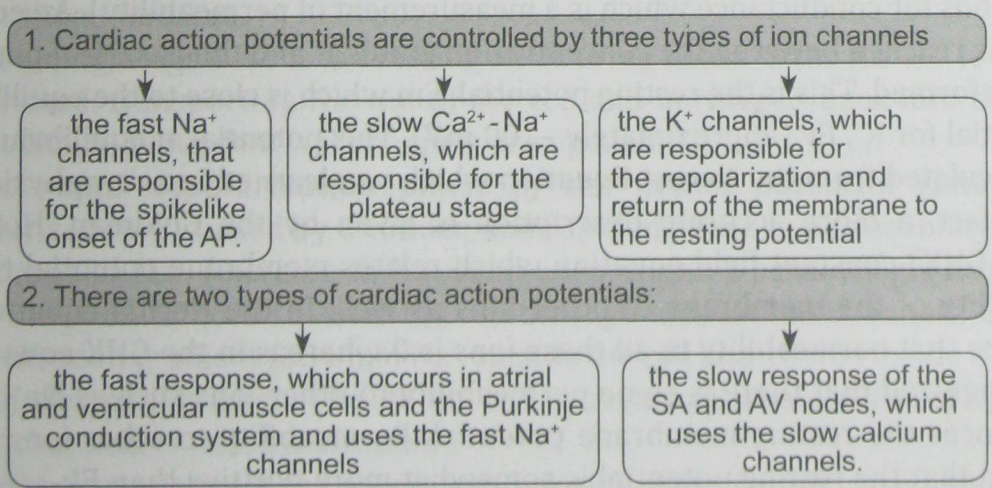


Fig. 17–5. The key features of cardiac excitability.

also called phase 4 of the action potential or the *stage of the slow diastolic depolarization*).

If we stimulate the cell by applying a short pulse of electricity to it, the membrane potential quickly changes so that the inside of the cell becomes approximately 30 mV positive to the outside, a phenomenon that is called *rapid depolarization* of the cell (phase 0, or the *upstroke*). The membrane potential gradually returns to the resting level, first rapidly (phase 1), then slowly (phase 2 or the plateau) and finally rapidly (phase 3) over a time period of about 200–300 msec (Fig. 17–4). The unusually slow repolarization distinguishes the cardiac action potential from the nerve action potential.

2. Cellular and Ionic Mechanisms for the resting potential

The basis for the resting potential of the ventricular muscle cell is almost identical to that of the nerve cell. Its origin lies in the differences in concentrations (activities) of the cations, K^+ and Na^+, in the intracellular and extracellular spaces. Inside the cell the concentration of K^+ is much higher than outside (140 mM versus about 4 mM in the blood and extracellular fluid). On the other hand, the concentration of Na^+ inside the cell is much lower than outside the cell, being about 10 mM inside and 154 mM outside.

During the resting state (stage 0 or slow diastolic depolarization), the membrane that separates the inside from the outside of the cell (the sarcolemma) is almost completely permeable to K^+ and almost completely impermeable to Na^+. As a result, K^+ ions diffuse down a concentration gradient from inside to outside the cell through membrane channels that have been designated as gK1 (G stands for conductance which is a measurement of permeability). An equilibrium is reached between the concentration gradient and the electrical potential that is formed. This is the resting potential, Em which is close to the equilibrium potential for K^+, Ev (approximately –100 mV). This potential at equilibrium can be calculated from the Nernst Equation which you learned in neurophysiology.

In fact, a more accurate description is given by the Goldman, Hodgkin, Kate (GHK) constant field equation which relates membrane potential to permeability of the membrane to other ions as well. In the Nernst equation we assume that permeability to all these ions is 0 whereas in the GHK equation it is recognized that there is some permeability to other ions such as Na^+ which influences the resting membrane potential. Permeability to other ions is the reason that the resting potential is somewhat more positive than Ek.

The Nernst equation shows us some very important properties of the resting membrane potential. Most importantly, it is obvious that the resting poten-

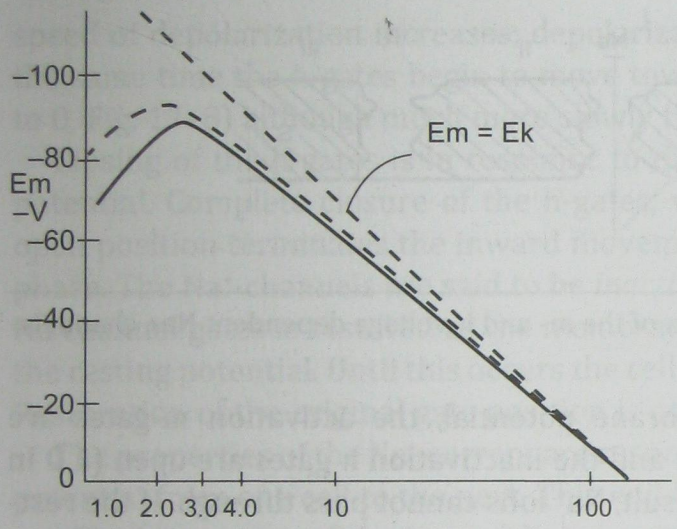


Fig. 17–6. Relationship between resting potential and extracellular K+ concentration. Interrupted line shows expected relationship from Ek-61.5 log [K+]Q/[K+]i;. Solid line shows data for ventricular muscle or Purkinje fibers. Xs show the increase in RP that would be expected if the permeability of the membrane to K+ were increased in relation to the permeability to other ions.

tial will change if the internal and external K^+ concentrations change. This can happen in some disease states. From the equation you can see that if the K^+ concentration in the extracellular space increa-ses, the membrane potential across the cell membrane decreases (Fig. 17–6).

As we reduce the extracellular K^+ concentration, membrane potential increases as expected until we get to about 2 mM. Reducing K^+ below this value causes an unexpected decrease in the membrane potential. The permeability of the membrane to K^+ decreases significantly at very low K^+ concentrations so that it is no longer freely permeable to K^+. When this occurs, the Na^+ permeability has a depolarizing influence and membrane potential decreases rather than increases.

Resuming, stage 4 — is the resting MP. During phase 4, the Na^+ — K^+ pump is activated, transporting Na^+ out of the cell and moving K^+ back into the cell. Slow voltage-dependent Na^+ channels are opened.

The rapid depolarization phase of the action potential (phase 0). Depolarization is caused by a sudden and rapid increase in permeability of the membrane to Na^+, (increase in Na^+ conductance, gNa^+) permitting a large and rapid influx of Na^+ ions to enter the cell through the fast Na^+ channels, down a concentration gradient (154 mM Na^+ outside the cell and 10 mM inside the cell) and an electrical gradient (inside the cell is negative). The process of opening of Na^+ channels is called *activation.*

The characteristics of the opening and closing of the fast Na^+ channels during the depolarization phase of the action potential can be described by the properties of the m- and h-voltage dependent "gates" around the channels (Fig. 17–7).

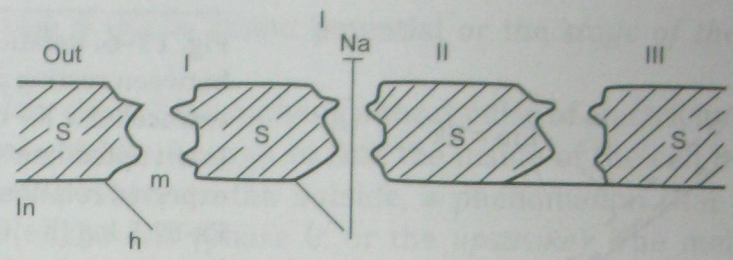


Fig. 17–7. Schematical representation of the m- and h-voltage dependent Na+ channels.

At the negative resting membrane potential, the activation m-gates are closed (at 0 in panel B, Fig. 20–7) and the inactivation h gates are open (1.0 in panels A and B, Fig. 17–8). As a result, Na^+ ions cannot pass through. If the resting membrane potential is depolarized by about -60 mV (the threshold potential), the m-gates begin to open rapidly (move from 0 to 1 in panel B, Fig. 17–8), Na^+ begins to enter the cell and the membrane potential starts to move in a positive direction. As membrane potential moves in a positive direction, more and more m-gates open and the rate of movement of Na^+ into the cell and the

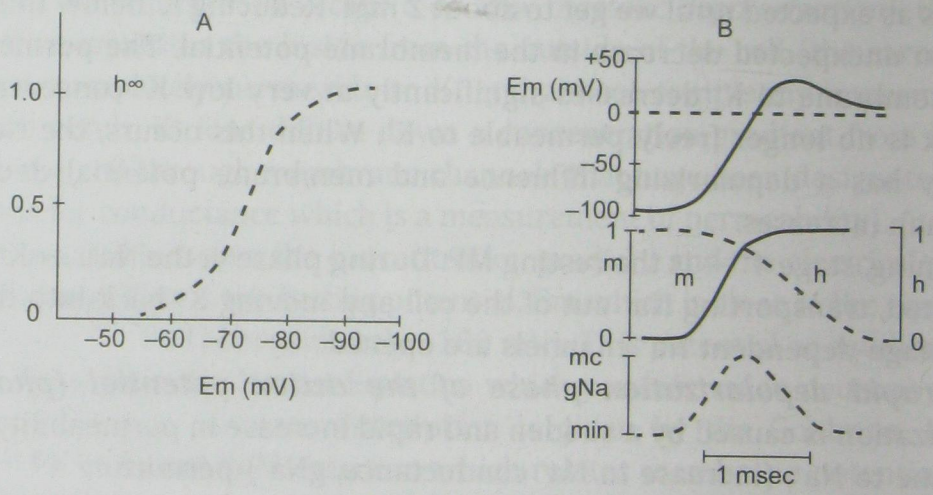


Fig. 17–8. The variables m (activation) and h (inactivation) vary with voltage and time. Sodium current is given by a relation of the following form: INa = gNm3h(Emj- ENa) where gNa is the conductance when all the sodium channels are open. A shows the steady-state voltage dependence of the inactivation variable h (h∞ = h infinity); B shows the changes in the activation variables, m, and h during the upstroke of the action potential and the resultant change in gNa.

speed of depolarization increases; depolarization is regenerative. However, at the same time the h-gates begin to move toward their closed position (from 1 to 0, Fig. 17–8) although much more slowly than the opening of the m-gates.

Closing of the h-gates is in response to the positive shift in the membrane potential. Complete closure of the h-gates, while the m gates are still in the open position terminates the inward movement of Na^+ and the depolarization phase. The Na^+-channels are said to be *inactivated*. The original position of the Na channel gates is restored as the membrane potential is repolarized back to the resting potential. Until this occurs the cell is refractory to additional stimuli. Restoration of the original gate position is called *removal of inactivation*.

The properties of the Na^+ current are responsible for conduction of the action potential from one cell to the next. This will be described in detail in the chapter. The properties of the inward Na^+ current and conduction are also dependent on the level of the resting potential at which the upstroke arises because this level determines the amount of inactivation of the h "gates" (Fig. 17–8, A).

The plateau phase (phase 2). There is a plateau phase to the action potential where membrane potential remains in the positive range, even after Na^+ current has ceased to flow.

This results for three reasons:

1. There is another inward depolarizing current, the L-type Ca^{2+} current flowing at this time (Fig. 20-8).
2. The positive shift in membrane potential decreases resting K^+ conductance (anomalous rectifier property of gKl).
3. Activation of the repolarizing K^+ current is delayed.

Current carried by Ca^{2+} through L-type calcium channels is inward because E_Q, is quite positive. The Ca^{2+} current is much smaller than the Na^+ current, develops less rapidly and inactivates quite slowly.

This Ca^{2+} current provides the trigger for contraction.

Repolarization (phase 3). Repolarization is the net result of a number of currents. It is aided by the inactivation of the Na^+ and Ca^{2+} currents. However, the main cause is the eventual activation of an outward K^+ current which flows through K channels distinct from the channels that cause the resting potential and which are called the delayed rectifier channels (Fig. 17–10). These channels open slowly on depolarization, do not inactivate and close on repolarization.

Overall the changes in cations in- and outward movements are summarized on the Fig. 17–10. Resuming, the rise in action potential (phase 0) is caused by

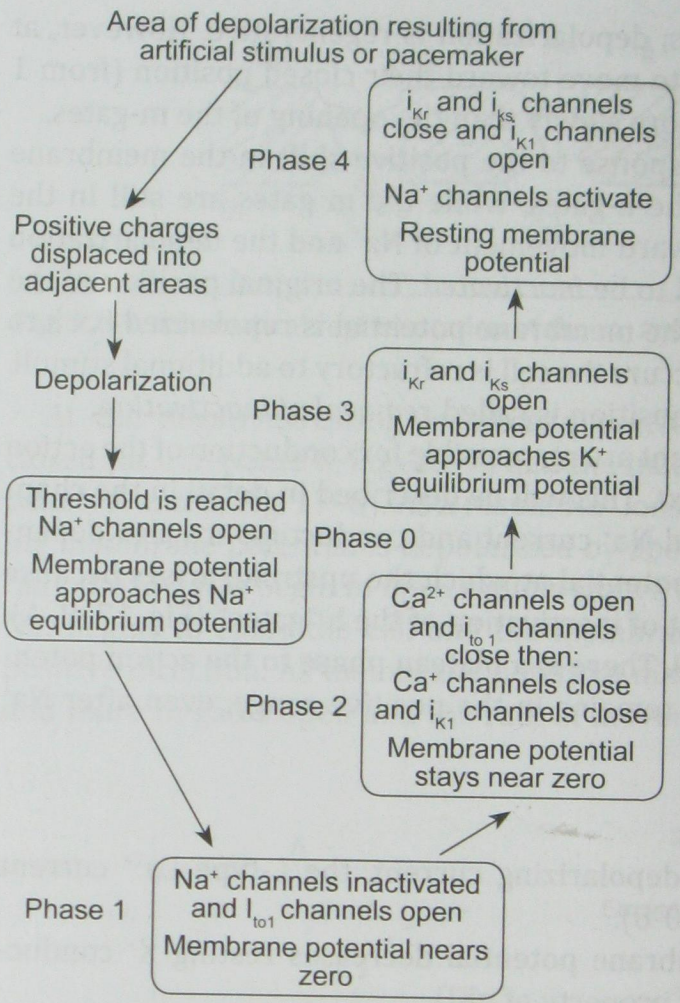


Fig. 17–9. Ionic and cellular events associated with the ventricular action potential (explanation see in the text).

rapidly increasing Na^+ current carried by voltage-gated Na^+ channels. Na^+ current falls rapidly because voltage-gated Na^+ channels are inactivated. K^+ current rises briefly because of opening of K^+ channels and then falls precipitously because iK1 channels are closed by depolarization. Ca^{2+} channels are opened by depolarization and are responsible, along with closed iK1 channels, for phase 2. K^+ current begins to increase because iKr and iKs channels are opened by depolarization, after a delay. Once repolarization occurs, Na^+ channels are activated. Reopened iK1 channels maintain phase 4.

Ion pumps. During the occurrence of the action potential there is movement of Na^+ and Ca^{2+} ions from outside to inside the cells and K^+ ions from inside to outside the cells. Over the course of many heart beats the cells would eventually fill up with Na^+ and Ca^{2+} and loose all their K^+ if there were not some physiological mechanisms to prevent this. Two mechanisms are the Na/K pump and the Na/Ca exchanger.

The Na/K pump in the sarcolemma continuously moves Na^+ out of the cell and K^+ into the cell, thereby enabling the cell to maintain a relatively constant intracellular and extracellular ionic environment. The pump moves 3 Na^+ ions out of the cell in exchange for 2 K^+ ions that it moves into the cell. Therefore, the pump is electrogenic; it generates an outward current. The movement of both ions is against both an electrical and concentration gradient and, there-

fore, requires the expenditure of energy. This energy comes from the splitting of ATP molecules that are part of the pump.

The Na/Ca exchanger in the cell membrane is the major mechanism that the cell has for removing Ca^{2+}. As the name implies, Ca^{2+} is moved out of the cell against an electrical and concentration gradient in exchange for Na^+. This movement of Na^+ is distinct from the Na^+ that moves into the cell during the upstroke of the action potential. Energy is required for the extrusion of Ca^{2+}.

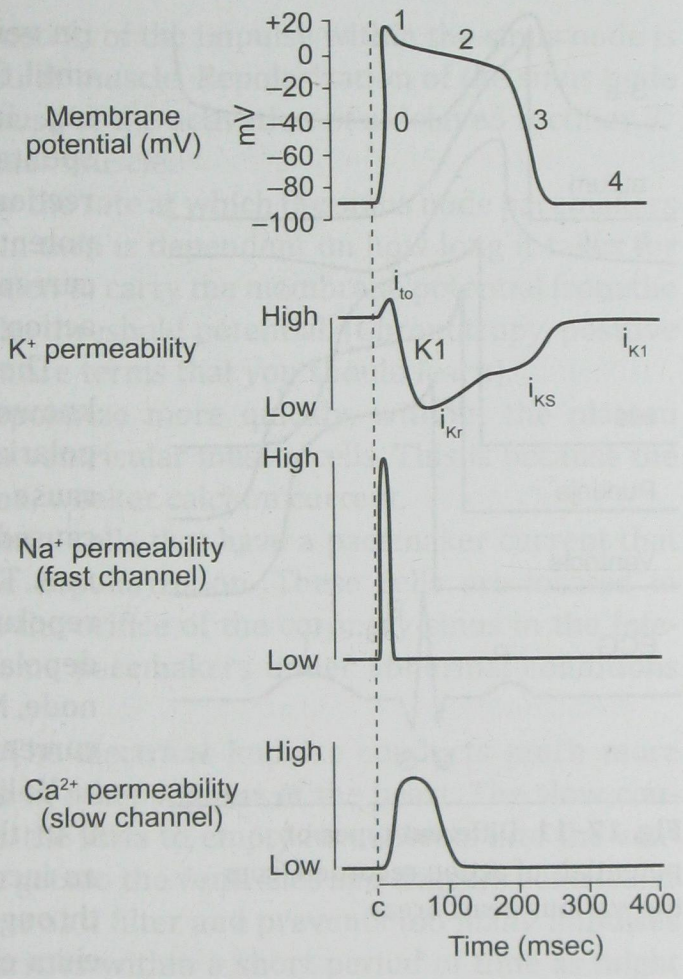


Fig. 17–10. Changes in cations permeabilities during a Purkinje fiber action potential.

3. Action potentials from other regions of the heart

Figures 17–4 and 17–11 show action potentials recorded from other regions of the heart beside the ventricular muscle. There are some special features of these action potentials that are important.

The sinus node. The sinus node is the normal pacemaker of the heart. Its ability to initiate the impulse is related to the special characteristics of its action potentials.

Sinus node cells have no true resting potential. The maximum level of negativity (the maximum diastolic potential, MDP) is only around –60 mV (Fig. 17–12). The membrane potential is further away from E_K because the cell membrane is permeable to other ions besides K^+, in particular there is a significant P_{Na}. Membrane potential does not remain at this negative level

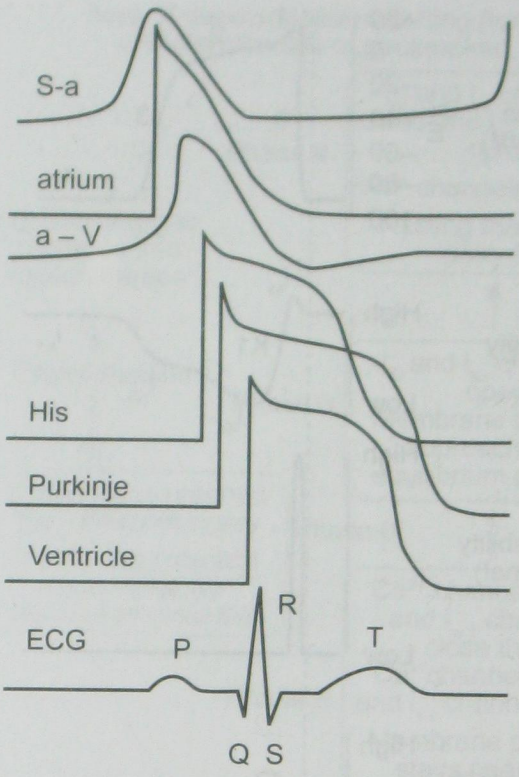


Fig. 17–11. Different types of potentials of action recorded from the various heart areas.

(in ventricular muscle it remains here until the cell is excited by a propagating impulse); membrane potential spontaneously moves in a positive direction until it reaches the threshold potential for a regenerative inward current (ENa), and the upstroke of the action potential occurs.

The spontaneous depolarization is known as "spontaneous diastolic depolarization", or "phase 4" and is the cause of sinus node automaticity. It is caused by the inward pacemaker current. This current slowly turns on upon repolarization and inactivates during depolarization to -60 mV. In the sinus node, Na^+ and Ca^{2+} carry the pacemaker current (Fig. 17–12).

Unlike in ventricular muscle, phase 0 in the sinus node is largely due to an increase in inward calcium current through activation of the L-type calcium channel. Because the inward calcium current is weaker than the fast sodium current, the depolarization phase of the sinus node action potential is much slower than the sodium dependent

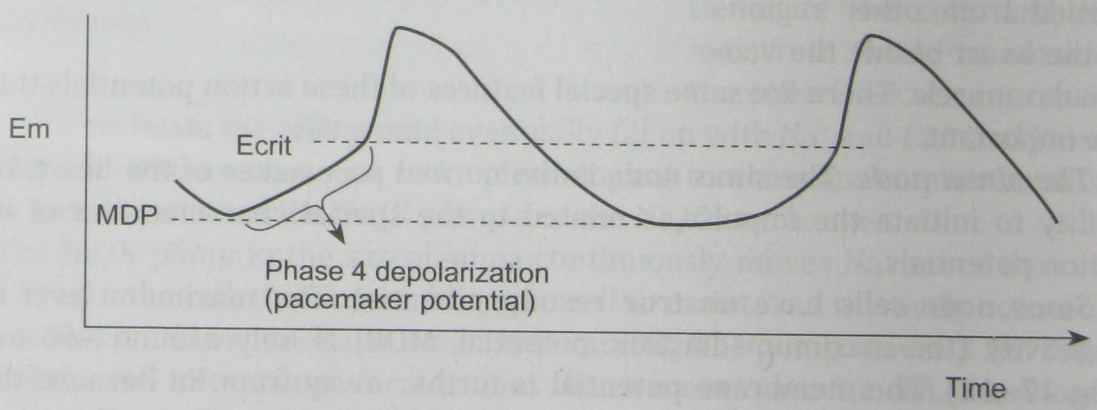


Fig. 17–12. Transmembrane potential of an automatic SA nodal fiber.

ventricular muscle cell and conduction of the impulse within the sinus node is also much slower than in ventricular muscle. Repolarization of the sinus node action potential than occurs because of the activation of a delayed rectifier K^+ current that is similar to ventricular muscle.

The heart rate is determined by the rate at which the sinus node pacemakers generate action potentials. This in turn is dependent on how long it takes for spontaneous diastolic depolarization to carry the membrane potential from the maximum diastolic potential to the threshold potential. (Chronotropy, positive and negative chronotropic effects are terms that you should learn).

Atrial muscle. Atrial cells repolarize more quickly, without the plateau phase (3) that is so prominent in ventricular muscle cells. This is because the atrial cells have a shorter duration, weaker calcium current.

There are also some special atrial cells that have a pacemaker current that can cause spontaneous diastolic depolarization. These cells are located in the crista terminalis and around the orifice of the coronary sinus in the inferior right atrium. They can become pacemakers under abnormal conditions (described later).

Atrioventricular (AV) node. The electrical impulse conducts much more slowly through the AV node than in other regions of the heart. The slow conduction allows sufficient time for the atria to empty its contents into the ventricles before the cardiac impulse gets to the ventricles and triggers ventricular contraction. The AV node also acts as a filter and prevents too many impulses from getting through to the ventricles within a short period of time as might occur during very rapid atrial rates. This prevents the ventricles from beating too rapidly which could diminish the cardiac output.

The membrane potential during diastole is low (around -60 mV) because of permeability to other cations (particularly Na^+ in addition to K^+) . There is a pacemaker current in some AV nodal cells but it is much weaker than in the sinus node and causes only a small drift in the positive direction before the cells are excited by a propagating impulse. The depolarization phase (upstroke) of the AV nodal action potential, like the sinus node, is mainly a result of an inward calcium current flowing through the L type calcium channel. Repolarization is a result of inactivation of this inward current and the delayed activation of an outward K^+ current. The slow conduction and the long effective refractory period of the AV node are a property of the Ca^{2+} current.

Purkinje fibers bundle of His and ventricular specialized conducting system. The function of the Purkinje system is to rapidly convey the electrical depolarization wave throughout the ventricles. The action potential is designed

for rapid conduction since it has a faster depolarization phase than working muscle as a consequence of a larger inward sodium current.

Purkinje fibers also have a pacemaker current. If, and under abnormal conditions can become the pacemaker of the heart.

3. THE NORMAL EXCITABILITY OF THE HEART

Excitability refers to a cell's ability to respond to a stimulus. It is measured in terms of the energy content of a stimulus that is needed to elicit a response. The need for a stronger stimulus indicates a reduction in excitability. Responses range from graded depolarizations up to fully propagated action potentials.

Excitability changes during cardiac activity and various terms are employed to describe these changes. We will use ventricular muscle as an example in explaining them, and consider the effects of depolarizing electrical stimuli of short (10 msec) duration (Fig. 17–13).

Absolute refractory period. During phase 4, membrane potential, threshold potential, and excitability are constant. After the generation of the action potential upstroke, excitability drops to zero because almost all of the Na^+ channels have been inactivated. This state is called *absolute refractoriness* because a response cannot be elicited no matter how strong a stimulus is applied (Fig. 17–13).

Relative refractory period. The *absolute refractory period* persist throughout the plateau until phase 3 repolarization has brought membrane potential to –55 to –60 mV. At this voltage some Na^+ channels have recovered from inactivation. Now a very strong stimulus can elicit a response. This signals the end of the absolute refractory period and the beginning of the *relative refractory period* (Fig. 17–13).

Supernormal period. As the cell continues to repolarize most Na^+ channels recover from inactivation and the stimulus requirement decreases. When membrane potential reaches –80 mV or thereabouts, the stimulus needed to excite is *smaller* than during phase 4 because the Na^+ channels have all recovered but the membrane potential is closer to threshold than it is during phase 4. When this occurs the cell is no longer in the relative refractory period but in the supernormal period. Propagated responses during this interval still have slower upstrokes and shorter durations than normal.

These changes in excitability reflect, in large part the inactivation and recovery from inactivation of Na channels.

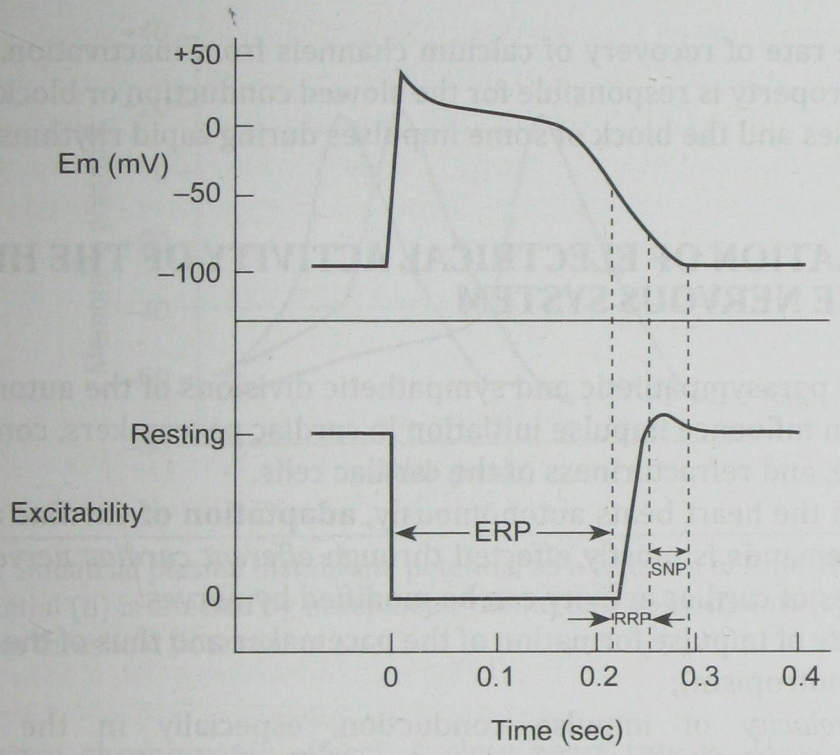


Fig. 17–13. Heart excitability cycle. ERP — effective refractory period RRP — relative refractory period, SNP — supernormal period

Responsiveness. What probably is of greater functional significance than small changes in electrical excitability are the changes in the ability of heart fibers to generate a response and in the characteristics of that response. This can be considered under the heading of *altered responsiveness.* Early during phase 3 strong stimuli elicit graded, non-propagated responses. When the membrane has repolarized more fully the responses propagate, although at reduced speed. The interval during which a *propagated response* cannot be generated is called the *effective refractory period.* Propagated responses elicited during phase 3 show slow upstrokes and reduced durations. When the resting potential has been attained, the response is fully normal; this signals the end of the *full recovery time.*

In most cardiac fibers under normal conditions transmembrane potential governs the time-course of recovery of excitability and responsiveness. The exceptions are the SA and AV node in which the excitatory current is a calcium current. In these cells recovery of responsiveness and excitability is not simultaneous with full repolarization but rather delayed after repolarization. This

reflects the rate of recovery of calcium channels from inactivation. In the AV node this property is responsible for the slowed conduction or block of premature impulses and the block of some impulses during rapid rhythms.

4. REGULATION OF ELECTRICAL ACTIVITY OF THE HEART BY THE NERVOUS SYSTEM

Both the parasympathetic and sympathetic divisions of the autonomic nervous system influence impulse initiation in cardiac pacemakers, conduction of the impulse, and refractoriness of the cardiac cells.

Although the heart beats autonomously, **adaptation of cardiac activity** to changing demands is mostly effected through *efferent cardiac nerves*. The following types of cardiac activity can be modified by nerves:

- *the rate* of impulse formation of the pacemaker and thus of the heart beat (chronotropism);
- *the velocity* of impulse conduction, especially in the AV node (dromotropism);
- the force of myocardial contraction at a given distension, i.e., the heart's *contractility* (inotropism);
- *the excitability* of the heart in the sense of changing its excitability threshold (bathmotropism);
- *the tone of cardiac muscle* during the heart cycle (tionotropism).

Parasympathetic nervous system. The fundamental mechanism through which the parasympathetics regulate cardiac electrical activity is by the action of acetylcholine (ACh) on membrane channels in the sarcolemma. ACh slows the heart rate by reducing the rate of spontaneous depolarization of pacemaker cells (Fig. 17–14), increasing the time required to reach threshold. Slowed heart rate is called *bradycardia*, or when the heart rate is below 60 beats/min. Acetylcholine increases the K^+ permeability (K^+ conductance) of the sarcolemma by activating special K-Ach channels. It also decreases the L-type Ca^{2+} conductance. Both actions hold the pacemaker potential. Both actions hold the pacemaker potentials closer to the K^+ equilibrium potential.

a. Sinus node.

The predominant effect of the parasympathetic nervous system on sinus node cells is to slow the rate of impulse initiation and, thereby, slow the heart

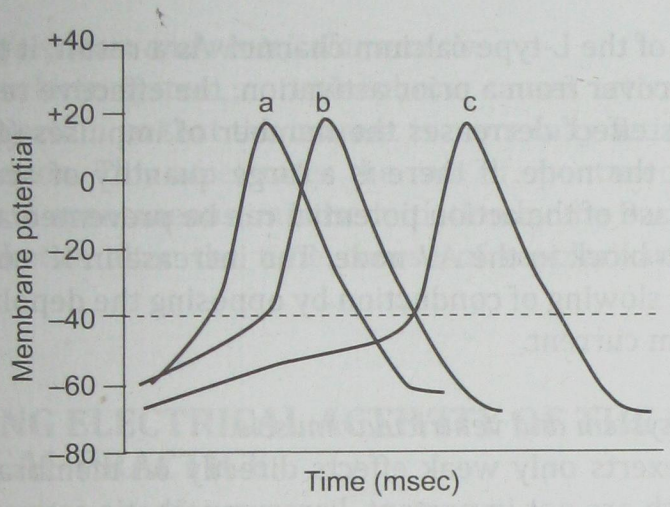


Fig. 17–14. Sinoatrial plasma membrane potential as a function of time. Normal pacemaker potential (b) is affected by norepinephrine (a) and acetylcholine (c). The dashed line indicates threshold potential.

rate (negative chronotropic effect). A slow heart rate is known as a *sinus bradycardia*. By increasing K^+ conductance, the maximum diastolic potential is shifted to a more negative level and as a result, it takes longer for spontaneous diastolic depolarization to bring membrane potential to threshold. This slows the rate. In addition, the increase in outward K^+ opposes the inward pacemaker current, reducing the rate at which the pacemaker current depolarizes the cell. The decrease in calcium conductance (T-type calcium channel) caused by acetylcholine also may contribute to the reduction in rate of spontaneous diastolic depolarization.

b. Atria.

An important effect is to accelerate the time course of repolarization. By accelerating repolarization, the refractory periods of the atrial cells are significantly shortened, meaning that the cells can be excited more frequently.

c. Atrioventricular node.

Acetylcholine slows conduction, and may even cause conduction to fail (conduction block) in the AV node (negative dromotropic effect). Acetylcholine decreases the inward L-type calcium current and, thereby, slows the rate of depolarization and decreases the amplitude of the action potential. This in turn slows conduction through the AV node. Acetylcholine also delays the removal

from inactivation of the L-type calcium channel. As a result, it takes longer for the AV node to recover from a prior activation; the effective refractory period is prolonged. This effect decreases the number of impulses (frequency) that can pass through the node. If there is a large quantity of acetylcholine, the depolarization phase of the action potential can be prevented altogether causing the impulse to block in the AV node. The increase in K^+ conductance also contributes to the slowing of conduction by opposing the depolarizing effect of the inward calcium current.

d. His-Purkinje system and ventricular muscle.

Acetylcholine exerts only weak effects directly on membrane channels of these tissues which are not important. Parasympathetic activation can inhibit the effects of the sympathetics (see below).

Sympathetic nervous system. Norepinephrine causes an increase in the slope of the pacemaker potential so that the threshold is reached more rapidly and the heart rate increases. Increased heart rate is called *tachycardia*, or when the heart rate is above 100 beats/min. Norepinephrine increases the slope of the pacemaker potential by opening channels carrying Na^+ and Ca^{2+} and closing K^+ channels. Both effects result in faster movement of the pacemaker potential toward the Na^+ and Ca^{2+} equilibrium potentials. The effects of the sympathetic nervous system on electrical activity of the heart are brought about primarily by the effects of norepinephrine released from sympathetic nerve terminals on β_1 receptors.

a. Sinus node.

The sympathetic nervous system (release of norepinephrine) accelerates the heart rate (positive chronotropic effect) by increasing the rate of impulse initiation by sinus node pacemaker cells. A fast heart rate is known as a *sinus tachycardia*. This results primarily from an increase in the rate of spontaneous depolarization. The increase occurs because the current is activated at more positive potentials and the inward current earned by calcium is increased during the phase of spontaneous diastolic depolarization.

b. Atrioventricular node.

Norepinephrine increases the L-type calcium current to increase the rate and amplitude of depolarization of AV nodal action potentials. This effect speeds conduction of the cardiac impulse through the AV node (*positive dromotropic effect*). The effective refractory period is also decreased. Norepinephrine also enhances the pacemaker current in AV nodal cells.

c. His-Purkinje system and ventricular muscle.

In the His-Purkinje system, norepinephrine causes activation to be shifted in a positive direction so that the current is more fully activated immediately after repolarization. This accelerates the rate of spontaneous depolarization and may result in spontaneous impulse initiation by the Purkinje cells.

The sympathetics alter the time course of repolarization of ventricular muscle.

5. MEASURING ELECTRICAL ACTIVITY OF THE HEART IN CLINICAL PRACTICE

Abnormalities in initiation and conduction of the electrical impulse in the heart are caused by cardiac diseases and by the effects of drugs that act on the heart. To detect these abnormalities it is necessary for the clinician to be able to measure the heart's electrical activity.

Cardiac Electrograms. Records of electrical activity obtained through leads in contact with the heart are called *cardiac electrograms*. The lead systems can be either *unipolar* (measuring the potential difference between a selected point and either ground or a reference potential) or *bipolar* (measuring the difference between two points at each of which potential varies with tune).

When the transmembrane potential is uniform throughout a segment of the atria or ventricles, no current flows. When there is a difference in transmembrane potential between adjacent segments of the membrane, current flows. This current flows from the depolarized membrane along the internal path (myoplasm and gap junctional channels) out across the membrane and back in the opposite direction in the extracellular path from the not-yet-depolarized membrane to the depolarized membrane. It is the extracellular current that generates the potential differences recorded as electrograms.

a. Unipolar electrograms

The electrograms easiest to derive are the unipolar records. For our model we will employ a bundle of ventricular muscle fibers with uniform properties (Fig. 17–15).

The muscle is stimulated at one end (site S). The unipolar recording lead is located at one of three positions: near the site of stimulation (Site A), at the midpoint of the muscle (Site B), or at the end distal from the stimulus site (Site C). The recording is set so that a positive shift in extracellular potential is

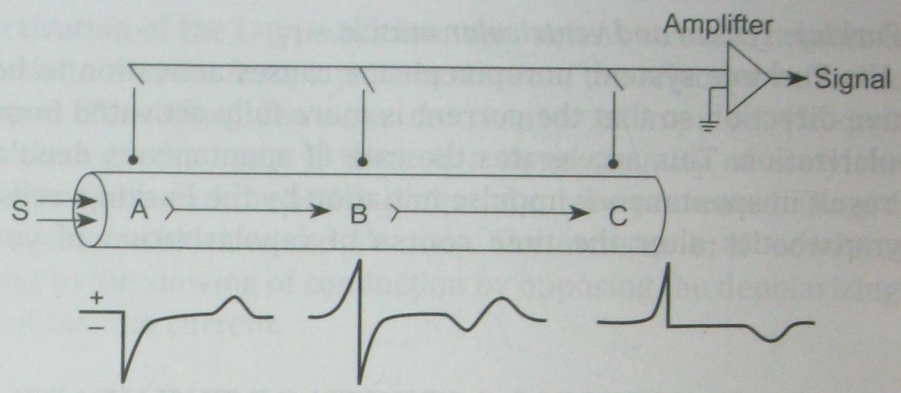


Fig. 17–15. Schematical illustration of strip of ventricular muscle, stimulated at one end (S). Unipolar leads are recorded from each of three sites, A, B and C. The resulting electrograms are shown for each site. The arrows show the direction of propagation of the impulse.

registered as an upward deflection. When the impulse is initiated at *Site A* the local electrogram shows a prompt negative deflection and then a slower return to the isoelectric line.

As the impulse approaches *Site B*, the trace shifts in a positive direction. When the wavefront is immediately under the electrode, the trace rapidly changes from a maximum positive to a maximum negative deflection.

Then as the impulse spreads towards *Site C* the trace returns to the isoelectric line. At *Site C* the record is the inverse of that at *Site A*. The trace goes slowly positive and, then when the impulse reaches the very end of the muscle, abruptly returns to the isopotential line.

In terms of the phenomenology of these records one can derive some helpful rules. If the impulse is propagating *towards* an electrode, the electrode senses a positive change in potential; if the impulse is propagating away from an electrode, that electrode senses a negative change in potential. One can explain this in terms of the change in intensity and direction of transmembrane current. In electrocardiography it is usual to explain it by representing the wavefront of activity during depolarization as a dipole or line charge with positive charges in front and negative charges behind.

In ventricular muscle, transmembrane potential remains near zero during the plateau and the plateau has a long duration. For these reasons, after the impulse has propagated from one end of the muscle to the other, the cells are again at a uniform (all depolarized) potential and no current flows.

When repolarization starts at Site A, it again causes a potential difference that can be detected by the unipolar leads. Because the rate of repolarization is less than the rate of depolarization, the electrogram deflections resulting from repolarization are smaller and of greater duration than those during depolarization. These deflections are also of opposite polarity because the membrane potential is changing from a depolarized to a repolarized state. The spread of repolarization can be represented by a dipole or line charge with negative charges in front and positive charges behind.

b. Bipolar electrograms

When recordings are obtained through bipolar leads the record shows the instantaneous difference in potential between the two recording sites. The easiest way to represent it is to construct the unipolar records for each recording site and then add them as the amplifier does. If conduction is slow enough and the electrode separation great enough, a bipolar lead will in fact record two typical "unipolar" traces.

Remember that a positive potential at one amplifier input gives an upward deflection and at the other input a downward deflection.

The bipolar electrogram thus typically is triphasic, showing a positive, then a negative and finally a positive deflection.

The Electrocardiogram (ECG). The electrical activity that propagates throughout the heart causes fluctuations in electrical potential in the body fluids which contain electrolytes (known as a volume conductor). These fluctuations can be detected at the body surface by placing electrodes on the skin. The potentials at the body surface are very small but can be electronically amplified.

The ECG is a recording of *potential differences* (in mV) that are generated by the excitation within the heart. It can provide information about the position of the heart and its rate and rhythm as well as the origin and spread of the action potential, but *not* about the contraction and pumping action of the heart.

Although, the electrical activity could be detected anywhere at the body surface, there are standard positions where the recording electrodes are placed so that electrical activity can be compared to a set of normal values. For this standard, the heart is considered to be in the center of an equilateral triangle formed by the left and right shoulders and the pubic region. This triangle, is oriented in the frontal plane of the body and is called Einthoven's triangle (Fig. 17–16) named after the Dutch electrocardiographer who developed the recording techniques. Theoretically, the recording electrodes should

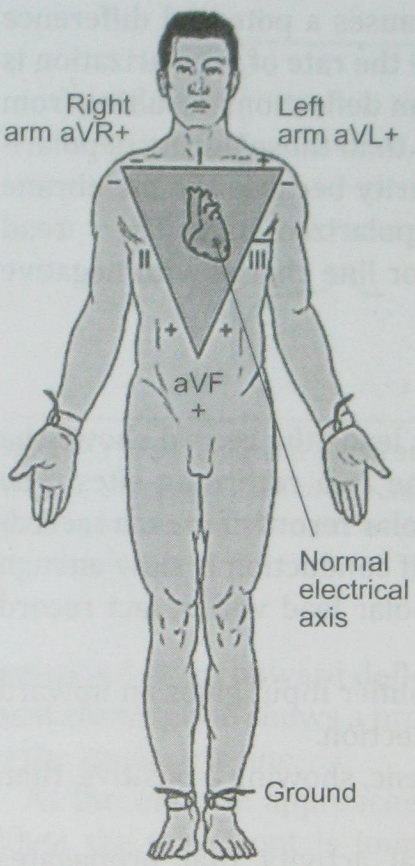


Fig. 17–16. Einthoven triangle.

be placed at each of the three apices of the triangle.

To understand the ECG, it is necessary to understand the behavior of electrical potentials in a three-dimensional conductor of electricity. Consider what happens when wires are run from the positive and negative terminals of a battery into a dish containing salt solution. Positively charged ions flow toward the negative wire (negative pole) and negatively charged ions simultaneously flow in the opposite direction toward the positive wire (positive pole).

The combination of two poles that are equal in magnitude and opposite in charge and located close to one another, is called *a dipole.* The flow of ions (current) is greatest in the region between the two poles, but some current flows at every point surrounding the dipole, reflecting the fact that voltage differences exist everywhere in the solution.

We can use this analysis of a dipole in a volume conductor to rationalize the waveforms of the ECG. Of course, the actual case of the heart located in the chest is not as simple as the dipole in the tub of salt solution for two main reasons.

First, excitation of the heart does not create one dipole; instead, there are many simultaneous dipoles. We will focus with the net dipole emerging as an average of all the individual dipoles. Second, the body is not a homogeneous volume conductor. The most significant problem is that the lungs are full of air, not salt solution. Despite these problems, the model is useful in an initial understanding of the generation of the ECG.

At rest, myocardial cells have a negative charge inside and a positive charge outside the cell membrane. As cells depolarize, the depolarized cells become negative on the outside, whereas the cells in the region ahead of the depolarized cells remain positive on the outside (Fig. 17–17). When the entire myocardium is depolarized, no voltage differences exist between any regions of myocardium because all cells are negative on the outside. When the cells in a given

region depolarize during normal excitation, that portion of the heart generates a dipole. The depolarized portion constitutes the negative side, and the yet to be depolarized portion constitutes the positive side of the dipole. The tub of salt solution is analogous to the rest of the body in that the heart is a dipole in a volume conductor. With electrodes located at various points around the volume conductor (i.e. the body) the voltage resulting from the dipole generated by the electrical activity of the heart can be measured.

To record a ***Lead I*** electrocardiogram, measurements are made between the right and left arms (Fig. 17–16). When the potential at the left arm exceeds the potential at the right arm, the deflection on the recording device will be upward (positive) from the isoelectric line.

Lead II records the potential difference between the right arm and left leg and there is an upward deflection when the potential at the left leg is greater than at the right arm.

Lead III records the potential difference between the left arm and the left leg and there is an upward deflection when the potential at the left leg is greater than at the arm (Fig. 17–16).

During conduction of the electrical excitatory wave over the heart, the movement of the dipoles through the heart can be represented by a series of instantaneous vectors (repre-

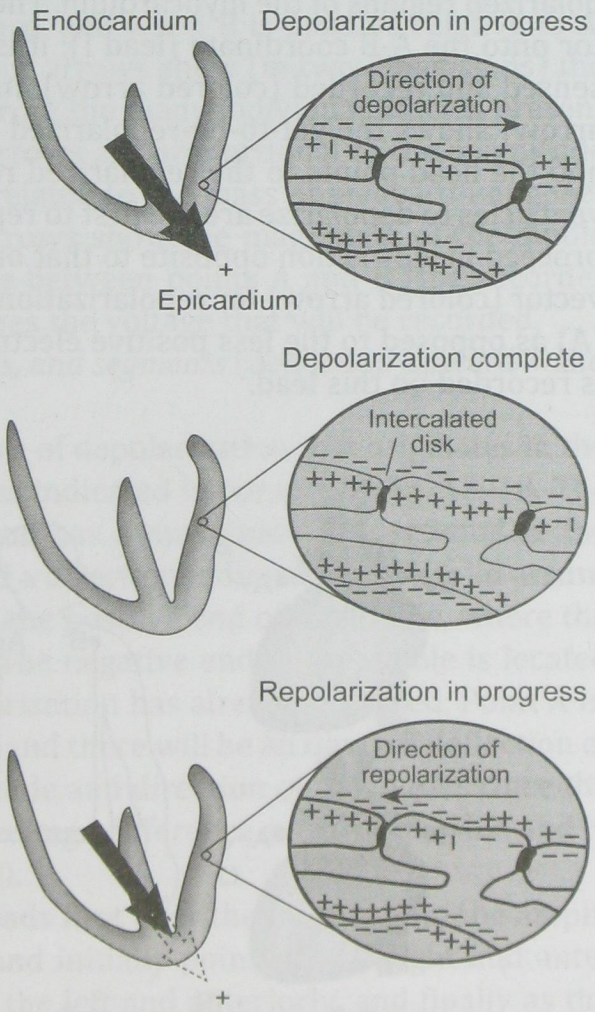


Fig. 17–17. Cardiac dipoles.
Partially depolarized or repolarized myocardium creates a dipole. Arrows show the direction of depolarization (or repolarization). Dipoles are present only when myocardium is undergoing depolarization or repolarization.

sented by arrows) that have both magnitude and direction and which are dependent on where the wavefront is at any moment in time. The point of the arrow represents the positive end of the dipole (Fig. 17–18).

The black arrows are vectors that represent the magnitude and direction of a major dipole. The magnitude is proportional to the mass of myocardium involved. The direction is determined by the orientation of depolarized and polarized regions of the myocardium. The vertical dashed lines project the vector onto the A-B coordinate (lead I); it is this component of the vector that is sensed and recorded (colored arrow). In panel 5, the tail of the vector (black arrow) shows the yet-to-be-repolarized region of the myocardium (negative) and the head points to the repolarized region (positive). The last areas of the ventricles to depolarize are the first to repolarize, i.e., repolarization appears to proceed in a direction opposite to that of depolarization. The projection of the vector (colored arrow) for repolarization points to the more positive electrode (A) as opposed to the less positive electrode (B), and so an upward deflection is recorded on this lead.

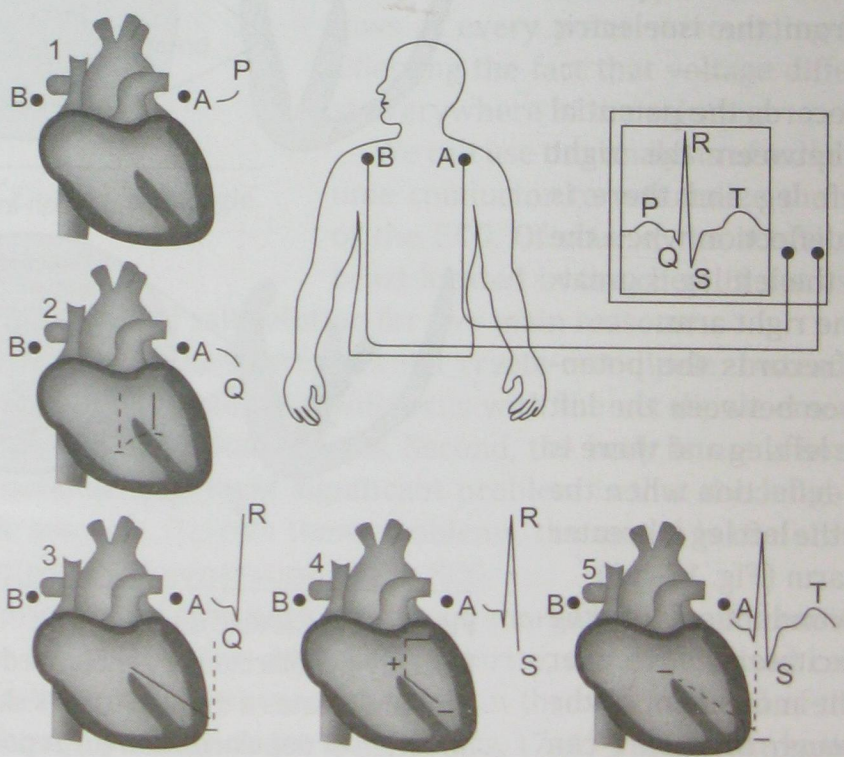


Fig. 17–18. The sequence of major dipoles giving rise to ECG waveforms.

The origin of each vector arrow is placed at a central point. The direction and magnitude of the projection of the vector on any of the electrocardiographic leads determines the electrocardiographic voltage in that lead. By convention, any dipole that is moving toward the positive recording pole creates an upright or positive deflection on the electrocardiographic tracing.

An electrocardiographic recorder (a voltmeter) is connected between points A and B (lead I, see Fig. 17–18). By convention, when point A is positive relative to point B, the ECG is deflected upward, and when B is positive relative to A, downward deflection results. The black arrows show (in two dimensions) the direction of the net dipole resulting from the many individual dipoles present at any one time. The lengths of the arrows are proportional to the magnitude (voltage) of the net dipole, which is related to the mass of myocardium generating the net dipole. The colored arrows show the magnitude of the dipole component that is parallel to the line between points A and B (the recorder electrodes); this component determines the voltage that will be recorded.

An **ECG tracing** has *waves, intervals, and segments* (deflection upward + and downward –).

Atrial excitation results from a wave of depolarization that originates in the SA node and spreads over the atria, as indicated in panel 1 of Fig. 17–18. The net dipole generated by this excitation has a magnitude proportional to the mass of the atrial muscle involved and a direction indicated by the solid arrow. The head of the arrow points toward the positive end of the dipole, where the atrial muscle is not yet depolarized. The negative end of the dipole is located at the tail of the arrow, where depolarization has already occurred. Point A is, therefore, positive relative to point B, and there will be an upward deflection of the ECG as determined by the magnitude and direction of the dipole. Once the atria are completely depolarized, no voltage difference exists between A and B, and the voltage recording returns to 0.

In the atria, as depolarization spreads first over the right atrium the amplitude of the vector increases in size and initially points to the right and anteriorly. The vector then shifts toward the left and anteriorly, and finally as the left atrium is depolarized, it becomes smaller and is directed to the left and posteriorly. The shifting vectors during atrial depolarization form the *P wave* on the electrocardiogram (Figs. 17–18, 17–19). Atrial repolarization is not seen on the ECG because it is masked by ventricular depolarization.

During passage of the impulse through the AV node and His bundle, electrical activity is not detected on the body surface because of the small size of these structures and the concomitant small size of the extracellular potentials.

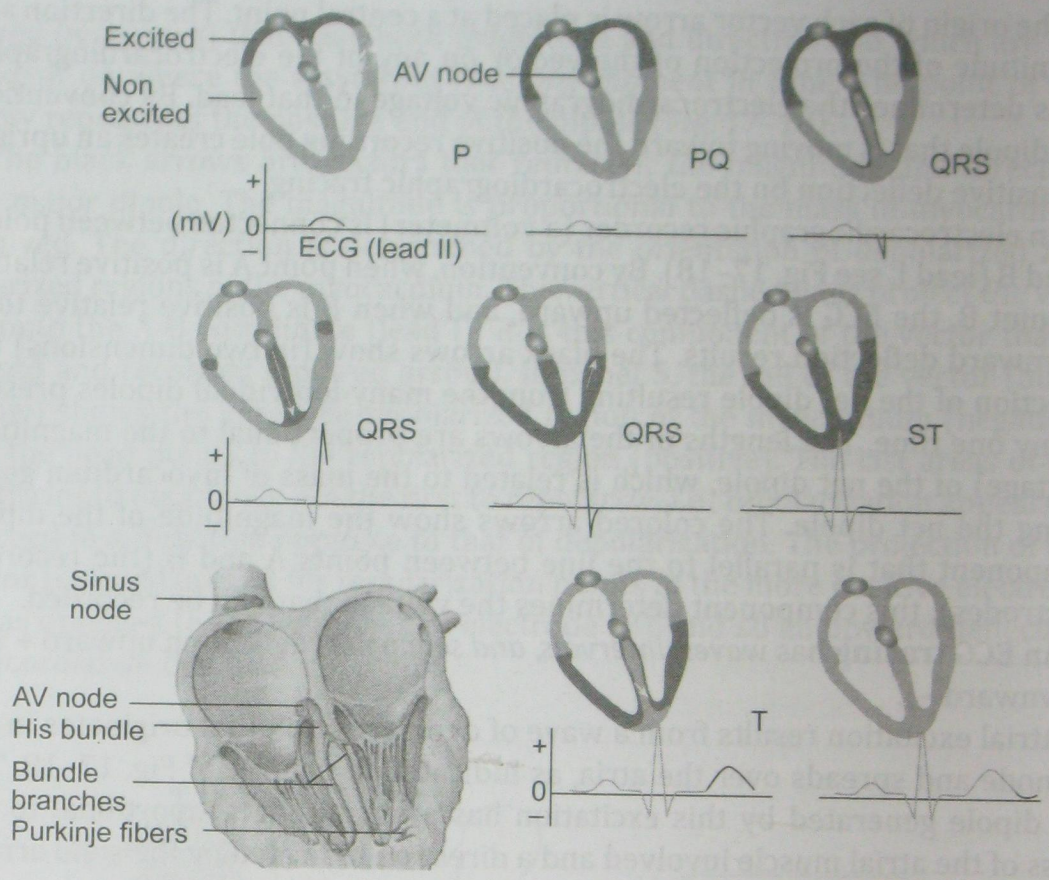


Fig. 17–19. Spread of excitation in the heart.

The dipoles created by depolarization of these structures are too small to produce a deflection on the ECG.

The isoelectric period between the end of the P wave and the beginning of the QRS complex, which signals ventricular depolarization is called the PQ interval. The P wave plus the PR segment is the PQ interval. The duration of the PQ interval is usually taken as an index of AV conduction time. Thus, there is an isoelectric segment after atrial depolarization, before ventricular activity can be detected (Fig. 17–19).

While the impulse is spreading through the AV node and His-Purkinje system the atria have been activated and the ventricles are still at rest; the ECG thus shows an isoelectric interval. The depolarization of the ventricular fibers begins in the left interventricular septum and spreads toward the right septum. This results in a small vector which projects on the standard limb leads

as a downward deflection marked *Q wave*. As a result of the rapid conduction through the bundle branches and the Purkinje fibers, the various parts of the endocardial surface of the ventricles depolarize almost simultaneously.

Depolarization then proceeds through the thickness of the ventricular myocardium from the endocardial surface to the epicardial surface. The summation of these activities gives rise to a large net vector pointing from the base (the atrioventricular junction) to the apex. This vector is normally tilted to the left because the thicker left ventricular wall contains more fibers than the thinner right ventricle. The projection of this vector on each of the three standard limb leads gives rise to the large upward *R wave*. This phase of ventricular activation includes most of the ventricular fibers except those at the base, which does not contain the fast-conducting Purkinje fibers.

Therefore, the last phase of ventricular depolarization proceeds toward the base. This leads to the downward deflection *S* on standard limb leads. The *QRS complex* represents the normal sequence of ventricular activation. When all parts of the ventricles are depolarized, the ECG again shows an isoelectric tracing.

The endocardial surface of the ventricular myocardium, which depolarizes first, has action potentials with longer durations than those at the epicardial surface. Therefore, during the early phase of ventricular recovery, when the epicardial region has already repolarized, the endocardial region is still depolarized; this results in the same orientation of surface potential as seen during ventricular excitation (R wave) because during repolarization (see above) the polarity of the vector is reversed. Hence, the ECG recording during the spread of ventricular repolarization, i.e., the T wave, is normally upright. The T wave is smaller in magnitude and longer in duration than the QRS, because repolarization occurs less synchronously and less rapidly than depolarization .

ECG leads. For appropriate ECG registration one could use the following leads that might be classified as following:

The *Einthoven* (or *standard limb*) *leads I, II, and III* are bipolar and lie in the frontal plane (Fig. 17–20, part 1).

For the *unipolar Goldberger* (*limb*) *leads*, aVL, aVR, and aVF (a = augmented) one limb electrode (e.g., the left arm in aVL) is connected to the junction of the two other limb electrodes (Fig. 17–20, part 3). These leads, too, lie in the frontal plane. (Fig. 17–21).

The *unipolar precordial leads* V_1–V_6 (*Wilson leads;* Fig. 17–20, part 4) lie approximately in the horizontal plane (Fig. 17–22). They mainly record those vectors that are directed posteriorly. As the mean QRS vector (see below) mainly points downward to the left and posteriorly, the thoracic cage is divided

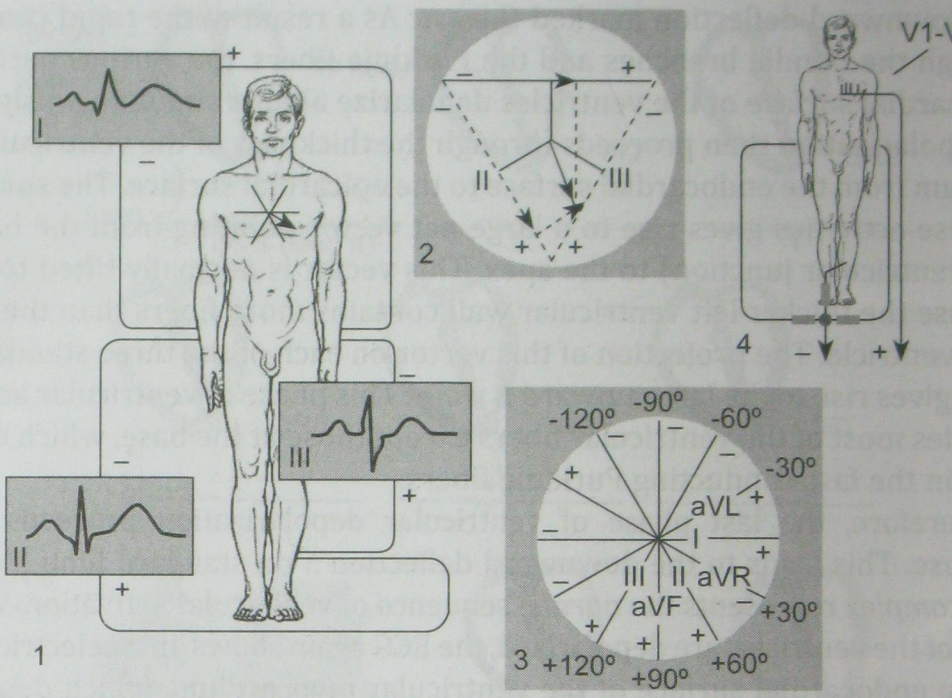


Fig. 17–20. Bipolar (Standard — I, II and III) and Unipolar (Goldberger — aVR, aVL and aVF; precordial or chest) leads.

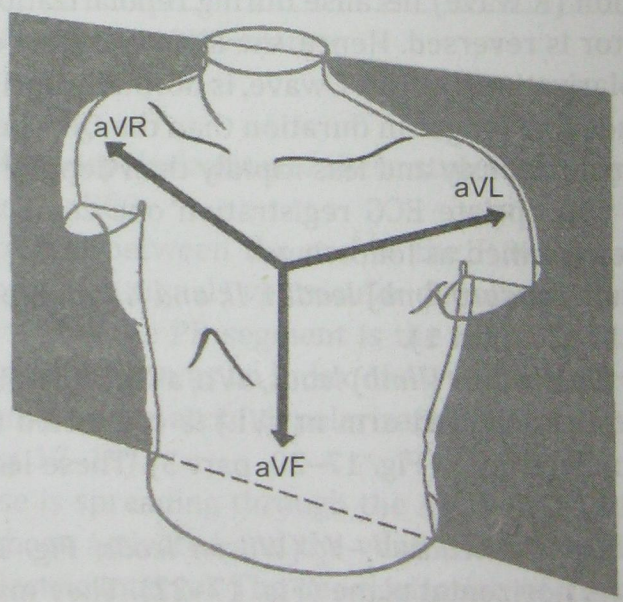


Fig. 17–21. The scheme of augmented unipolar leads frontal localization.

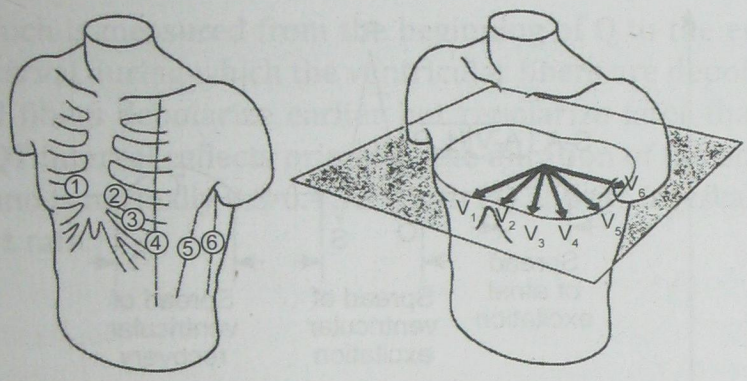


Fig. 17–22. Chest (Wilson) leads projections on the horizontal axis.

into a positive and a negative half by a plane which is vertical to this vector. As a result, the QRS vector is usually negative in V_1–V_3, positive in V_5–V_6.

Usual normal standard 12-lead ECG includes six limb leads and six chest leads. The ECG is calibrated so that two dark horizontal lines (1 cm) represent 1 mV, and five dark vertical lines represent 1 second. This means that one light vertical line represents 0.04 sec.

The normal ECG in relation to activation and recovery for various cardiac fibers. As indicated above, the ECG recording reflects the spread of electrical activity through the heart. The P wave represents the spread of excitation through the atrial fibers, the QRS complex denotes the spread of excitation in the ventricular fibers, and the T wave indicates the spread of recovery in the ventricular fibers (Fig. 17–23).

Recovery in the atrial fibers is not identifiable on the normal ECG tracing because (a) the associated surface signal is very small and (b) its timing often coincides with the large QRS complex and hence is masked by the latter.

The spread of excitation through the specialized fibers, due to their small mass, does not lead to any signal in the standard ECG. The timing of activation of the specialized fibers, however, can be inferred from the ECG tracing (Fig. 17–19).

Thus, activation of the SA node occurs prior to the beginning of the P wave. The AV node is activated shortly after the peak of the P wave and propagation in it extends through most of the isoelectric period after the P wave. The activation of the bundle of His and the Purkinje fibers, which are rapidly conducting, begins toward the end of this isoelectric period and ends in the beginning portion of the QRS complex.

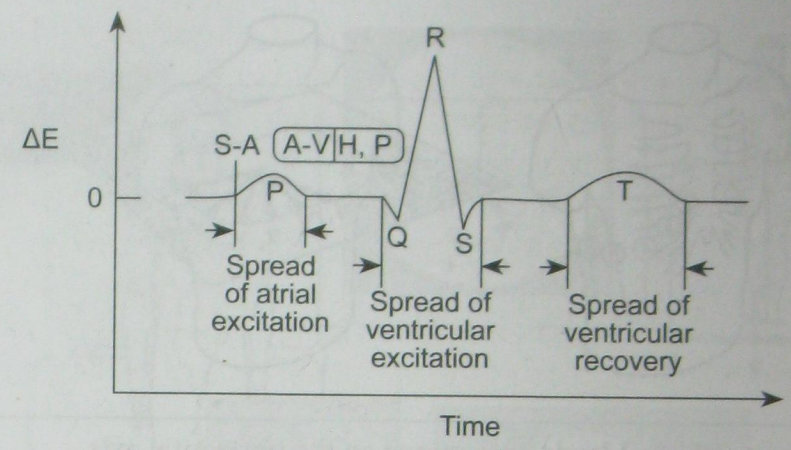


Fig. 17–23. ECG part pertaining to different cardiac fibers refractoriness.

Normal parameters of the ECG. The duration of the P wave is 0.1 sec under normal conditions, amplitude less 0.25 mV (Fig. 17–24). The PQ interval represents the time required for impulses to be conducted from the atrial margin of the SA node to the ventricular fibers. The average normal duration of the PQ interval is 0.16 sec (range 0.12–0.20 sec). A lengthening of the PQ interval beyond 0.20 sec usually signifies an abnormal conduction delay in the AV node.

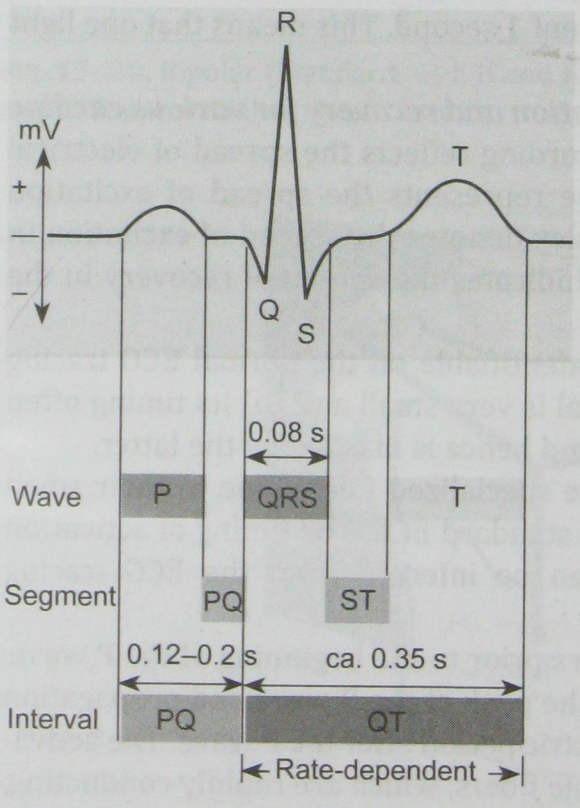


Fig. 17–24. Normal ECG values

The duration of the QRS complex, which signifies the spread of excitation through the ventricular fibers, is normally 0.06–0.09 sec. Abnormalities in conduction through the ventricular specialized fibers will effect the synchrony and time of activation of the ventricular fibers.

As a result, in the presence of such an abnormality the QRS duration is lengthened and the QRS complex may become notched or show other bizarre shapes. The QT

interval, which is measured from the beginning of Q to the end of T, signifies the total interval during which the ventricular fibers are depolarized. Since the endocardial fibers depolarize earlier but repolarize later than the epicardial fibers, the QT interval reflects primarily the duration of the endocardial action potentials and is normally 0.3-0.4 sec. Both PQ and QT intervals vary as a function of heart rate.

Chapter 18. Neurohumoral regulation of cardiac activity

In humans and other mammals, multiple cardiovascular regulatory mechanisms have evolved. These mechanisms increase the blood supply to active tissues and increase or decrease heat loss from the body by redistributing the blood. In the face of challenges such as hemorrhage, they maintain the blood flow to the heart and brain. When the challenge faced is severe, flow to these vital organs is maintained at the expense of the circulation to the rest of the body.

Being functioning together, both cardiac system and the circulatory one undergone to the related regulatory mechanisms in different stress- , emotion- and physical exertion — induced situations. It's very difficult to separate the cardiac regulatory mechanisms from the vascular and vice versa. It's important to understand that these adaptative and/or regulative mechanisms are similar and serve to maintain the oxygen delivery to the organs and tissues for their adequate functioning.

That's why this chapter will introduce you with the only cardiac-related regulatory mechanisms which are primary connected with myocardium functioning. Thus, we are artificially divide this regulatory aspect of the cardiovascular system physiology into two parts. The second, closely related to the circulatory system you may see later.

Heart rate (HR), blood pressure (BP), blood flow to individual organs, velocity of the blood, all remain within a narrow limit. For example, at rest, the HR and BP are, usually, between 60-72 per min. and 110-140 (systolic) and 60-90 (diastolic) mm Hg, respectively. There are many situations in our daily life, which threaten to destabilize this equilibrium Or, there are many conditions, where some cardiovascular readjustment is necessary. The forces of destabilization are opposed by various reflexes and other control mechanisms. These also are required to achieve the readjustment mentioned higher.

The healthy cardiovascular system is capable of providing appropriate blood flow to each of the organs and tissues of the body under a wide range of conditions. This is done by BP maintaining within normal limits, adjusting the heart output to the appropriate level and adjusting the resistance to blood flow in specific organs and tissues to meet special functional needs.

The heart regulatory mechanism includes *intracardiac* mechanism and *extracardiac* mechanism which also could be divided into some special subtypes (Fig. 18–1).

1. The intracardiac mechanism known also as the **"Law of the Heart"** or **"Frank-Starling Law of the Heart"** represents also the mechanism of cardiac activity self-regulation.

The heart *autonomously* responds to changes in ventricular volume load or aortic pressure load by adjusting the stroke volume in accordance with the myocardial *preload* (resting tension). The Frank-Starling Law of the Heart also functions to maintain an equal stroke volume in both ventricles to prevent congestion in the pulmonary or systemic circulation. This mechanism of self-regulation works to adjust the ventricular end-diastolic fiber length and stroke volume.

End-diastolic fiber length is determined by EDV, which is dependent on end-diastolic pressure. End-diastolic pressure is the force that expands the ventricle to a particular volume. For the intact heart, *preload* can be defined as end-diastolic pressure. For a given ventricular compliance (change in volume caused by a given change in pressure), a higher end-diastolic pressure (preload) increases both diastolic volume and fiber length. The end-diastolic pressure depends on the degree of ventricular filling during ventricular diastole, which is influenced largely by atrial pressure.

The curve expressing the relationship between myocardium sarcomere length (that in turn represent the marker of the ventricular filling) and ventricular contractile performance (the relative tension) is called a *Starling curve* (Fig. 18–2, left graph). The passive tension line on this graph means ventricular tension during the diastole.

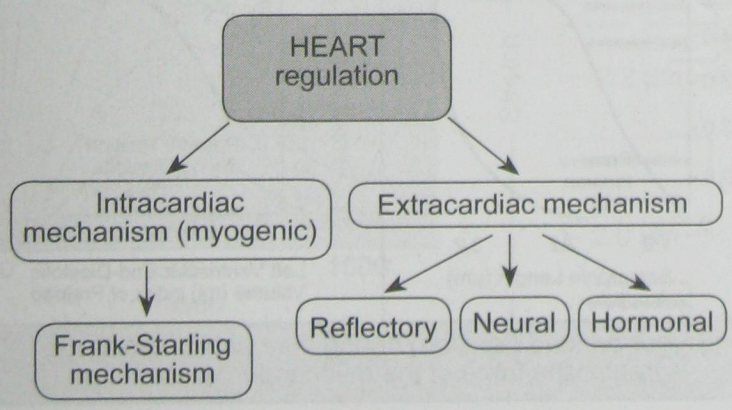


Fig. 18–1. The principal scheme of cardiac regulatory mechanisms.

The very important conclusion is that — *bigger preload (tension) during the ventricles diastole — greater the force of the myocardium contraction during the ventricles systole*. That is known as the Frank-Starling Law of the Heart description.

The above curve can be plotted with EDV (as well as end-diastolic pressure, or atrial pressure as the abscissa, as proxies for end-diastolic fiber length). The ordinate on the plot of Starling's law can also be a variable other than stroke volume (Fig. 18–2).

The interesting is the regulatory adjustment of two cardiac functional activity criteria — the degree to which the myocardium is stretched and the EDV. As the EDV rises within the physiological range, the myocardium is increasingly stretched and, as a result, contracts more forcefully. As discussed in Chapter 2, stretch can also increase the contraction strength of skeletal muscles (see Fig. 2-9). The resting length of skeletal muscles, however, is close to ideal, so that significant stretching decreases contraction strength. This is not true of the heart. Prior to filling with blood during diastole, the sarcomere lengths of myocardial cells are only about 1.5 μm (Fig. 18–2, left graph). At this length, the actin filaments from each side overlap in the middle of the sarcomeres, and the cells can contract only weakly (Fig. 18–3).

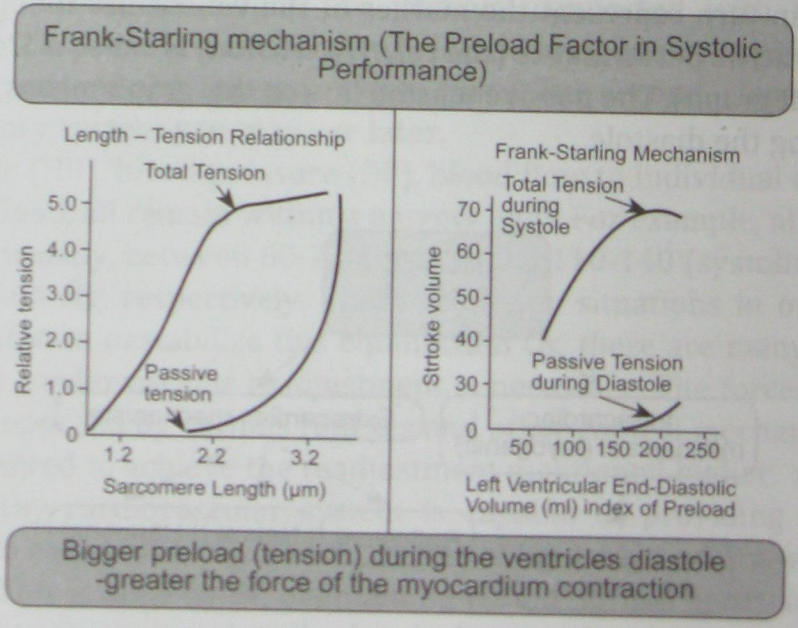


Fig. 18–2. The principal scheme of intracardiac regulatory mechanisms.

As the ventricles fill with blood, the myocardium stretches so that the actin filaments overlap with myosin only at the edges of the A bands (Fig. 18–3). This increases the number of interactions between actin and myosin, allowing more force to be developed during contraction. Since this more advantageous overlapping of actin and myosin is produced by stretching of the ventricles, and since the degree of stretching is controlled by the degree of filling (the EDV), the strength of contraction is intrinsically adjusted by the EDV.

Therefore, when the heart muscle is subjected to an increasing degree of stretch (*a through d*), it contracts more forcefully. The contraction strength is indicated on the y-axis as the tension. Notice that the time required to reach maximum contraction remains constant, regardless of the degree of stretch.

From this point of view important is that muscle length has a more pronounced effect on contraction strength in cardiac muscle than in skeletal muscle. That is, a particular increase in sarcomere length will stimulate contraction strength more in cardiac muscle than in skeletal muscle. This is believed to be due to an increased sensitivity of stretched cardiac muscle to the stimulatory effects of Ca^{2+} (Fig. 18–4).

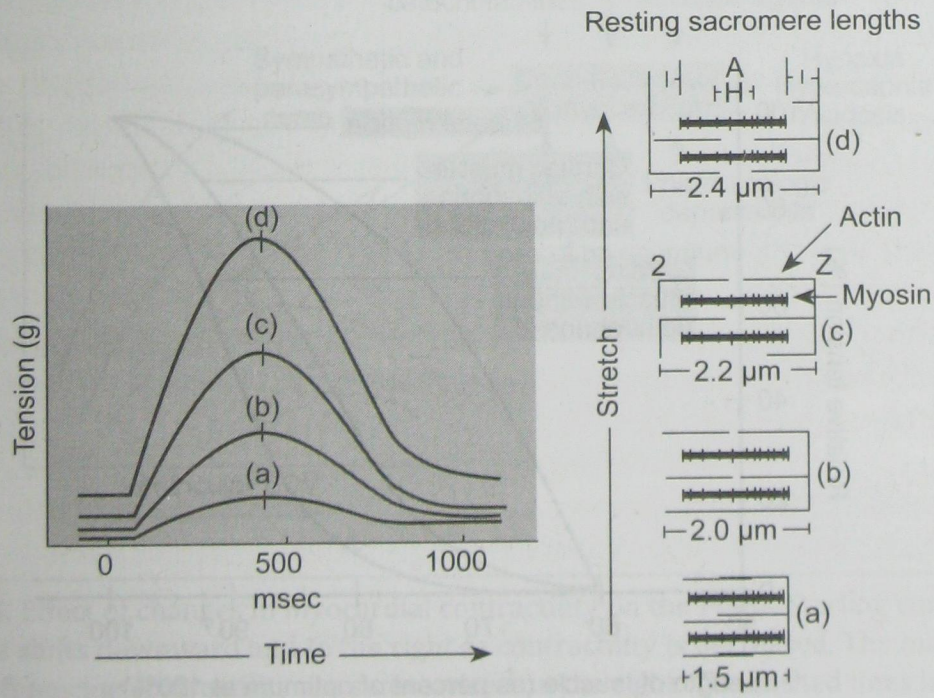


Fig. 18–3. The Frank–Starling Law of the Heart.

In this schematic comparison, all three curves demonstrate that each muscle contracts with its maximum force (100 % relative tension) at its own optimum length (100 % optimum length). As the length is decreased from optimum, each curve demonstra-tes a decreased contraction strength. Notice that the decline is more abrupt for cardiac muscle than for skeletal one, demonstrating the importance of the Frank–Starling relationship in heart physiology. At any length, however, epinephrine increases the strength of myocardial contraction, demonstrating a positive inotropic effect.

Regulation of cardiac output as a result of changes in cardiac muscle fiber length is sometimes called *heterometric regulation,* whereas regulation due to changes in contractility independent of length is sometimes called *homometric regulation.*

Factors other than end-diastolic fiber length can influence the force of ventricular contraction. Important that the myocardium contractility exerts a major influence on stroke volume. Different conditions produce different relationships between stroke volume (or work) to end-diastolic fiber length. For example, increased sympathetic nerve activity causes release of norepinephrine (NE). When the sympathetic nerves to the heart are stimulated, the whole length-tension curve shifts upward and to the left (Fig. 18–5).

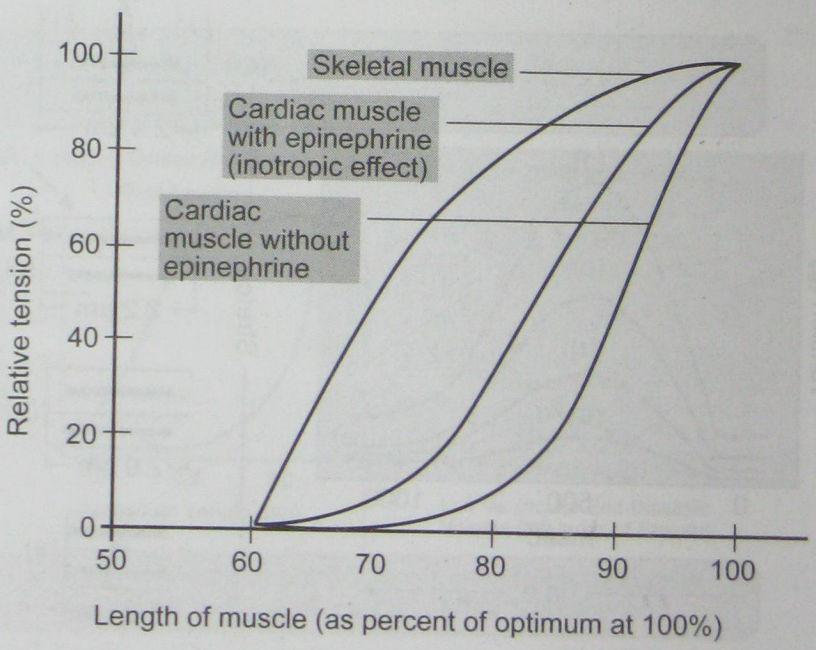


Fig. 18–4. The effect of muscle length and epinephrine on contraction strength.

The positive inotropic effect of the epinephrine liberated at the nerve endings is augmented by circulating epinephrine. There is a negative inotropic effect of vagal stimulation on the atrial muscle and a small negative inotropic effect on the ventricular muscle.

The catecholamines exert their inotropic effect via an action on cardiac β_1-adrenergic receptors and G_s, with resultant activation of adenylyl cyclase and increased intracellular cAMP. Xanthines such as caffeine and theophylline that inhibit the breakdown of cAMP are positively inotropic. Glucagon, which increases the formation of cAMP, is positively inotropic, and it has been recommended for use in the treatment of some heart diseases.

Cardiac glycosides, such as digitalis, tend to normalize contractility; that is, they shift the ventricular curve of the failing heart back to the left (Fig. 18–5). The positively inotropic effect of digitalis and related drugs is due to their inhibitory effect on the Na^+-K^+ ATPase in the myocardium.

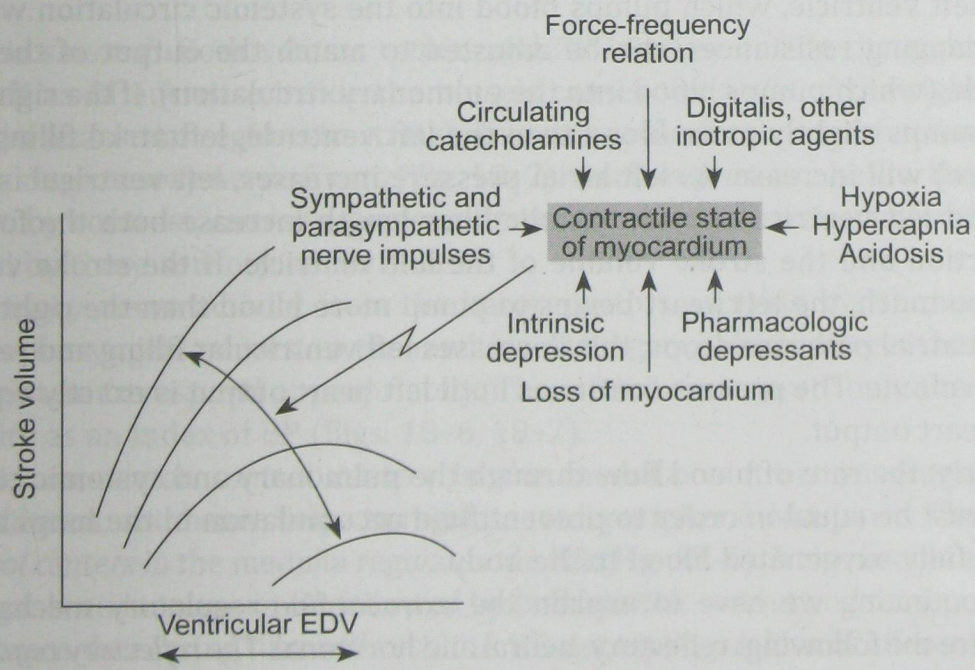


Fig. 18–5. Effect of changes in myocardial contractility on the Frank-Starling curve. The curve shifts downward and to the right as contractility is decreased. The major factors influencing contractility are summarized on the right. The dashed lines indicate portions of the ventricular function curves where maximum contractility has been exceeded; i.e., they identify points on the "descending limb" of the Frank-Starling curve.

Hypercapnia, hypoxia, acidosis, and drugs such as quinidine, procainamide, and barbiturates depress myocardial contractility. The contractility of the myocardium is also reduced in heart failure (intrinsic depression). During heart failure, the ventricular function curve is shifted to the right, causing a particular end-diastolic fiber length to be associated with less force of contraction and/or shortening and a smaller stroke volume.

The collection of ventricular function curves reflecting changes in contractility in a particular heart is known as a family of ventricular function curves.

The Frank–Starling law explains how the heart can adjust to a rise in total peripheral resistance: (i) a rise in peripheral resistance causes a decrease in the stroke volume of the ventricle, so that (ii) more blood remains in the ventricle and the EDV is greater for the next cycle; as a result, (iii) the ventricle is stretched to a greater degree in the next cycle and contracts more strongly to eject more blood. This allows a healthy ventricle to sustain a normal cardiac output.

A very important consequence of these events is that the cardiac output of the left ventricle, which pumps blood into the systemic circulation with its ever-changing resistances, can be adjusted to match the output of the right ventricle (which pumps blood into the pulmonary circulation). If the right ventricle pumps slightly more blood than the left ventricle, left atrial filling (and pressure) will increase. As left atrial pressure increases, left ventricular pressure and left ventricular end-diastolic fiber length increase both the force of contraction and the stroke volume of the left ventricle. If the stroke volume rises too much, the left heart begins to pump more blood than the right heart and left atrial pressure drops; this decreases left ventricular filling and reduces stroke volume. The process continues until left heart output is exactly equal to right heart output.

Clearly, the rate of blood flow through the pulmonary and systemic circulations must be equal in order to prevent fluid accumulation in the lungs and to deliver fully oxygenated blood to the body.

2. Continuing we have to explain the *extracardiac* regulatory mechanisms which are the following: reflectory, neural and hormonal. *The reflectory regulation* of cardiac function takes place with the help of the baroreceptor reflex (Fig. 18–6).

This reflex (named also as *Bainbridge reflex*) starts from the *stretch* receptors and *baroreceptors* activation. Stretch receptors are mainly located within the right atrium[1] and baroreceptors are located in the *aortic arch* and in the

[1] This is complete Bainbridge reflex.

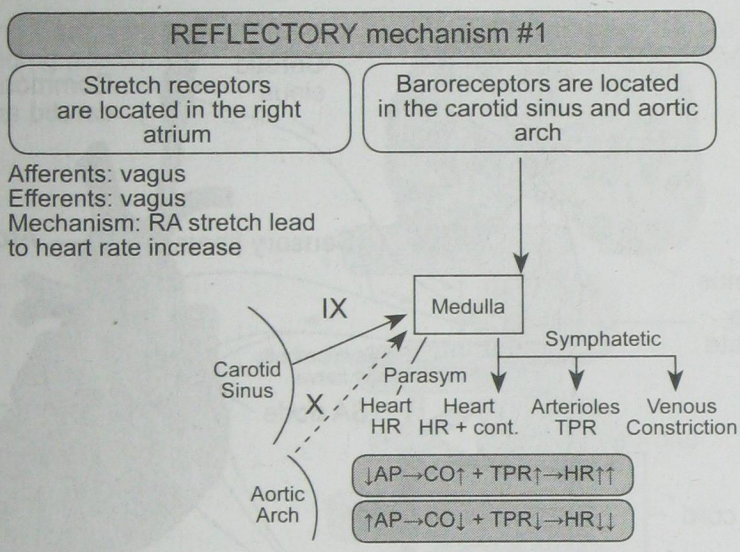


Fig. 18–6. The scheme of the baroreceptor reflex.

carotid sinuses[2]. Both venous return and BP increase inside the right atrium as well as BP increase inside the vessels causes the walls of the right atrium and these arterial regions to stretch, increasing the frequency of action potentials along sensory nerve fibers (Fig. 18–7). A fall in pressure below the normal range, by contrast, causes a decrease in the frequency of action potentials produced by these sensory nerve fibers.

Sensory nerve activity from the baroreceptors ascends, via the vagus and glossopharyngeal nerves, to the medulla oblongata, which directs the autonomic system to respond appropriately. The medulla interprets only the afferent activity as an index of BP (Figs. 18–6, 18–7).

The *vasomotor control centers* in the medulla controls vasoconstriction/vasodilation, and hence helps to regulate total peripheral resistance. The *cardiac control centers* in the medulla regulates the HR (Fig. 18–6). A rise in afferent activity signals an increase in BP followed by both cardiac output and total peripheral resistance decreasing that will result in HR decreasing. A loss of afferent activity signals a decrease of BP that will pursued by both cardiac output and total peripheral resistance increase. Finally, the HR index reflectory will increase (Fig. 18–6). Full description about the vasomotor center will be done in the chapter devoted to the circulatory system functional activity regulation (see chapter N 21).

[2] This reflex is known also as Hering reflex. Explanation see below.

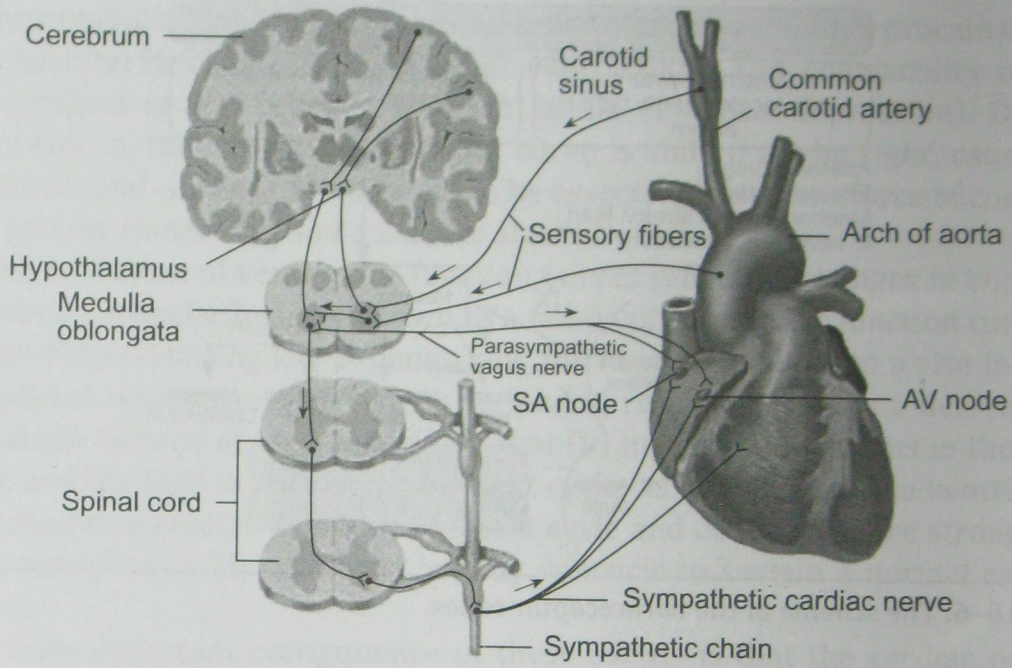


Fig. 18–7. Structures involved into the baroreceptor reflex.

Acting through the activity of motor fibers within the vagus and sympathetic nerves controlled by these brain centers, the baroreceptors function to counteract BP and HR changes so that fluctuations in HR and pressure are minimized. The baroreceptor reflex is activated whenever BP increases or decreases. The reflex is somewhat more sensitive to decreases in pressure than to increases, and is more sensitive to sudden changes in pressure than to more gradual changes. A good example of the importance of the baroreceptor reflex in normal physiology is its activation whenever a person goes from a lying to a standing position.

When a person goes from a lying to a standing position, there is a shift of 500 to 700 ml of blood from the veins of the thoracic cavity to veins in the lower extremities, which expand to contain the extra volume of blood. This pooling of blood in the lower extremities reduces the venous return and cardiac output, but the resulting fall in BP is almost immediately compensated for by the baroreceptor reflex. A decrease in baroreceptor sensory information, traveling in the glossopharyngeal nerve (IX) and the vagus nerve (X) to the medulla oblongata, inhibits parasympathetic activity and promotes sympathetic nerve activity. This produces an increase in cardiac rate and vasoconstriction, which help to maintain an adequate BP upon standing (Fig. 18–8).

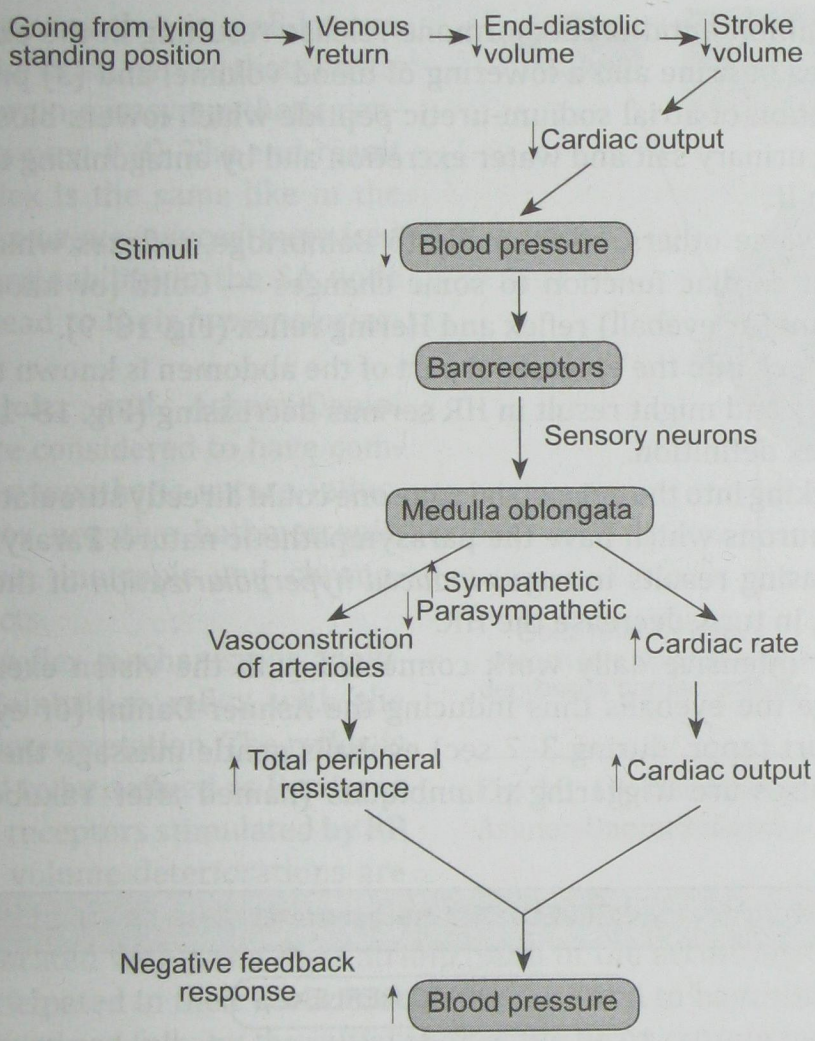


Fig. 18–8. The negative feedback control of blood pressure by the baroreceptor reflex.

Since the Bainbridge reflex may require a few seconds before it is fully effective, many people feel dizzy and disoriented if they stand up too quickly. If the baroreceptor sensitivity is abnormally reduced, perhaps by atherosclerosis, an uncompensated fall in pressure may occur upon standing. This condition — called *postural (orthostatic) hypotension* — can make a person feel extremely dizzy or even faint because of inadequate perfusion of the brain.

Resuming Bainbridge reflex, some words should be told concerning *right atrium stretch* receptors. Recall that these receptors are activated by venous return to the heart increase and, in response trigger the following processes: (1) induce reflectory tachycardia as a result of increased sympathetic nerve

activity; (2) inhibit antidiuretic hormone release resulting in the excretion of larger volumes of urine and a lowering of blood volume; and (3) promote increased secretion of atrial sodium-uretic peptide which lowers blood volume by increasing urinary salt and water excretion and by antagonizing the actions of angiotensin II.

There are some others, additionally to Bainbridge, reflexes which participate to adjust cardiac function to some changes — Goltz (or knock) reflex, Ashner-Daninni (or eyeball) reflex and Hering reflex (Fig. 18–9).

Weighty knock into the epigastrial part of the abdomen is known to depress cardiac activity and might result in HR serious decreasing (Fig. 18–10). This is the Holtz reflex definition.

While knocking into the epigastrial zone one could directly stimulate the plexus celiacus neurons which have the parasympathetic nature. Parasympathetic activity increasing results in *vagus-induced hyperpolarization* of the SA node cells that will, in turn, decrease the HR.

Often after intensive daily work connected with the vision exercises one could massage the eyeballs thus inducing the Ashner-Danini (or eyeball) reflex. After short (appr. during 3–7 sec) eyeballs gentle massage the intensive afferent impulses are triggering n. ambiquus (named after Yakubovich) via

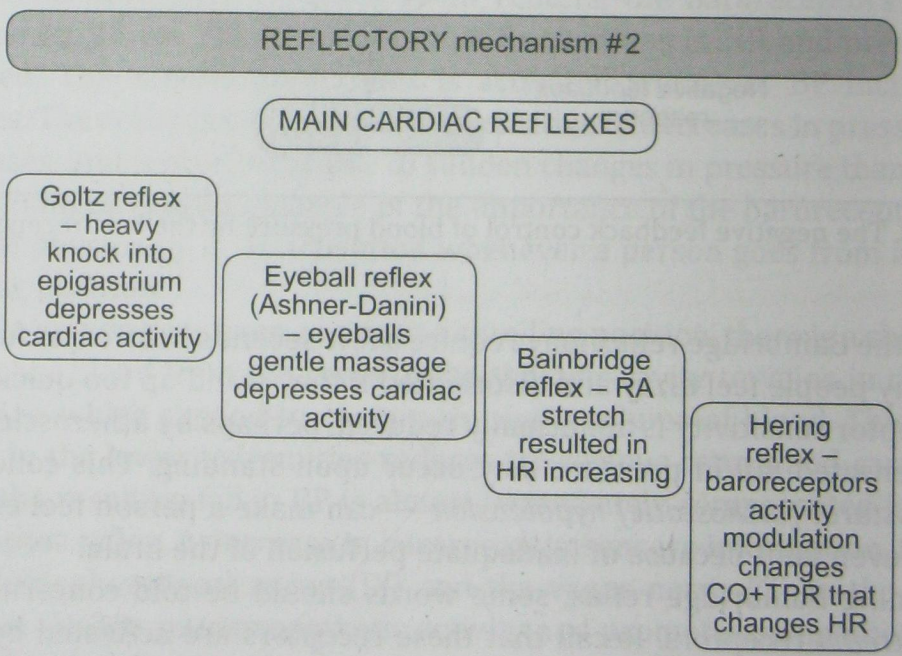


Fig. 18–9. The scheme of the main cardiac regulatory reflexes.

n. oculumotorius that is well-known as one of the brain nerves that constitutes the brain parasympathetic center (see Chapter # 8). The end-result of this reflex is the same like in the Holtz reflex *vagus-induced* increased K^+ ions permeability in the SA node cells that lead to their *hyperpolarization*.

Both Holtz and Ashner-Danini reflexes are considered to have complete parasympathetic nature inducing complex negative bathmotropic, dromotropic, inotropic and chronotropic effects.

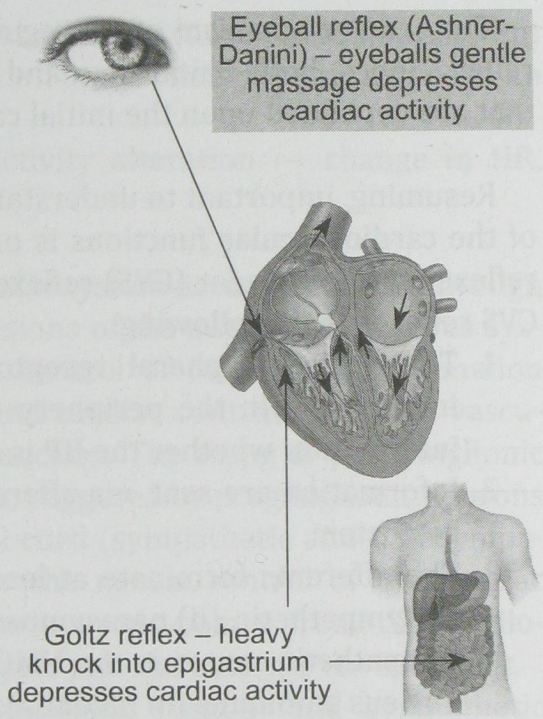


Fig. 18–10. The scheme of Holtz and Ashner-Danini reflexes.

Hering reflex mechanism is analogous to Bainbridge reflex with the following interpretation. The reflex is considered to be defined as Hering in case when receptors stimulated by BP and blood volume deteriorations are located within aortic arch and carotid sinus. Bainbridge reflex starts from the impulses located within the right atrium. Both of the according to the mechanisms participated in their realization are considered to have mixed character that is determined fully by the initial shift of the heart criteria indexes.

Imagine, for instance, that cardiac output, stroke volume will be decreased because of the different reasons and reflectory decreased BP will trigger the intravascularly located baroreceptors (on the aortic arch and carotid sinus) — these events are mimicking primarily decreased sympathetic and increased parasympathetic activity. Reflectory response in the form of HR increase (that is the sympathetic tone increase) resulting in cardiac output, stroke volume, venous return and BP increase will be evident. Therefore, originally happened parasympathetic tone activation results in reflectory sympathetic response. But if one could assume the cardiac output and stroke volume increase because of the stress stimuli, for instance (occurred due to sympathetic activity increase) the reflectory response will be seen in the form of the cardiac activity indexes decrease — all negative tropic effects. Initially happened sympathetic tone activation in this example results in reflectory parasympathetic response.

The same mechanism of the regulatory end-results supplies the Bainbridge reflex. Hence, both Bainbridge and Hering reflexes have mixed mechanisms that are depended upon the initial cardiac activity change.

Resuming, important to understand the following. The reflectory regulation of the cardiovascular functions is on the safe and quick. Several well known reflexes cardiovascular (CVS) reflexes were given above. General plan of these CVS reflexes is the following.

1. There are peripheral receptors (also called 'sensors') which collect information in the periphery (e.g., carotid sinus, the sensor, collecting information whether the BP is normal/high/or low).
2. Information are sent, via afferent ('sensory') nerves to the center in the brainstem.
3. The afferents terminate at least at two (sometimes three) centers, via, (i) sympathetic, (ii) parasympathetic, and sometimes (ii) respiratory. The sympathetic center is the VMC (in medulla). Parasympathetic center is nucleus ambiquus (in medulla). Respiratory center also locates within the medulla.
4. From VMC (vasomotor center), descending fibers, which usually remain intermingled with the reticulospinal fibers (an extra pyramidal group of fibers), descend to the lateral horn cells, (LHC) of Th_1 to L_2 segments of the spinal cord.
5. LHC cells of the thoracic and upper two lumbar segments of spinal cord rise to the preganglionic sympathetic fiber → which emerges out of the spinal cord → relies in a ganglion (=heart and blood vessels). Pre- and postganglionic sympathetic fibers are efferent or motor fibers.
6. Conclusion is that WMC activity alteration → alteration of sympathetic efferent activity → changes in cardiac activity as well as change in the tone of vascular smooth muscles (=vasodilatation or vasoconstriction) → change in HR, cardiac output, BP, redistribution of blood etc. etc.
7. Afferent impulses also reach the parasympathetic center, that is, nucleus ambiquus. From nucleus ambiquus fibers arise to join the vagal nerve fibers and thus these fibers ultimately become vagal parasympathetic fibers → supply the heart (=efferent parasympathetic to the heart).

Parasympathetic fibers supply the blood vessels too, but for ail practical purposes, they have no effect on the blood vessels, that is parasympathetic stimulation cannot produce vasodilatation or vasoconstriction (only in salivary

glands, or in penis, parasympathetic stimulation can cause vasodilatation; but they cause local salivary gland vasodilatation or erection of penis and have no effect on general circulation).

Conclusion is that parasympathetic activity alteration → change in HR, cardiac output → change in BP.

3. Neural regulation of the cardiovascular system occurs primarily through the *sympathetic* and *parasympathetic* divisions of the autonomic nervous system (ANS). The ANS contributes to the control of cardiovascular function through modulation of cardiac (*i.e.,* HR and cardiac contractility) and vascular (*i.e.,* peripheral vascular resistance) function. The firing of postganglionic parasympathetic and sympathetic neurons, triggered by preganglionic neurons in the brain (parasympathetic) and spinal cord (sympathetic and parasympathetic) are of the crucial importance for the neural control of the cardiac function. Afferent input influencing these neurons comes from the cardiovascular system, as well as from other organs and the external environment.

The heart is innervated by parasympathetic (vagus) and sympathetic (cardioaccelerator) nerve fibers (Fig. 18–11). Parasympathetic fibers release acetylcholine (ACh), which binds to muscarinic receptors of the SA node, the AV node and specialized conducting tissues. Stimulation of parasympathetic fibers causes a slowing of the heart rate and conduction velocity. The ventricles are only sparsely innervated by parasympathetic nerve fibers, and stimulation of these fibers has little direct effect on cardiac contractility.

Some cardiac parasympathetic fibers end on sympathetic nerves and inhibit the release of NE from sympathetic nerve fibers. Therefore, in the presence of sympathetic nervous system activity, parasympathetic activation reduces cardiac contractility.

The sympathetic fibers arise from the LHC of the thoracic and upper two lumbar segments of the spinal cord; therefore these spinal cord segments [LHC] are under the direct influence of the VMC.

Sympathetic fibers to the heart release NE, which binds to β_1-adrenergic receptors in the SA node, the AV node and specialized conducting tissues and cardiac muscle. Stimulation of these fibers causes increased HR (positive chronotropic effect), conduction velocity (positive dromotropic effect) and contractility (positive inotropic effect).

It has to be remembered, that in the intact man, whenever there is sympathetic stimulation, there is bound to be a concomitant stimulation of adrenal medulla. When stimulated, the adrenal medulla secretes both NE and epineph-

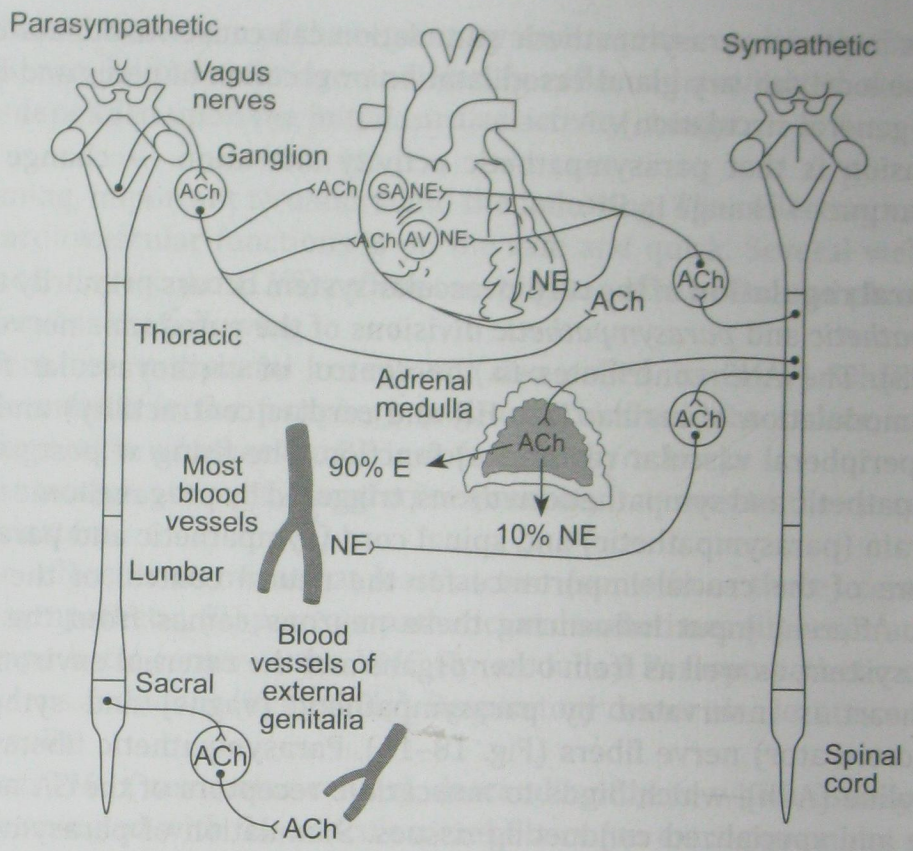


Fig. 18–11. Autonomic innervation of the cardiovascular system.

rine, but mostly epinephrine. Therefore, in intact man, sympathetic stimulation causes availability of both epinephrine and NE.

At rest, the frequency of the tonic discharge of the sympathetic system is between 1–2 per sec, this shoots up to about 10 per sec during intensive sympathetic stimulation.

The two divisions of the autonomic nervous system tend to oppose each other in their effects on the heart, and activities along these two pathways usually change in a reciprocal manner.

We need to indicate the other CNS structures (i.e., spinal cord) importance in the cardiovascular function regulation (Fig. 18–12).

Related to this it should be stressed that preganglionic sympathetic neurons normally generate a steady level of background postganglionic activity (tone). This *sympathetic tone* produces a background level of sympathetic vasoconstriction, cardiac stimulation, and adrenal medullary catecholamine secretion,

all of which contribute to the maintenance of normal blood pressure. This tonic activity is generated by excitatory signals from the medulla oblongata. When the spinal cord is acutely transected and these excitatory signals can no longer reach sympathetic preganglionic fibers, their tonic firing is reduced and blood pressure falls — an effect known as *spinal shock*.

Humans have spinal reflexes of cardiovascular significance (see Chapter # 6). For example, the stimulation of pain fibers entering the spinal cord below the level of a chronic spinal cord transection can cause reflectory vasoconstriction and increased both HR and BP.

The upperly located medulla oblongata has three major cardiovascular functions: (i) tonic excitatory signals generation to spinal sympathetic preganglionic fibers; (ii) integrating cardiovascular reflexes and (iii) integrating signals from supramedullary neural networks and from circulating hormones and drugs (Fig. 18–12).

Specific pools of neurons are responsible for elements of these functions. Neurons in the *rostral ventrolateral nucleus* (RVL) are normally active and provide tonic excitatory activity to the spinal cord (Fig. 18–12). Specific pools of neurons within the RVL have actions on heart and blood vessels. RVL neurons are critical in mediating reflex inhibition or activating sympathetic firing to the heart and blood vessels.

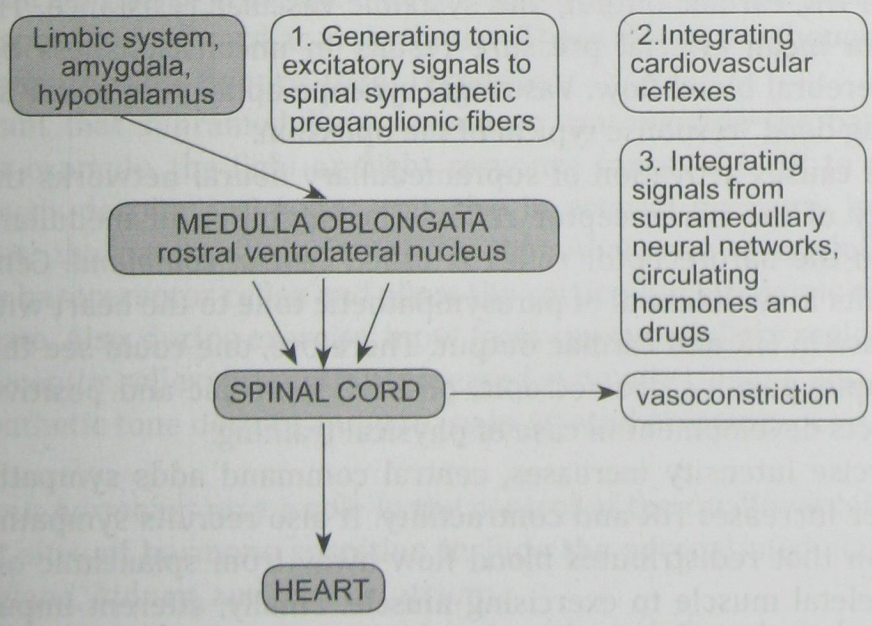


Fig. 18–12. Schematic representation of the cardiac activity neural regulation.

The cell bodies of cardiac preganglionic parasympathetic neurons are located in the *n. ambiguus*; the activity of these neurons is influenced by reflex input, as well as input from respiratory neurons.

The highest levels of organization in the ANS are the *supramedullary networks* of neurons with way stations in the limbic cortex, amygdala, and hypothalamus (Fig. 18–12). These supramedullary structures orchestrate cardiovascular correlates of specific patterns of emotion and behavior by their projections to the autonomic nervous system. Unlike the medulla, supramedullary networks do not contribute to the tonic maintenance of blood pressure, nor are they necessary for most cardiovascular reflexes, although they modulate reflex reactivity.

Upon stimulation of certain areas in the *hypothalamus*, cats demonstrate a stereotypical rage response, with spitting, clawing, tail lashing, back arching, and so on. This is accompanied by the autonomic *fight-or-flight response*. Cardiovascular responses include elevated HR and BP.

Vasovagal syncope (fainting) is a somatic and cardiovascular response to certain emotional experiences. Stimulation of the *cerebral cortex specific areas* can lead to a sudden relaxation of skeletal muscles, depression of respiration, and loss of consciousness. The cardiovascular events accompanying these somatic changes include profound parasympathetic-induced bradycardia and withdrawal of resting sympathetic vasoconstrictor tone. There is a dramatic drop in HR, cardiac output, and systemic vascular resistance. The overall decrease in mean arterial pressure results in unconsciousness because of lowered cerebral blood flow. Vasovagal syncope appears in lower animals as the "playing dead" response typical of the opossum.

Exercise causes activation of supramedullary neural networks that inhibit the activity of the baroreceptor reflex. The inhibition of medullary regions involved in the baroreceptor reflex is called *central command*. Central command results in withdrawal of parasympathetic tone to the heart with a resulting increase in HR and cardiac output. Therefore, one could see the positive bathmotropic, positive dromotropic, positive inotropic and positive chronotropic effects development in case of physical training.

As exercise intensity increases, central command adds sympathetic tone that further increases HR and contractility. It also recruits sympathetic vasoconstriction that redistributes blood flow away from splanchnic organs and resting skeletal muscle to exercising muscle. Finally, afferent impulses from exercising skeletal muscle terminate in the RVL where they further augment sympathetic tone.

The *diving response* is best observed in seals and ducks but it also occurs in humans. An experienced diver can exhibit intense HR slowing (parasympathetic) and peripheral vasoconstriction (sympathetic) of the extremities and splanchnic regions when his or her face is submerged in cold water. With breath holding during the dive, arterial PO_2 and pH fall and PCO_2 rises, and the chemoreceptor reflex reinforces the diving response. The arterioles of the brain and heart do not constrict and, therefore, cardiac output is distributed to these organs. This heart-brain circuit makes use of the oxygen stored in the blood that would normally be used by the other tissues, especially skeletal muscle. Once the diver surfaces, the HR and cardiac output increase substantially; peripheral vasoconstriction is replaced by vasodilatation, restoring nutrient flow and washing out accumulated waste products.

We want to stress the attention once more on the fact that the abovementioned reflexes and brain structures determined changes in cardiac functions are the same also for the vessels functional activity regulation.

Cardiovascular responses can be conditioned (as can other autonomic responses, such as those observed in Pavlov's famous experiments). Both classical and operant conditioning techniques have been used to raise and lower the BP and HR of animals. Humans can also be taught to alter their HR and BP using a variety of behavioral techniques. Behavioral conditioning of cardiovascular responses has significant clinical implications. Animal and human studies indicate that psychological stress can raise BP, increase atherogenesis and predispose to fatal cardiac arrhythmias. These effects are thought to result from an inappropriate fight-or-flight response.

Important that supramedullary responses can override the baroreceptor reflex. For example, the fight-or-flight response causes the HR to rise above normal levels despite a simultaneous rise in arterial pressure. In such circumstances, the neurons connecting the hypothalamus to medullary areas inhibit the baroreceptor reflex and allow the cortico-hypothalamic response to predominate. Also, during exercise, input from supramedullary regions inhibits the baroreceptor reflex, promoting increased sympathetic tone and decreased parasympathetic tone despite an increase in arterial pressure.

4. *Various hormones* play a role in the control of the cardiovascular system. Important sites of hormone secretion include the adrenal medulla, posterior pituitary gland, kidney, and cardiac atrium.

When the sympathetic nervous system is activated, the adrenal medulla releases *epinephrine* (> 90 %) and *NE* (< 10 %).

Changes in the circulating NE concentration are small relative to changes in NE resulting from the direct release from nerve endings close to vascular smooth muscle and cardiac cells. In the heart, circulating epinephrine binds to cardiac cell β_1- adrenergic receptors and reinforces the effect of NE released from sympathetic nerve endings.

A comparison of the responses to infusions of epinephrine and norepinephrine illustrates not only the different effects of the two hormones but also the different reflex response each one elicits (Fig. 18–13).

Epinephrine and NE have similar direct effects on the heart, but NE elicits a powerful baroreceptor reflex because it causes systemic vasoconstriction and increases mean arterial pressure. The reflex masks some of the direct cardiac effects of NE by significantly increasing cardiac parasympathetic tone. In contrast, epinephrine causes vasodilation in skeletal muscle and splanchnic beds. Systemic vascular resistance may actually fall and mean arterial pressure does

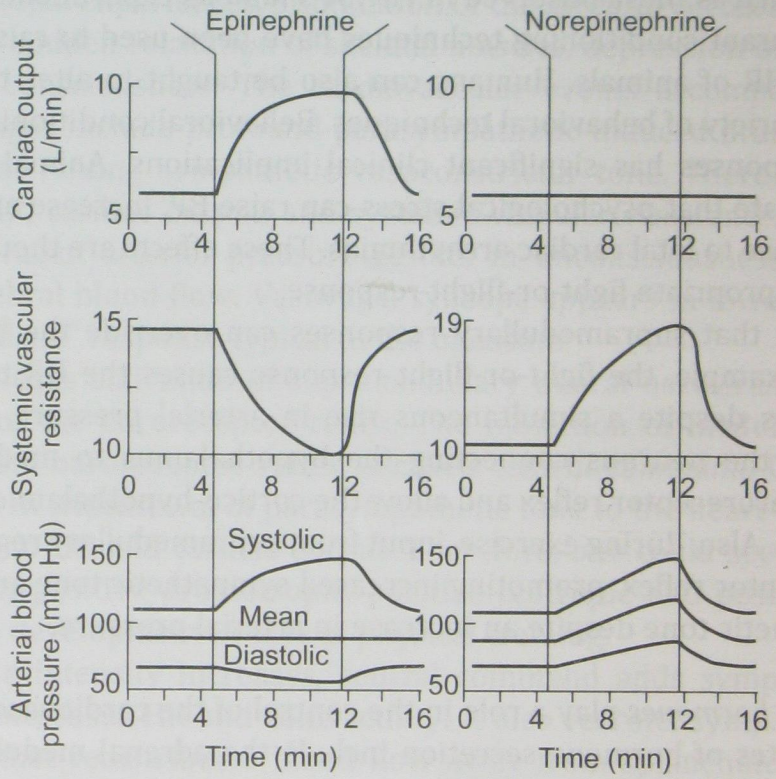


Fig. 18–13. A comparison of epinephrine and NE influence on the cardiovascular system activity.

not rise. The baroreceptor reflex is not elicited, parasympathetic tone to the heart is not increased, and the direct cardiac effects of epinephrine are evident.

At high concentrations, epinephrine binds to α_1-adrenergic receptors and causes peripheral vasoconstriction; this level of epinephrine is probably never reached except when it is administered as a drug.

During exercise, circulating levels of norepinephrine and epinephrine increase. Because of their enhanced response to circulating catecholamines, transplanted hearts can perform almost as well as normal hearts.

The following hormones are also important for the cardiovascular control — *renin* released from the kidney, *angiotensin II* formated from the renin using some biochemical transformations, *antidiuretic hormone, aldosterone* and in the less extent the *sodium-uretic peptide* produced in the right atrium. Being responsible for the both cardiac and vascular functions modulations these hormones effects will be discussed later (see Chapter # 21).

5. From the regulatory options *some ions concentration* change may significantly modifies the function of the cardiac muscle. These ions influence of cardiovascular activity is schematically shown on the Fig. 18–14.

Extracellular calcium ions concentration increase results in cardiac muscle excitability (positive bathmotropic effect) and contractility increase (positive inotropic effect). The mechanism of this Ca^{2+} effect analogous with the calcium-determined cardiac muscle contraction induction which was described earlier (see Chapter #16). The opposite situation is used in patients with ischemic heart disease to whom the calcium channel blockers are prescribed with the therapeutical aim.

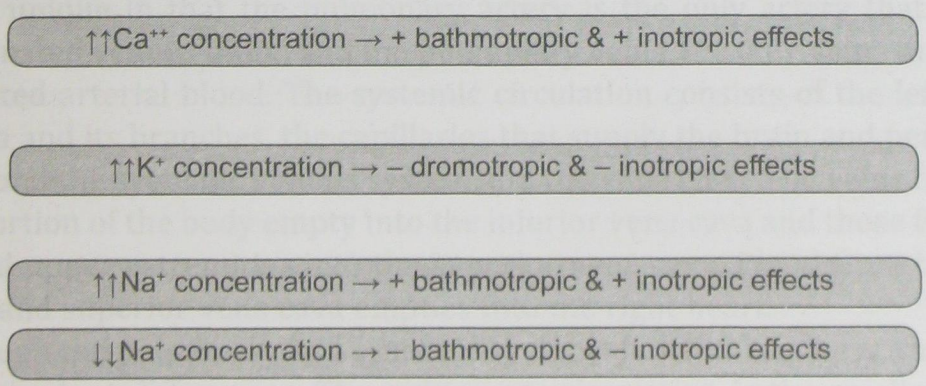


Fig. 18–14. Ionic influence on cardiac activity.

The increased blood plasma potassium concentration results in SA and AV nodes neurons hyperpolarization that, in turn, followed by negative dromotropic and inotropic effects. This situation masks the parasympathetic fibers influence on the heart.

Finally, the sodium ions concentration increasing resulted in heart excitability (dromotropic effect) and contractility (inotropic effect) enhancement. In case of decreased Na^+ ions concentration in the blood one could observe the opposite (negative dromotropic and inotropic) effects.

Chapter 19. Circulatory system. Hemodynamics. Classification of vessels

The main function of the *circulatory system,* which consists of the heart and blood vessels, is transport. The circulatory system delivers oxygen and nutrients needed for metabolic processes to the tissues, carries waste products from cellular metabolism to the kidneys and other excretory organs for elimination, and circulates electrolytes and hormones needed to regulate body function. This process of nutrient delivery is carried out with exquisite precision so that the blood flow to each tissue of the body is exactly matched to tissue need.

1. ORGANIZATION OF THE CIRCULATORY SYSTEM

Pulmonary and Systemic Circulations. The circulatory system can be divided into two parts: the *pulmonary circulation,* which moves blood through the lungs and creates a link with the gas exchange function of the respiratory system, and the *systemic circulation,* which moves blood throughout all the other tissues of the body (Fig. 19–1). The blood that is in the heart and pulmonary circulation is sometimes referred to as the *central circulation,* and that outside the central circulation as the *peripheral circulation.*

The pulmonary circulation consists of the right heart, the pulmonary artery, the pulmonary capillaries, and the pulmonary veins. The large pulmonary vessels are unique in that the pulmonary artery is the only artery that carries deoxygenated venous blood and the pulmonary veins, the only veins that carry oxygenated arterial blood. The systemic circulation consists of the left heart, the aorta and its branches, the capillaries that supply the brain and peripheral tissues, and the systemic venous system and the vena cava. The veins from the lower portion of the body empty into the inferior vena cava and those from the head and upper extremities into the superior vena cava. Blood from both the inferior and superior vena cava empties into the right heart.

Although the pulmonary and systemic systems function similarly, they have some important differences. The pulmonary circulation is the smaller of the two and functions with a much lower pressure. Because the pulmonary circulation is located in the chest near to the heart, it functions as a low pressure

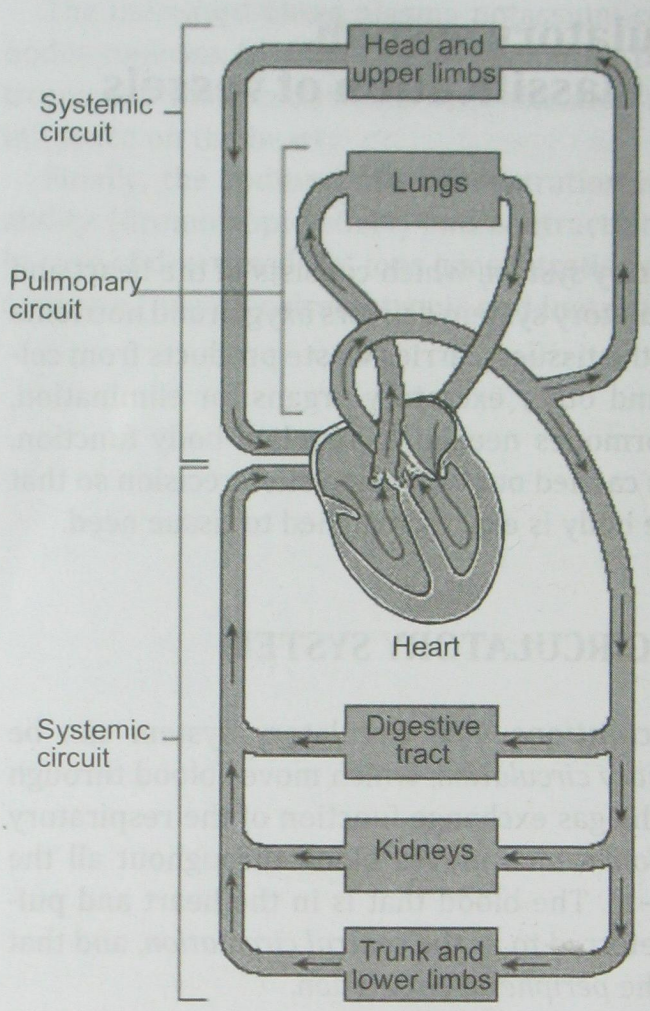


Fig. 19–1. Systemic and pulmonary circulations.

system with a mean arterial pressure of approximately 12 mm Hg. The low pressure of the pulmonary circulation allows blood to move through the lungs more slowly, which is important for gas exchange. Because the systemic circulation must transport blood to distant parts of the body, often against the effects of gravity, it functions as a high-pressure system, with a mean arterial pressure of 90 to 100 mm Hg.

The heart, which propels the blood through the circulation, consists of two pumps in series — the right heart which propels blood through the lungs and the left heart which propels blood to all other tissues of the body. The effective function of the circulatory system requires that the outputs of both sides of the heart pump the same amount of blood over time. If the output of the left heart were to fall below that of the right heart, blood would accumulate in the pulmonary circulation. Likewise, if the right heart were to pump less effectively than the left heart, blood would accumulate in the systemic circulation.

Volume and Pressure Distribution. Blood flow in the circulatory system depends on a blood volume that is sufficient to fill the blood vessels and a pressure difference across the system that provides the force that is needed to move blood forward. As shown in Fig. 19–2, approximately 4 % of the blood at any given time is in the left heart, 16 % is in the arteries and arterioles, 4 % is in the capillaries, 64 % is in the venules and veins, and 4 % is in the right heart.

The arteries and arterioles, which have thick, elastic walls and function as a distribution system, have the highest pressure. The capillaries are small, thin-walled vessels that link the arterial and venous sides of the circulation. They serve as an exchange system where transfer of gases, nutrients, and wastes take place. Because of their small size and large surface area, the capillaries contain the smallest amount of blood. The venules and veins, which contain the largest amount of blood, are thin-walled, distensible vessels that function as a reservoir to collect blood from the capillaries and return it to the right heart.

Blood moves from the arterial to the venous side of the circulation along a pressure difference, moving from an area of higher pressure to one of lower pressure. The pressure distribution in the different parts of the circulation is almost an inverse of the volume distribution (Fig. 19–2). The pressure in the arterial side of the circulation, which contains only approximately one sixth of the blood volume, is much greater than the pressure on the venous side of the circulation, which contains approximately two thirds of the blood. This pressure and volume distribution is due in large part to the structure and relative elasticity of the arteries and veins. It is the pressure difference between the arterial and venous sides of the circulation (approximately 84 mm Hg) that

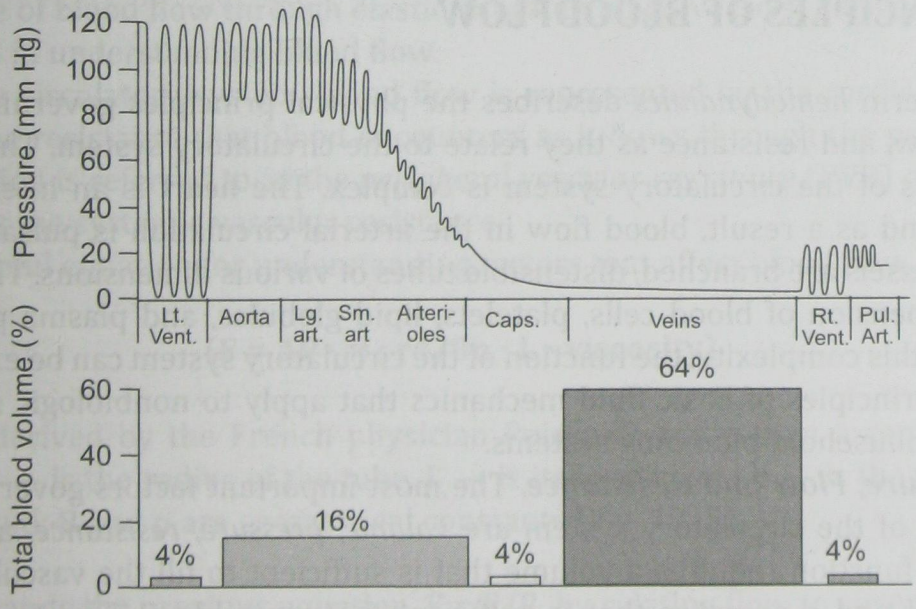


Fig. 19–2. Pressure and volume distribution in the systemic circulation. The graphs show the inverse relation between internal pressure and volume in different portions of the circulatory system.

provides the driving force for flow of blood in the systemic circulation. The pulmonary circulation has a similar arterial-venous pressure difference, although of a lesser magnitude, that facilitates blood flow.

Because the pulmonary and systemic circulations are connected and function as a closed system, blood can be shifted from one circulation to the other. In the pulmonary circulation, the blood volume (approximately 450 ml in the adult) can vary from as low as 50 % of normal to as high as 200 % of normal. An increase in intrathoracic pressure, such as occurs when exhaling against a closed glottis, impedes venous return to the right heart. This can produce a transient shift from the central to the systemic circulation of as much as 250 ml of blood. Body position also affects the distribution of blood volume. In the recumbent position, approximately 25 % to 30 % of the total blood volume is in the central circulation. On standing, this blood is rapidly displaced to the lower part of the body because of the forces of gravity. Because the volume of the systemic circulation is approximately seven times that of the pulmonary circulation, a shift of blood from one system to the other has a much greater effect in the pulmonary than in the systemic circulation.

2. PRINCIPLES OF BLOODFLOW

The term *hemodynamics* describes the physical principles governing pressure, flow, and resistance as they relate to the circulatory system. The hemodynamics of the circulatory system is complex. The heart is an intermittent pump, and as a result, blood flow in the arterial circulation is pulsatile. The blood vessels are branched, distensible tubes of various dimensions. The blood is a suspension of blood cells, platelets, lipid globules, and plasma proteins. Despite this complexity, the function of the circulatory system can be explained by the principles of basic fluid mechanics that apply to nonbiologic systems, such as household plumbing systems.

Pressure, Flow and Resistance. The most important factors governing the function of the circulatory system are *volume, pressure, resistance,* and *flow.* Optimal function requires a volume that is sufficient to fill the vascular compartment and a pressure that is sufficient to ensure blood flow to all body tissues. Blood flow is determined by two factors: (1) a pressure difference between the two ends of a vessel or group of vessels and (2) the resistance that blood must overcome as it moves through the vessel or vessels (Fig. 19–3).

The relation between pressure, resistance and flow is expressed by the equation

$$F = P/R,$$

in which F is the blood flow,

P is the difference in pressure between the two ends of the system, and

R is the resistance to flow through the system.

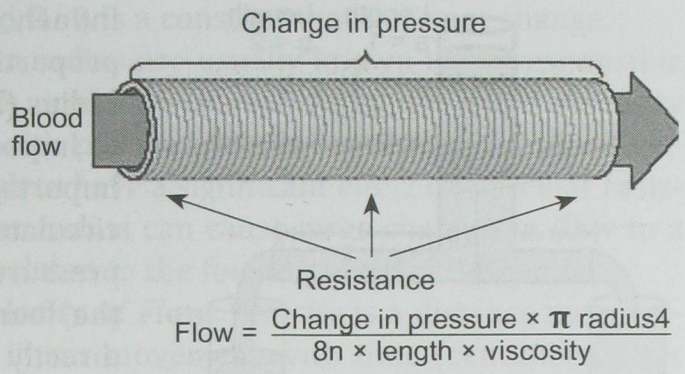


Fig. 19–3. Factors that affect blood flow (Poiseuille's law).
Increasing the pressure difference between the two ends of the vessel increases flow. Flow diminishes as resistance increases. Resistance is directly proportional to blood viscosity and the length of the vessel and inversely proportional to the fourth power of the radius.

In other words, the volume of fluid flowing through a rigid tube per unit time (Q) is proportional to the pressure difference (ΔP) between the ends of the tube and inversely proportional to the resistance to flow (R) (Fig. 19–4). While not exactly descriptive of blood flow through elastic, tapering blood vessels, Poiseuille's law is useful in understanding blood flow.

In the circulatory system, blood flow is represented by the cardiac output (CO). The resistance that blood encounters as it flows through the peripheral circula tion is referred to as the *peripheral vascular resistance* (PVR) or, sometimes, as the systemic vascular resistance.

A helpful equation for understanding factors that affect blood flow

$$(F = \Delta P \cdot \pi \cdot r4/8n \cdot L \cdot \text{viscosity})$$

was derived by the French physician Poiseuille more than a century ago where **r** — is the radius of the tube, **L** — is its length, and **h** — is the viscosity of the fluid; **8** and **p** are geometrical constants (Fig. 19–5).

It expands the previous equation, F = P/R, by relating flow to several determinants of resistance—radius, length, and viscosity. The last equation shows that the resistance to blood flow increases proportionately with increases in fluid viscosity or tube length. In contrast, radius changes have a much greater

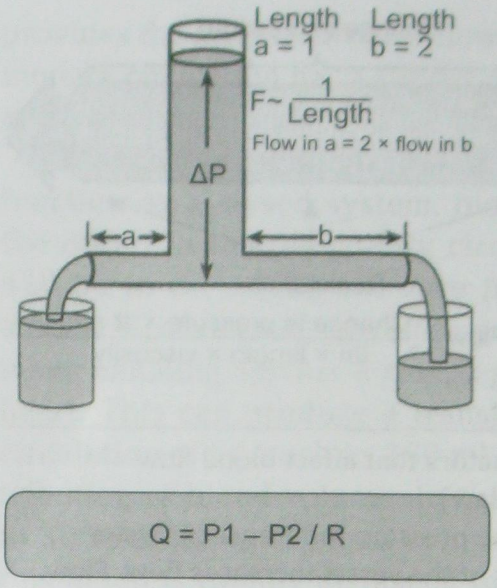


Fig. 19–4. The influence of the tube length on the blood flow (Poiseuille's law).

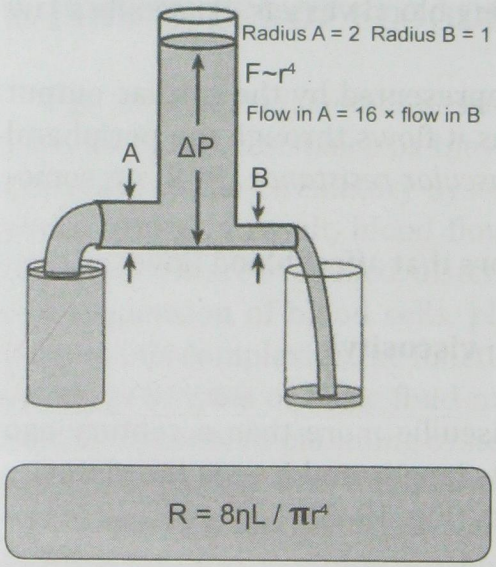


Fig. 19–5. The influence of the tube radius on the blood flow (Poiseuille's law).

influence because resistance is inversely proportional to the fourth power of the radius (Fig. 19–5).

Important is that there are two most important determinants of flow in the circulatory system — difference in pressure (ΔP) and the vessel radius to the fourth power (r^4). Because flow is directly related to the fourth power of the radius, small changes in vessel radius can produce large changes in flow to an organ or tissue. For example, if the pressure remains constant, the rate of flow is 16 times greater in a vessel with a radius of 2 mm ($2 \cdot 2 \cdot 2 \cdot 2$) than in a vessel with a radius of 1 mm.

Blood flow is also affected by the *viscosity of blood.* Viscosity is the resistance to flow caused by the friction of molecules in a fluid. The viscosity of a fluid is largely related to its thickness. The more particles that are present in a solution, the greater the frictional forces that develop between the molecules. Unlike water that flows through plumbing pipes, blood is a nonhomogeneous liquid. It contains blood cells, platelets, fat globules, and plasma proteins that increase its viscosity. The red blood cells, which constitute 40 % to 45 % of the formed elements of the blood, largely determine the viscosity of the blood. Under special conditions, temperature may affect viscosity. There is a 2 % rise in viscosity for each 1 °C decrease in body temperature, a fact that helps explain the sluggish blood flow seen in persons with hypothermia. The length (L) of ves-

sels does not usually change, and $8 \cdot \eta$ is a constant that does not change.

Resuming, only the changes in radius are usually known to be responsible for bloodflow variations. Length does not change. Although blood viscosity increases with hematocrit and with plasma protein concentration, blood viscosity only rarely changes enough to have a significant effect on vascular resistance. Small changes in arteriolar radius can cause large changes in flow to a tissue or organ because flow is related to the fourth power of the radius.

Cross-sectional Area and Velocity of Flow. *Velocity* is a distance measurement; it refers to the speed or linear movement with time (centimeters per second) with which blood flows through a vessel. *Flow* is a volume measurement (ml/second); it is determined by the cross-sectional area of a vessel and the velocity of flow (Fig. 19–6). When the flow through a given segment of the circulatory system is constant — as it must be for continuous flow — the velocity is inversely proportional to the cross-sectional area of the vessel (*i.e.,* the smaller the cross-sectional area, the greater the velocity of flow). This phenomenon can be compared with cars moving from a two-lane to a single-lane section of a highway. To keep traffic moving at its original place, cars would have to double their speed in the single-lane section of the highway. So it is with flow in the circulatory system.

Speaking about the cross-sectional area, one has to remember that aorta is a large-diameter vessel but it still represents the systemic segment with the smaller cross-sectional area. As aorta branches, the cross-sectional area of each individual vessel decreases, but collectively the cross-sectional area increases to reach a maximum in the capillaries. Cross-sectional area then decreases through the venous system (Fig. 19–7).

The linear velocity of blood flow in the circulatory system varies widely from 35 to 50 cm/sec in the aorta to 0.2 to 0.3 mm/sec in the capillaries (Fig. 19–8). This is because

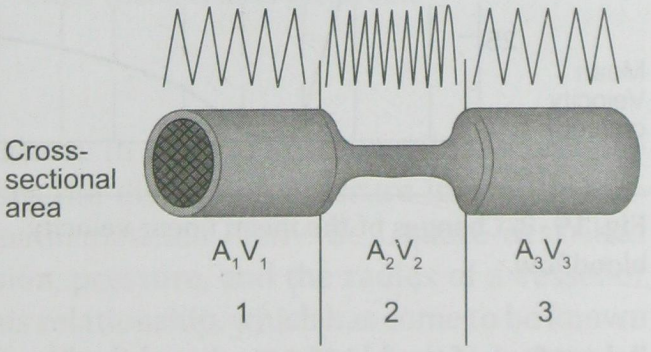


Fig. 19–6. Effect of cross-sectional area (A) on velocity (V) of flow. In section 1, velocity is low because of an increase in cross-sectional area. In section 2, velocity is increased because of a decrease in cross-sectional area. In section 3, velocity is again reduced because of an increase in cross-sectional area. Flow is assumed to be constant.

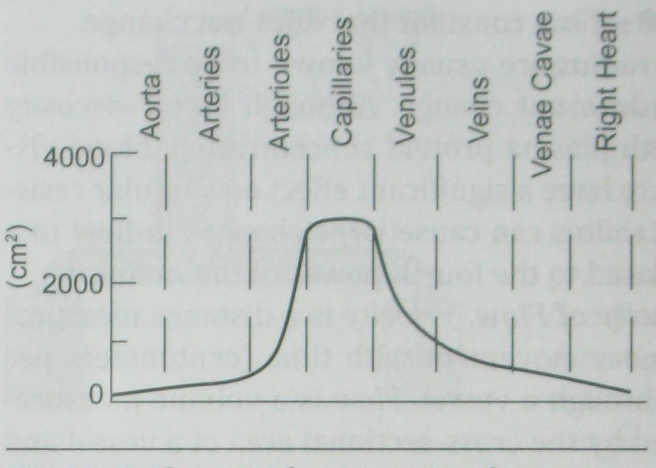


Fig. 19–7. Changes of cross-sectional area throughout the whole vascular network in the organism.

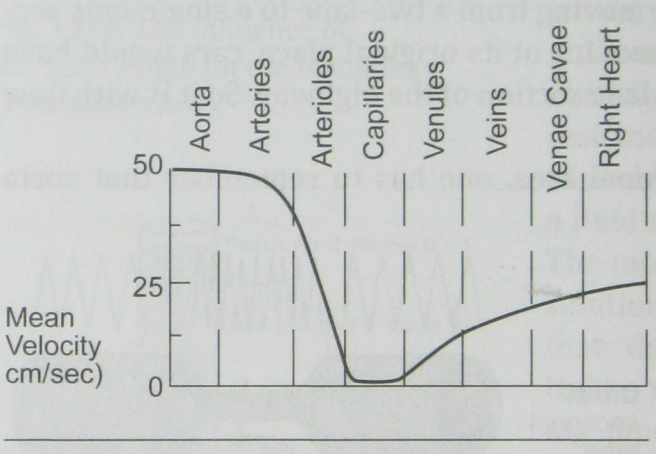


Fig. 19–8. Changes of the mean linear velocity bloodflow.

even though each individual capillary is very small, the total cross-sectional area of all the systemic capillaries greatly exceeds the cross-sectional area of other parts of the circulation. As a result of this large surface area, the slower movement of blood allows ample time for exchange of nutrients, gases, and metabolites between the tissues and the blood.

Therefore, the bloodflow velocity is inversely related to the cross-sectional area of all vessels in a segment. The average velocity is the greatest in aorta, then decreases to a minimum in the capillaries and then increases from the venules, veins toward the right atrium (Fig. 19–8).

Laminar and Turbulent Flow. Blood flow normally is *laminar,* with the blood components arranged in layers so that the plasma is adjacent to the smooth, slippery endothelial surface of the blood vessel, and the blood cells, including the platelets, are in the center or *axis* of the bloodstream (Fig. 19–9).

This arrangement reduces friction by allowing the blood layers to slide smoothly over one another, with the axial layer having the most rapid rate of flow. Under certain conditions, blood flow switches from *laminar* to *turbulent* flow (Fig. 19–9).

Turbulent flow can be caused by a number of factors, including high velocity of flow, change in vessel diameter, and low blood viscosity. The tendency for turbulence to occur increases in direct proportion to the velocity of flow.

Imagine the chaos as cars from a two- or three-lane highway converge on a single-lane section of the highway. The same type of thing happens in blood vessels that have been narrowed by disease processes, such as atherosclerosis. Low blood viscosity allows the blood to move faster and accounts for the transient occurrence of heart murmurs in some persons who are severely anemic. Turbulent flow may predispose to clot formation as platelets and other coagulation factors come in contact with the endothelial lining of the vessel. Turbulent flow often produces sounds that can be heard through the use of a stethoscope. For example, a heart murmur results from turbulent flow through a diseased heart valve.

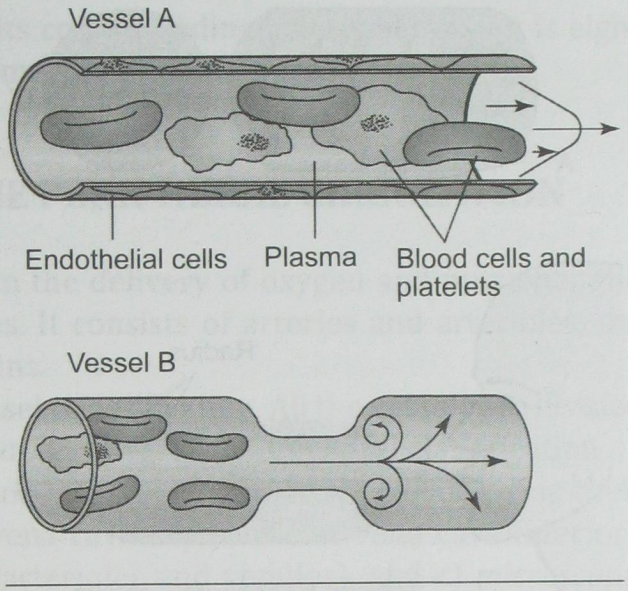


Fig. 19–9. Laminar and turbulent flow in blood vessels. Vessel A shows streamlined or laminar flow in which the plasma layer is adjacent to the vessel endothelial layer and blood cells are in the center of the bloodstream. Vessel B shows turbulent flow in which the axial location of the platelets and other blood cells is disturbed.

Wall Tension, Radius, and Pressure. In a blood vessel, *wall tension* is the force in the vessel wall that opposes the distending pressure inside the vessel. The French astronomer and mathematician Pierre de Laplace described the relationship between wall tension, pressure, and the radius of a vessel or sphere more than 200 years ago. This relationship, which has come to be known as *Laplace's law,* can be expressed by the equation, $\mathbf{P = T/r}$, in which T is wall tension, P is the intraluminal pressure, and r is vessel radius (Fig. 19–10, A).

Accordingly, the internal pressure expands the vessel until it is exactly balanced by the tension in the vessel wall. The smaller the radius, the greater the pressure needed to balance the wall tension. Laplace's law can also be used to express the effect of the radius on wall tension **(T = Pr).** This correlation can be compared with a partially inflated balloon (Fig. 19–10, B).

Because the pressure is equal throughout, the tension in the part of the balloon with the smaller radius is less than the tension in the section with the

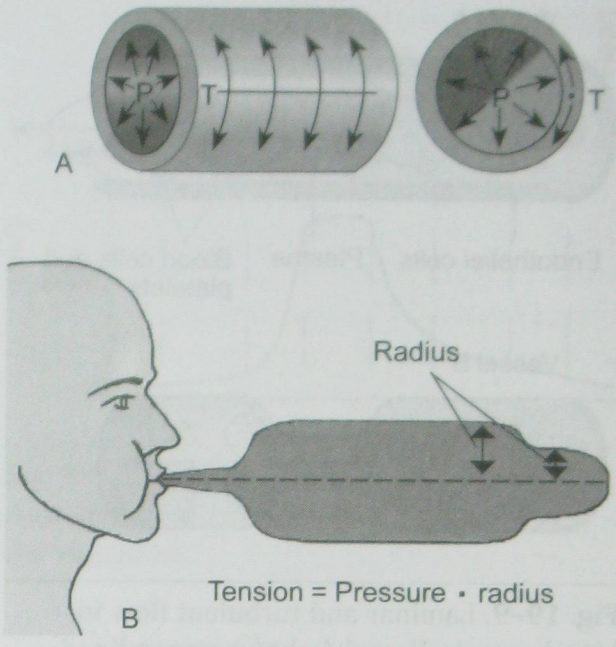


Fig. 19–10. Laplace's law relates pressure (P), tension (T), and radius in a cylindrical blood vessel (A).
The pressure expanding the vessel is equal to the wall tension multiplied by the vessel radius. (B) Effect of the radius of a cylinder on tension. In a balloon, the tension in the wall is proportional to the radius because the pressure is the same everywhere inside the balloon. The tension is lower in the portion of the balloon with the smaller radius.

larger radius (Fig. 19–10, B). The same holds true for an arterial aneurysm in which the tension and risk of rupture increase as the aneurysm grows in size.

Laplace's law was later expanded to include wall thickness

(T = P · r/wall thickness).

Wall tension is inversely related to wall thickness, such that the thicker the vessel wall, the lower the tension, and vice versa. In hypertension, arterial vessel walls hypertrophy and become thicker, thereby reducing the tension and minimizing wall stress.

Laplace's law can also be applied to the pressure required to maintain the potency of small blood vessels. Providing that the thickness of a vessel wall remains constant, it takes more pressure to overcome wall tension and keep a vessel open as its radius decreases in size. The critical closing pressure refers to the point at which vessels collapse so that blood can no longer flow through them.

Distention and Compliance. *Compliance* refers to the total quantity of blood that can be stored in a given portion of the circulation for each millimeter rise in pressure. Compliance reflects the *distensibility* of the blood vessel. The distensibility of the aorta and large arteries allows them to accommodate the pulsatile output of the heart. The most distensible of all vessels are the veins, which can increase their volume with only slight changes in pressure, allowing them to function as a reservoir for storing large quantities of blood that can be returned to the circulation when it is needed. The compliance of a vein

is approximately 24 times that of its corresponding artery, because it is eight times as distensible and has a volume three times as great.

3. BLOOD VESSELS AND THE PERIPHERAL CIRCULATION

The vascular system functions in the delivery of oxygen and nutrients and removal of wastes from the tissues. It consists of arteries and arterioles, the capillaries, and the venules and veins.

Firstly, you have to know the vessels classification. All the vessels are divided into the *arterial*, *venous* and *lymphatic* vessels. This is general classification.

According to the vessels diameter one could distinguish the following their types: a) *big-sized vessels* (aorta, vena cava superior and vena cava inferior), b) *vessels of the medial diameter* (arterioles and venules), and c) *microscopic vessels* (capillaries).

And the last one, functional classification in accordance to which there are the following types of the vessels: a) *elastic* (or amortizing) vessels, for instance, aorta, a.a. carotic communis and subclavius; b) *resistant* vessels with the big expression of the muscular layer in their walls (f.i., arterioles); c) *shunt-vessels* (or vessels-anastomosis) which promote wide vascular net; d) *exchange vessels* (capillaries due to their thin vascular wall that allows rapid O_2, CO_2, hormones, and other substrates exchanges with the surrounding tissues), and e) *capacitance vessels* (venules and veins of the different caliber).

Blood vessels are dynamic structures that constrict and relax to adjust blood pressure and flow to meet the varying needs of the many different tissue types and organ systems. Structures such as the heart, brain, liver, and kidneys require a large and continuous flow to carry out their vital functions. In other tissues, such as the skin and skeletal muscle, the need for blood flow varies with the level of function. For example, there is a need for increased blood flow to the skin during fever and for increased skeletal muscle blood flow during exercise.

Blood Vessels. All blood vessels, except the capillaries, have walls composed of three layers, or coats, called *tunicae*. The *tunica externa,* or *tunica adventitia,* is the outermost covering of the vessel. This layer is composed of fibrous and connective tissues that support the vessel. The *tunica media,* or middle layer, is largely a smooth muscle layer that constricts to regulate and control the diameter of the vessel.

The *tunica intima,* or inner layer, has an elastic layer that joins the media and a thin layer of endothelial cells that lie adjacent to the blood. The endo-

thelial layer provides a smooth and slippery inner surface for the vessel. This smooth inner lining, as long as it remains intact, prevents platelet adherence and blood clotting.

The layers of the different types of blood vessels vary with vessel function. The walls of the arterioles, which control blood pressure, have large amounts of smooth muscle. Veins are thin-walled, distensible, and collapsible vessels. Capillaries are single-cell–thick vessels designed or the exchange of gases, nutrients, and waste materials.

Vascular Smooth Muscle. Smooth muscle contracts slowly and generates high forces for long periods with low energy requirements; it uses only 1/10 to 1/300 the energy of skeletal muscle. These characteristics are important in structures, such as blood vessels, that must maintain their tone day in and day out. Although vascular smooth muscle contains actin and myosin filaments, these contractile filaments are not arranged in striations as they are in skeletal and cardiac muscle. The smooth muscle fibers are instead linked together in a strong cable-like system that generates a circular pull as it contracts.

In addition, smooth muscle has less well developed sarcoplasmatic reticulum for storing intracellular calcium than do skeletal and cardiac muscle, and it has very few fast sodium channels. Instead, depolarization of smooth muscle relies largely on extracellular calcium, which enters through calcium channels in the muscle membrane. These channels respond to changes in membrane potential or receptor-activated responses to chemical mediators such as norepinephrine. Sympathetic nervous system control of vascular smooth muscle tone occurs by way of receptor-activated channels. In general, α-adrenergic receptors are excitatory and produce vasoconstriction, and β-adrenergic receptors are inhibitory and produce vasodilatation. Calcium-channel blocking drugs cause vasodilatation by blocking calcium entry through the calcium channels.

Arterial System. The arterial system consists of the large and medium-sized arteries and the arterioles. Arteries are thick-walled vessels with large amounts of elastic fibers. The elasticity of these vessels allows them to stretch during cardiac systole, when the heart contracts and blood is ejected into the circulation, and to recoil during diastole, when the heart relaxes. The arterioles, which are predominantly smooth muscle, serve as resistance vessels for the circulatory system. They act as control valves through which blood is released as it moves into the capillaries. Changes in the activity of sympathetic fibers that innervate these vessels cause them to constrict or to relax as needed to maintain blood pressure.

Aortic Pressure Pulse. The delivery of blood to the tissues of the body is dependent on pressure pulsations or waves of pressure that are generated by the intermittent ejection of blood from the left ventricle into the distensible aorta and large arteries of the arterial system. The aortic pressure pulse represents the energy that is transmitted from molecule to molecule along the length of the vessel (Fig. 19–11). In the aorta, this pressure pulse is transmitted at a velocity of 4 to 6 m/ sec, which is approximately 20 times faster than the flow of blood. Therefore, the pressure pulse has no direct relation to blood flow and could occur if there was no flow at all.

When taking a pulse, it is the pressure pulses that are felt, and it is the pressure pulses that produce the Korotkov sounds heard during blood pressure measurement. The tip or maximum deflection of the pressure pulsation coincides with the systolic blood pressure, and the minimum point of deflection coincides with the diastolic pressure. Both the pressure values and the conformation of the pressure wave changes as it moves though the peripheral arteries (Fig. 19–11).

As the pressure wave moves out through the aorta into the arteries, it changes as it collides with reflected waves from the periphery. This is why the systolic pressure is higher in the medium-sized arteries than in the aorta even though the diastolic pressure is lower. After its initial amplification, the pressure pulse becomes smaller and smaller as it moves through the smaller arteries and arterioles,

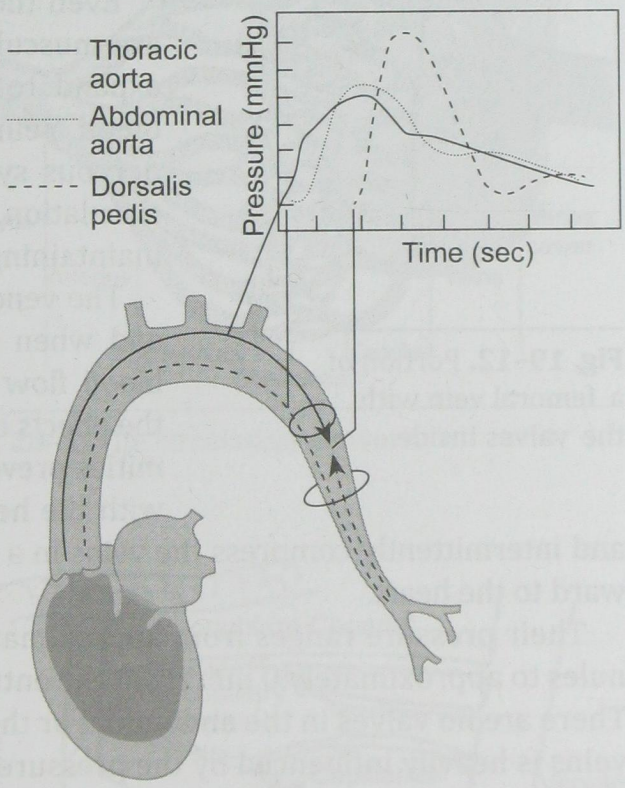


Fig. 19–11. Amplification of the arterial pressure wave as it moves forward in the peripheral arteries. This amplification occurs as a forward-moving pressure wave merges with a backward moving reflected pressure wave. (Inset) The amplitude of the pressure pulse increases in the thoracic aorta, abdominal aorta and a. dorsalis pedis.

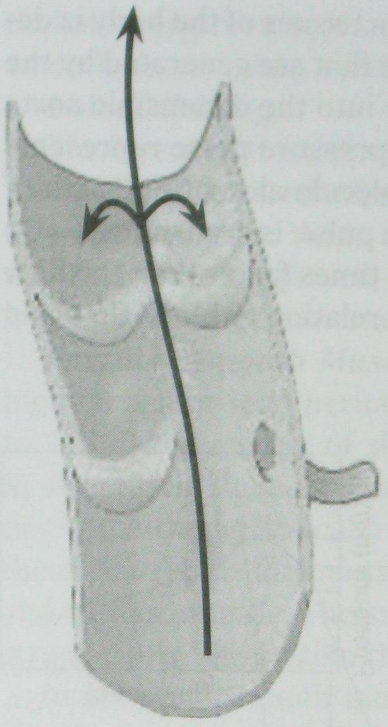


Fig. 19–12. Portion of a femoral vein with the valves inside.

until it disappears almost entirely in the capillaries. This allows for continuous, rather than pulsatile, flow in the capillary beds.

Venous System. The veins and venules are thin-walled, distensible, and collapsible vessels. The venules collect blood from the capillaries, and the veins transport blood back to the heart. The veins are capable of enlarging and storing large quantities of blood, which can be made available to the circulation as needed.

Even though the veins are thin walled, they are muscular. This allows them to contract or expand to accommodate varying amounts of blood. Veins are innervated by the sympathetic nervous system. When blood is lost from the circulation, the veins constrict as a means of maintaining intravascular volume.

The venous system is a low-pressure system, and when a person is in the upright position, blood flow in the venous system must oppose the effects of gravity. Valves in the veins of extremities prevent retrograde flow (Fig. 19–12), and with the help of skeletal muscles that surround and intermittently compress the veins in a milking manner, blood is moved forward to the heart.

Their pressure ranges from approximately 10 mm Hg at the end of the venules to approximately 0 mm Hg at the entrance of the vena cava into the heart. There are no valves in the abdominal or thoracic veins, and blood flow in these veins is heavily influenced by the pressure in the abdominal and thoracic cavities, respectively.

Chapter 20. Microcirculation and lymph flow

1. MICROCIRCULATION

The vessels of the microcirculation consist of the arterioles, the capillaries and the venules (Fig. 20–1). Exchange of gases, nutrients, and metabolites between the tissues and the circulating blood takes place here. The lymphatic system represents an accessory system that removes excess fluid, proteins, and large particles from the interstitial spaces and returns them to the circulation. Because of their size, these particles cannot be reabsorbed into the capillaries.

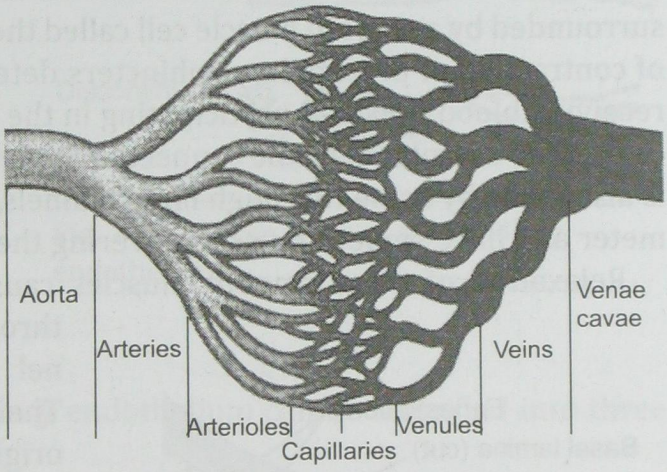


Fig. 20–1. The circulation network.

The detailed architecture of the microcirculation varies in different organs and tissues. An idealized diagram of the vessels in the microcirculation is shown in the Fig. 20–2.

The *arterioles*, as mentioned in Chapter #19, are the major resistance vessels; the capillaries are exchange vessels

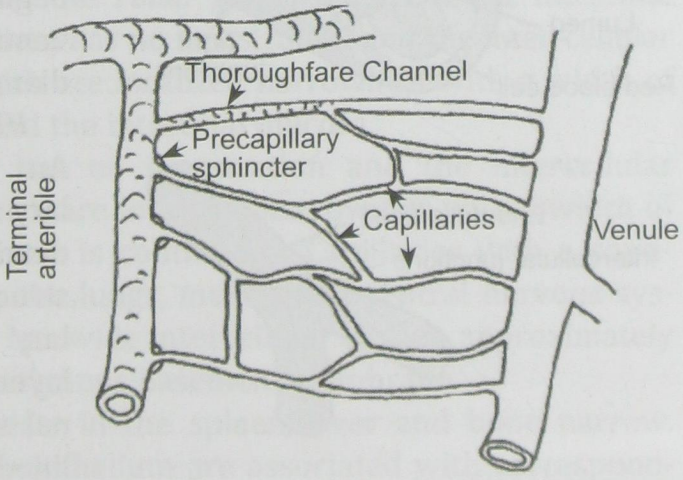


Fig. 20–2. The principal diagram of the microcirculation vessels organization.

responsible for ransport of materials between blood and tissue; the *venules* serve as collecting vessels, volume reservoirs and the site of postcapillary resistance. The arterioles are approximately 50 microns in diameter with a single layer of smooth muscle, which provides the mechanism for control of vascular hindrance. The last part of the arterioles which feed into the capillaries are called terminal arterioles.

The *capillaries*, which have a diameter of approximately 4–8 microns, possess only a single layer of endothelium and a basement membrane and are devoid of smooth muscle cells. In some tissues the entrance to the capillary is surrounded by a smooth muscle cell called the precapillary sphincter. The state of contraction of precapillary sphincters determines the number of capillaries receiving blood flow and participating in the transcapillary exchange process.

In some vascular beds the connection between the arterioles and the venules is also made by some thorough fare channels, which are 10–15 microns in diameter and have smooth muscles covering their arteriolar end.

Relaxation of these smooth muscles cause some shunting of blood flow through the thoroughfare channel away from the true capillaries. There are, however, also capillaries originating from the thoroughfare channel. On the venous side of the circulation, the capillaries coalesce into venules. A smooth muscle coat begins to appear in the medium sized venules, approximately 50 microns in diameter and larger.

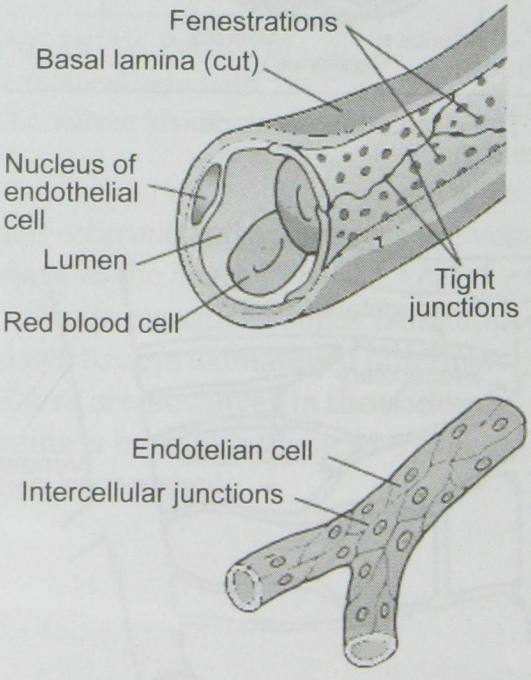


Fig. 20–3. Endothelial cells and intercellular junctions in a section of capillary.

Ultrastructure of capillaries. An understanding of transcapillary exchange necessitates a rather detailed knowledge of the ultrastructure of the capillary. The capillary wall is composed of a single layer of endothelial cells. The luminal surface of the endothelial cell is lined by a thin layer of glycoprotein, and the tissue side is covered by the basement membrane (Fig. 20–3). The roles of the glycoprotein coating and

the basement membrane in controlling capillary permeability have not been established.

The endothelial cell layer is probably the principal factor in the regulation of transcapillary movement, although the basement membrane may play a significant role in the renal glomeruli. Intracellular junctions join the capillary endothelial cells; these are called the *capillary pores.*

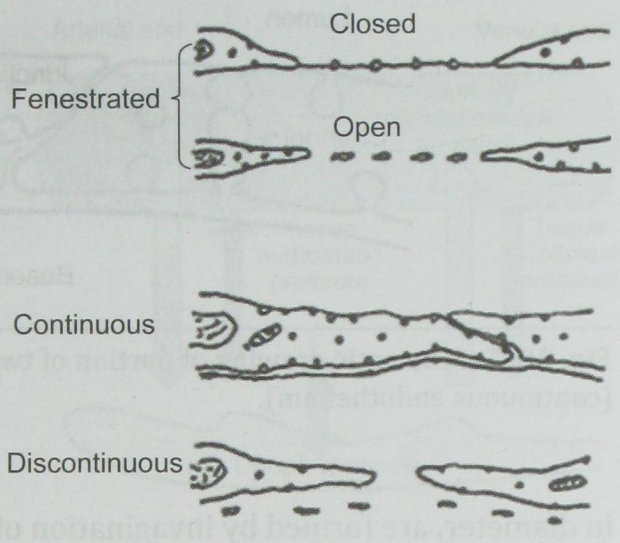


Fig. 20–4. Three types of the capillary endothelium.

The adjacent endothelial cells in a capillary are separated by an inter-cellular space, the width of which shows regional differences. Depending on its morphological features, the capillary endothelium can be classified into three types (Fig. 20–4).

The *fenestrated* endothelium has regional thinning (fenestrations) of the capillary wall. Some of the fenestrations are completely open, whereas others are covered by a thin diaphragm (closed).

Fenestrated capillaries exist in the renal glomerulus and the intestinal mucosa. The *continuous* endothelium has no fenestration and the inter-cellular space contains gap junctions, which are localized narrowings with a width of 5–10 nm. The renal glomerulus and the intestinal mucosa.

The *continuous* endothelium has no fenestration and the intercellular space contains gap junctions, which are localized narrowings with a width of 5–10 nm. The basement membrane is continuous. Capillaries with a continuous endothelium are found in the lungs, muscle and central nervous system. Discontinuous endothelium has wide intercellular spaces, approximately 50–100 nm in width, and a discontinuous basement membrane.

Discontinuous endothelium exists in the spleen, liver and bone narrow. These structural differences in endothelium are associated with corresponding variations in capillary permeability.

An important ultrastructural feature of the capillary endothelium is the presence of plasmalemmal vesicles. These vesicles, which are approximately 70 nm

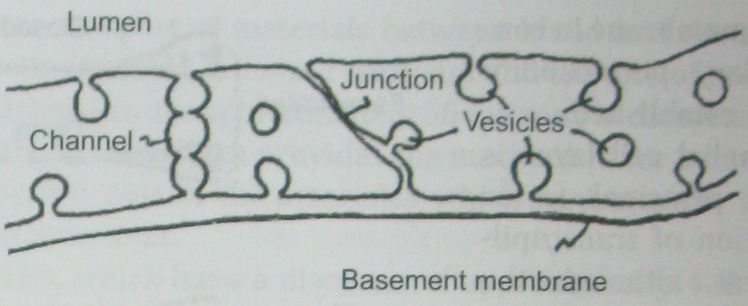


Fig. 20–5. Schematic drawing of portion of two endothelial cells (continuous endothelium).

in diameter, are formed by invagination of the plasmalemmal membrane of the endothelial cell. Some of the vesicles are attached to the plasmalemmal membrane on both the blood side and the tissue side, whereas others appear to be free bodies in the cytoplasm but are in reality connected to the outside of the cell (Fig. 20–5). When the vesicle is attached to the plasmalemmal membrane, it has an opening of approximately 20–30 nm.

The transport of materials across capillary walls occurs mainly by two processes: diffusion and filtration. The driving potential for diffusion is the concentration gradient The driving potential for filtration (or reabsorption) is the pressure gradient (fluid mechanical pressure and colloidal osmotic pressure). For a given potential, the rate of transport is a function of the characteristics of the membrane in relation to the properties of the substance being transported. Therefore, the ultrastructure of the capillary endothelium is an important determinant of transcapillary transport

2. CAPILLARY/INTERSTITIAL FLUID EXCHANGE

The transfer of water between the vascular and interstitial compartments occurs at the capillary level. Four forces control the movement of water between the capillary and interstitial spaces. Important is that these four forces are acting on each side of the capillary — arterial and venous ones:

1) the *capillary filtration pressure*, which pushes water out of the capillary into the interstitial spaces;
2) the *capillary colloidal osmotic pressure*, which pulls water back into the capillary;

3) the *interstitial tissue hydrostatic pressure,* which opposes the movement of water out of the capillary;
4) the *tissue colloidal osmotic pressure,* which pulls water out of the capillary into the interstitial spaces (Fig. 20–6).

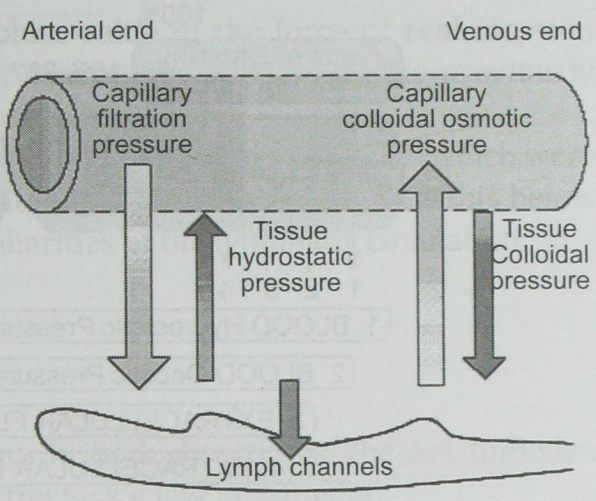


Fig. 20–6. Exchange of fluid on the capillary level.

Ernest Starling was the first who described the relations between hydrostatic and colloid pressures. The relationship is defined for a single capillary by the Starling-Landis equation:

$$V = Kh \cdot A\,[(Pc - Pt) - \sigma \cdot (COPp - COPt)]$$

where V is the net volume of fluid moving across the capillary wall per unit of time,

Kh is the hydraulic conductivity for water, which is the fluid permeability of the capillary wall.

- A is the vascular surface area;
- Pc is the capillary hydrostatic pressure;
- Pt is the tissue hydrostatic pressure.

COPp and COPt represent the plasma and tissue colloid osmotic pressures, respectively, and σ is the reflection coefficient for plasma proteins.

Normally, the combination of these four forces is such that only a small excess of fluid remains in the interstitial compartment. This excess fluid is removed from the interstitium by the lymphatic system and returned to the systemic circulation.

Capillary filtration refers to the movement of water through capillary pores because of a mechanical, rather than an osmotic, force. The capillary filtration pressure, sometimes called the *capillary (or blood) hydrostatic pressure,* is the pressure pushing water out of the capillary into the interstitial spaces.

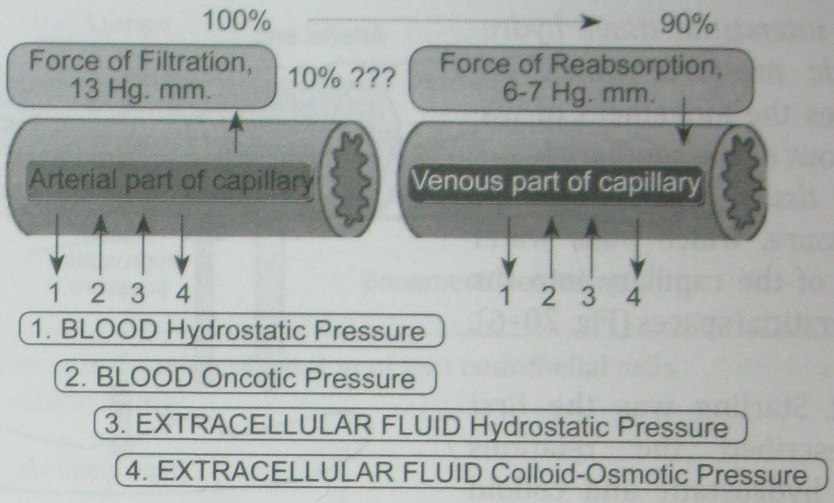


Fig. 20–7. The principal scheme of the microcirculation.

The *capillary (or blood) colloidal osmotic pressure* is the osmotic pressure generated by the plasma proteins that are too large to pass through the pores of the capillary wall. The term *colloidal osmotic pressure* differentiates this type of osmotic pressure from the osmotic pressure that develops at the cell membrane from the presence of electrolytes and non-electrolytes. Because plasma proteins do not normally penetrate the capillary pores and because their concentration is greater in the plasma than in the interstitial fluids, it is capillary colloidal osmotic pressure that pulls fluids back into the capillary.

The *interstitial (or extracellular) fluid pressure* and the tissue *(or extracellular fluid)* colloidal osmotic pressure contribute to movement of water into and out of the interstitial spaces. The interstitial fluid pressure opposes the outward movement of water from the capillary into the interstitial spaces. The tissue colloidal osmotic pressure pulls water out of the capillary into the tissue spaces. It reflects the small amount of plasma proteins that normally escape from the capillary to enter the interstitial spaces.

The *Force of Filtration* develops in the arterial part of the capillary as the result of these four forces action on the capillary and blood solvents (Fig. 20–7, the red arrow). The absolute value of the force of filtration equals to 13 Hg mm and it determines 100 % of all the blood plasma solvents to penetrate the capillary walls and enter the extracellular fluid.

The resulting force on the venous end of the capillary that is the mathematical sum of the mentioned four forces action is the *Force of Reabsorption*

(Fig. 20–7, the blue arrow). The absolute value of the force of reabsorption is within the range 6–7 Hg mm and it determines 90 % of all the previously filtrated blood plasma solvents reabsorption.

The question arises what is the fate of the rest 10 % of the solvents which were filtrated into the interstitial fluid but not reabsorbed. We will discuss this below while explaining the physiological peculiarities of the lymphatic circulation.

3. DIFFUSION

Diffusion across a membrane occurs in both directions. The net diffusion flux (Q), in moles per min, is given by the Fick's law of diffusion:

$$Q = DA(\Delta C) / x$$

where D is the diffusion coefficient of a substance through the membrane,

A is the membrane area,

ΔC is the concentration difference across the membrane (plasma vs. interstitial fluid) and x is the membrane thickness.

The parameters on the right side of the equation are often grouped: (ΔC)/x is the concentration gradient, D/x is referred to as permeability and, DA/x is the permeability-surface area product.

The endothelial cell membrane is lipid in nature, and substances with a high lipid solubility, such as O_2 and CO_2, have a high diffusion coefficients *through the endothelium.* This also means that the entire endothelial cell membrane area is available for diffusion. Therefore, a large diffusion flux occurs in response to a given concentration gradient.

Water soluble materials generally cannot diffuse through the membrane proper. Studies on the transcapillary transport of water soluble materials up to a molecular dimension of approximately 4 nm indicate that their diffusion takes place through *aqueous channels* (on the order of 10 nm) occupying a very small fraction of the endothelial surface (~0.1 % in the skeletal muscle and ~1 % in the kidney). This small area available for diffusion leads to a smaller flux of water soluble substances in comparison to soluble *materials.*

The intercellular gap junction is the ultrastructural basis for these small aqueous channels or pores. The geometric relationship between the diffusing molecule and the channel is such that diffusion is severely restricted when the

molecular dimension approaches *Vi* of the channel size. Therefore, the fact that protein molecules larger than 5 nm can diffuse across the capillary wall indicates the presence of a *large pore system.*

The ultrastructural basis of the large pore system is probably the plasmalemmal vesicles and the intercellular leaky junctions which are associated with endothelial cell turn-over processes. Because of the much higher protein concentration in the plasma, the large pore transport mechanism brings about a net transfer from the plasma to the interstitial fluid.

The above discussions apply primarily to capillaries with continuous endothelium. In capillaries with discontinuous endothelium or open-fenestrated endothelium, the wide intercellular space or open fenestration serve as the major large pore pathway for protein transport

Physiological regulation of transcapillary diffusion is achieved mainly by controlling the available capillary surface area In many tissues a significant fraction of the capillaries are unperfused in the normal condition. When there is an increased metabolic demand,.the previously closed capillaries can be opened by relaxation of the terminal arterioles or the precapillary sphincters.

4. FILTRATION AND REABSORPTION

The role of hydrostatic and colloid osmotic pressures in determining fluid movement across capillaries was first postulated by the English physiologist Ernest Starling at the end of the nineteenth century. As formulated in Starling's hypothesis, the fluid filtered across a capillary membrane is proportional to the net filtration pressure which is a combination of the fluid mechanical pressure (usually referred to as hydrostatic pressure, but actually is the sum of hydrostatic and hydrodynamic pressures) and the colloidal osmotic pressure.

The rate of fluid filtration (V, in ml/min) can be expressed as

$$\mathbf{Vf \cdot kf \cdot [(Pc - Pif) - (\pi c - \pi if)],}$$

where k, is the filtration coefficient,

P_c the capillary pressure,

P_{tf} the interstitial fluid pressure,

n, the colloidal osmotic pressure in plasma,

and π_{if} is the colloidal osmotic pressure in interstitial fluid.

V_r is positive for outward filtration from capillary to interstitial fluid and negative for inward reabsorption in the opposite direction.

The above equation is analogous to the Poiseuille-Hagen law, and k, is the reciprocal of filtration resistance, or the hydraulic conductance. Filtration occurs mainly through the intercellular gap junctions of the small pore system. Hence k, varies directly with the number of these junctions with the fourth power of the pore radius, and inversely with the pore length. It follows that k, is much higher in organs with discontinuous endothelium, e.g., the liver, than in structures with continuous endothelium, e.g. skeletal muscle.

The capillary pressure decreases continuously from the arteriolar end to the venous end. The higher pressure in the arteriolar capillaries favors filtration, whereas the lower pressure in the venous capillaries tends to cause fluid reabsorption. For the sake of simplicity, we shall examine the mean capillary pressure (P_c), equivalent to the pressure value somewhere in the middle of the capillary network, as a function of the arterial pressure (Pa), venous pressure (Pv), precapillary resistance (Ra) and postcapillary resistance (Rv).

It should be mentioned that the blood flow (V) through a region is given by

$$\mathbf{V = (Pa - Pc)/Ra = (Pc - Pv)/Rv}$$

Normally Pa — Pc is approximately 4 times the value of Pc — Pv in most tissues, hence Ra is also approximately 4 times the value of Rv.

The capillary exchange of fluid is bidirectional because capillaries and venules may filter or absorb fluid, depending on the balance of hydrostatic and colloid osmotic pressures. It is possible that filtration occurs primarily at the arteriolar end of capillaries, where filtration forces exceed absorptive forces. It is equally likely that fluid absorption occurs in the venous end of the capillary and small venules because the friction of blood flow in the capillary has dissipated the hydrostatic blood pressure. Based on directly measured capillary hydrostatic and plasma colloid osmotic pressures, the entire length of the capillaries in skeletal muscle filters slightly all of the time, while the lower capillary pressures in the intestinal mucosa and brain primarily favor absorption along the entire capillary length. However, as each of these organs does filter fluid, some of the capillaries and, probably, the smaller arterioles are filtering fluid most of the time.

The extrapolation of fluid filtration or absorption for a single capillary to fluid exchange in a whole tissue is difficult. Within organs, there are regio-

nal variations in microvascular pressures, possible filtration and absorption of fluid in vessels other than capillaries, and physiologically and pathologically induced variations in the available surface area for capillary exchange. Therefore, for whole organs, a measurement of total fluid movement relative to the mass of the tissue is used.

To take into account the various hydraulic conductivities and total surface areas of all vessels involved, the volume (ml) of fluid moved per minute for a change of 1 mm Hg in capillary pressure for each 100 g of tissue is determined. This value is called the *capillary filtration coefficient* (CFC), although it is likely that fluid exchange occurs in both venules and capillaries. CFC values in tissues such as skeletal muscle and the small intestine are typically in the range of 0.025 to 0.16 ml/min per mm Hg per 100 g.

The hydrostatic and colloid osmotic pressure differences across capillary walls — the **Starling forces** — cause the movement of water and dissolved solutes into the interstitial spaces. These movements are, however, normally quite small and contribute minimally to tissue nutrition. Most solutes transferred to the tissues move across capillary walls by simple diffusion, not by bulk flow of fluid.

5. LYMPHATIC SYSTEM

The lymphatic system, commonly called the *lymphatics,* serves almost all body tissues, except cartilage, bone, epithelial tissue, and tissues of the central nervous system. However, most of these tissues have prelymphatic channels that eventually flow into areas supplied by the lymphatics. Lymph is derived from interstitial fluids that flow through the lymph channels. It contains plasma proteins and other osmotically active particles that rely on the lymphatics for movement back into the circulatory system.

The lymphatic system begins as blind sacs in the interstitial space (Fig. 20–8). The endothelium of lymphatic capillaries is discontinuous, and the basement membrane is not prominent and can even be absent altogether. Proteins which have escaped from the blood capillaries into interstitial space can enter the lymphatic capillaries. The larger lymphatic vessels do not have significant permeability for proteins.

Therefore, the proteins are returned with water and other solutes via the lymphatic system (thoracic duct and other lymphatic ducts) into the large veins. The lymph protein concentration in different regions varies with the

blood capillary permeability to proteins; it is high in regions with discontinuous endothelium and low in areas with continuous endothelium.

We remind now the question aroused while the mechanisms of the blood plasma solvents filtration and reabsorption explaining (see page 514, Fig. 20–7). The question concerned the nearest outcome of those 10 % of the solvents that were filtrated out the arterial part of the capillary and were not reabsorbed in the venous part of the capillary. The answer is the following — these 10 % of the solvents are entering the lymphatic circulation.

The total lymph flow is probably over 3 liters per day, and approximately 100 gm of proteins pass through the lymphatic circulation per day to be returned to the systemic circulation. Therefore, as indicated in the Fig. 20–9, lymphatic circulation provides a circulatory path for fluids and solutes (including proteins) from capillaries through interstitial space to the large veins.

The lymphatic system is also the main route for absorption of nutrients, particularly fats, from the gastroin-

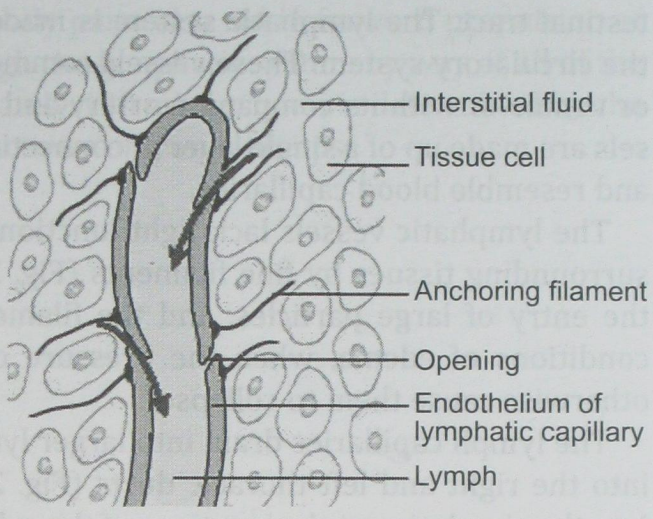


Fig. 20–8. Lymphatic capillary ultrastructure.

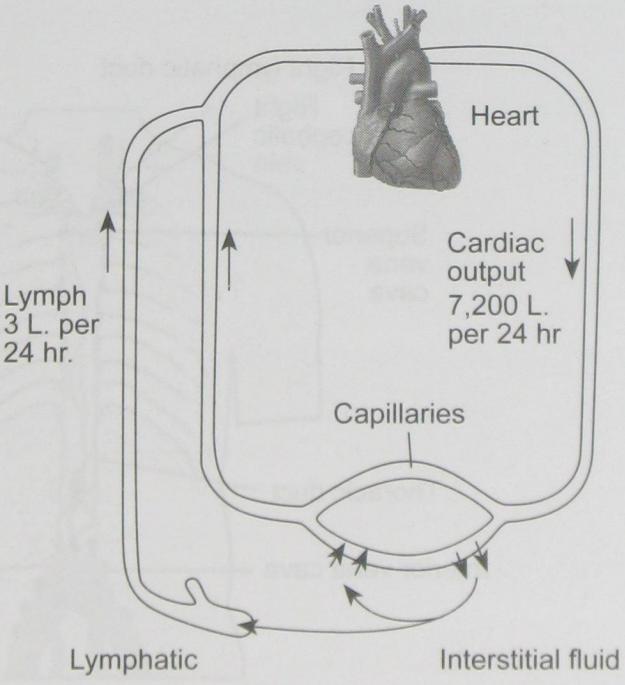


Fig. 20–9. The scheme of lymphatic system drainage function.

testinal tract. The lymphatic system is made up of vessels similar to those of the circulatory system. These vessels commonly travel along with an arteriole or venule or with its companion artery and vein. The terminal lymphatic vessels are made up of a single layer of connective tissue with an endothelial lining and resemble blood capillaries.

The lymphatic vessels lack tight junctions and are loosely anchored to the surrounding tissues by fine filaments (Fig. 20–8). The loose junctions permit the entry of large particles, and the filaments hold the vessels open under conditions of edema, when the pressure of the surrounding tissues would otherwise cause them to collapse.

The lymph capillaries drain into larger lymph vessels that ultimately empty into the right and left thoracic ducts (Fig. 20–10). The thoracic ducts empty into the circulation at the junctions of the subclavian and internal jugular veins.

Although the divisions are not as distinct as in the circulatory system, the larger lymph vessels show evidence of having intimal, medial, and adventitial layers similar to those of blood vessels. Contraction of the smooth muscle in the medial layer of the larger collecting lymph channels assists in propelling lymph fluid toward the thorax. External compression of the lymph channels by

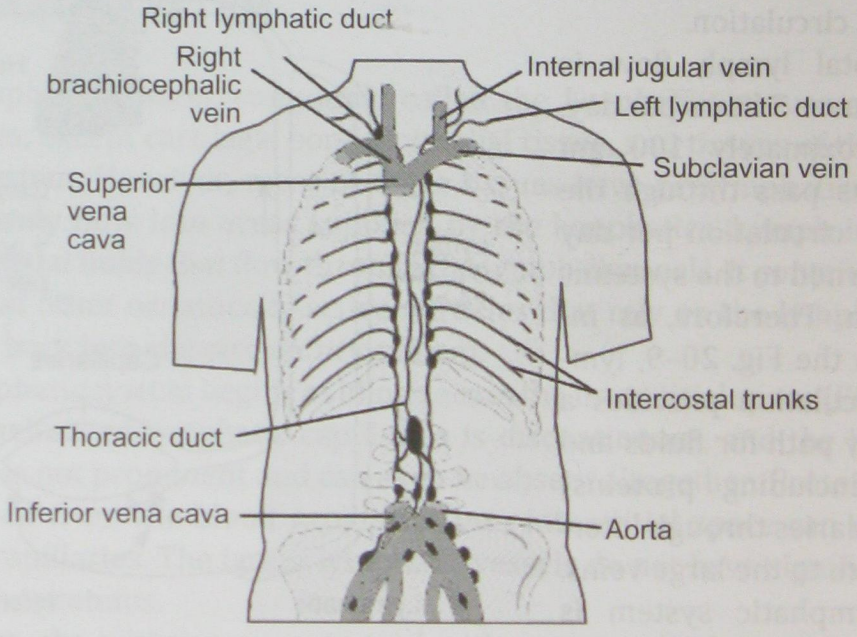


Fig. 20–10. Lymphatic system scheme with the indication of thoracic duct and both right and left lymphatic ducts localizations.

active and passive movements of body parts also aid in forward propulsion of lymph fluid. The rate of flow through the lymphatic system by way of all of the various lymph channels, approximately 120 ml per hour, is determined by the interstitial fluid pressure and the activity of lymph pumps.

Chapter 21. Arterial pressure. Neurohumoral regulation of circulatory system functional activity

The mechanisms controlling the circulation can be divided into *neural, hormonal* and *local* control mechanisms. Cardiac performance and vascular tone at any time are the result of the integration of all three control mechanisms. To some extent, this categorization is artificial because each of the three categories affects the other two. To a certain degree, this chapter may be understood easily after Chapter # 18.

1. FACTORS REGULATING ARTERIAL PRESSURE, CARDIAC OUTPUT AND PERIPHERAL RESISTANCE

1.1. The Arterial Pressure

Measurement of Arterial Pressure. The systemic arterial blood pressure rises to a maximum (*the systolic pressure*), during the ejection period, while it falls to a minimum (*the diastolic pressure*) during diastole and the iso-(volu)-metric period of systole (aortic valve closed; Fig. 21–1).

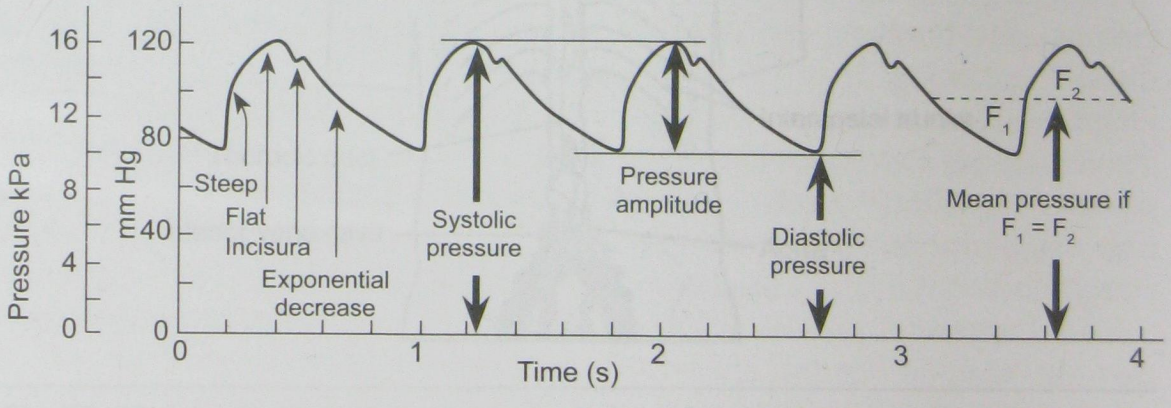


Fig. 21–1. Aortic pressure curve (direct measurement).

The arterial pressure (AP) can be determined *directly* by introducing into a systemic artery a needle or a catheter that is connected to a pressure measuring device. The pressure measuring device may be a pressure transducer connected to an amplifier-recorder system which allows the recording of phasic changes in AP with each cardiac cycle, or it may be a mercury manometer which gives only the mean arterial pressure due to damping by the mercury.

The direct method is used in investigations on experimental animals and clinically in intensive care units, in the operating room, and in the cardiac catheterization laboratory. When direct pressure measurement is made by introducing the needle in a direction perpendicular to the flow, one measures the lateral pressure exerted radially on the vessel wall, i.e. the internal pressure which provides the transmural pressure gradient. If the needle or the catheter is pointed in a direction against the flow, then the end pressure measured also includes an additional pressure component due to the conversion of the kinetic energy. In the systemic arteries, however, the kinetic component is generally small, and the end pressure is not appreciably higher than the lateral pressure.

Routine determination of AP in human subjects or patients is performed *indirectly* according to the *Riva-Rocci* method using a sphygmomanometer (Fig. 21–2).

In this method, an inflatable cuff is wrapped around the upper arm, and a stethoscope head is pressed gently against the antecubital fossa for auscultation

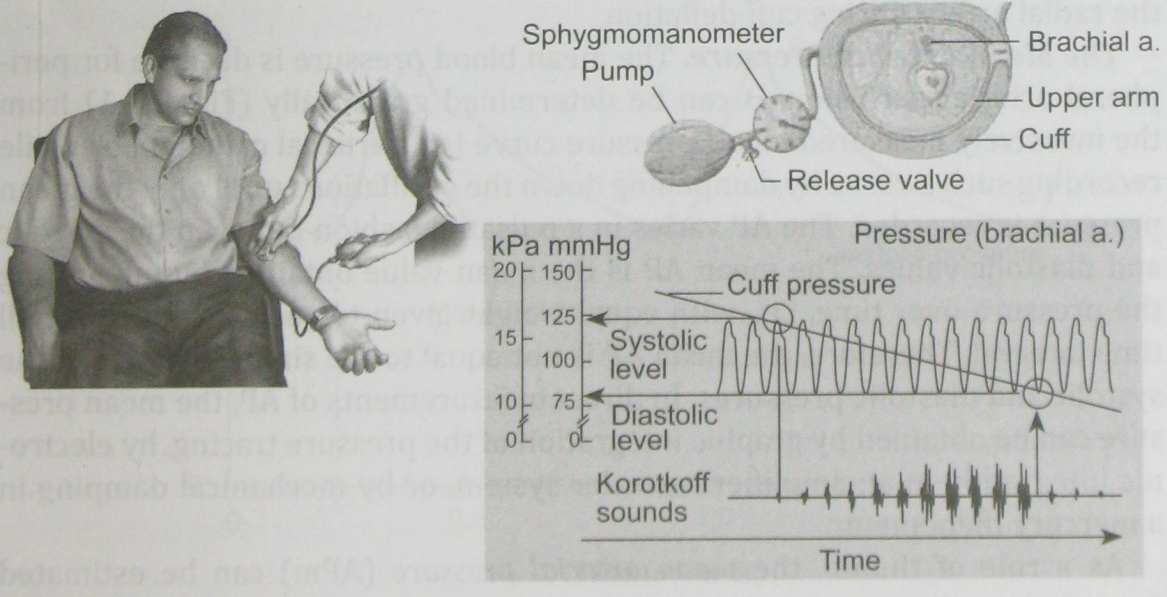


Fig. 21–2. Blood pressure measurement with sphygmomanometer (after Riva-Rocci).

of sounds in the brachial artery (Fig. 21–2). Before inflation of the cuff, blood flow through the artery is laminar and, therefore, there is no audible sound. When the cuff pressure is raised appr. 10–15 Hg mm above the systolic pressure, there is total cessation of flow through the arm (palpated radial pulse disappears) and again there is no sound. As the cuff is gradually deflated, a sharp, tapping sound can be heard when the cuff pressure is reduced to the level of the systolic pressure, thus allowing the occluded artery to open briefly. The occurrence of the first pulse-synchronous sound (clear, tapping sound; the first tone of Korotkov) represents systolic pressure and is recorded. Therefore, the reading of the manometer at the *first occurrence of this sound corresponds to the systolic pressure.*

Normally, this sound at first becomes softer before getting louder, then becomes muffled and disappears completely. With progressive reduction in cuff pressure, the sound recurs with each systolic peak, but the tone becomes softer and the volume louder, as the arterial opening becomes wider and stays open longer. It has been found by comparison with direct pressure recording that the sound becomes muffled when the cuff pressure equals the diastolic pressure. A further reduction of cuff pressure (full disappearance of the noise in radial artery, the second tone of Korotkov) to *approximately 5 mm Hg below the diastolic pressure* causes complete disappearance of sound (Fig. 21–2). Instead of using the auscultatory method, the systolic pressure can also be estimated, with lesser accuracy, by observing the onset of oscillation of the mercury column in the manometer or by palpating for the onset of a pulse in the radial artery during cuff deflation.

The Mean Arterial Pressure. The mean blood pressure is decisive for peripheral arterial perfusion. It can be determined graphically (Fig. 21–1) from the invasively measured blood pressure curve (e.g., arterial catheter) or while recording such a curve by dampening down the oscillations until only the mean pressure is recorded. The AP varies in a pulsatile fashion between the systolic and diastolic values. The mean AP is the mean value obtained by integrating the pressure over time, i.e., with equal weight given to each extremely small time interval. Therefore, the mean AP is not equal to the simple average of the systolic and diastolic pressures. In direct measurements of AP, the mean pressure can be obtained by graphic integration of the pressure tracing, by electronic integration in an amplifier-recorder system, or by mechanical damping in a mercury manometer.

As a rule of thumb, the *mean arterial pressure* (APm) can be estimated from the *systolic pressure* (Pa) and *diastolic pressure* (Pd) obtained in indirect measurements by the following equation:

$$APm = 1/3\ Pa + 2/3Pd$$

The greater weight given to Pd reflects the fact that the arterial pressure is closer to Pa for longer periods than it is to Pa (Fig. 21–3).

Pressure Change Along the Arterial Tree. The mean AP decreases progressively from the aorta toward the peripheral branches, reflecting the energy dissipation of flow. In comparison to the root of the aorta the large and medium-sized arteries have a slightly lower *mean pressure,* but a slightly higher systolic pressure. The higher systolic pressure in the large and medium arteries is best explained by the reflection of pressure waves from the small arteries. With each systole, the arterial wall propagates a pressure wave peripherally at a rate (approximately 500 cm/sec) much faster than the velocity of blood (approximately 20 cm/sec). The pressure wave cannot be propagated easily through the high resistance small arteries and is thus reflected back in a retrograde fashion. As a result of the summation of the reflected wave with the forward wave, the systolic pressure in the more centrally located medium-sized arteries is enhanced. In contrast, the diastolic pressure shows a continuous decrease from the aorta toward the periphery.

Therefore, the *pulse pressure,* which is the difference between systolic and diastolic pressure, gradually increases from the aorta toward the medium-sized arteries, even as the *mean pressure* declines (Fig. 21–4).

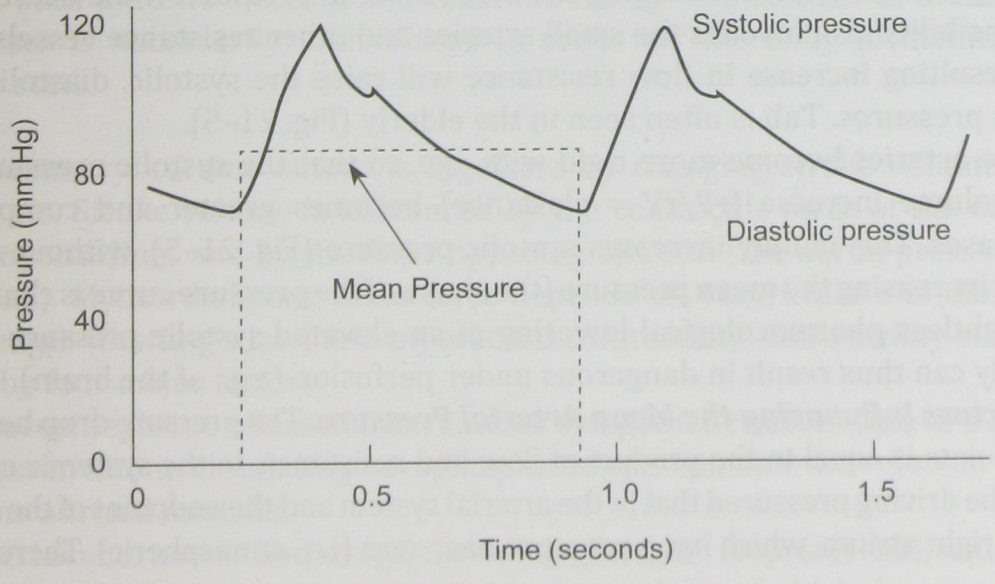


Fig. 21–3. Graphical illustration of the mean AP.

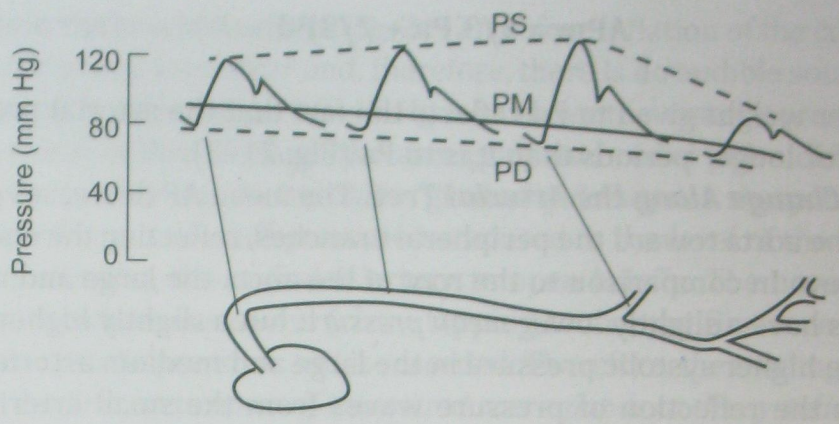


Fig. 21–4. Graphical illustration of the pulse pressure.

Influence of Arterial Capacitance on Arterial Pressure. Due to the capacitance of the arterial tree and to the peripheral resistance to flow, a portion Of the blood ejected during systole is stored in the arterial tree and gradually released during diastole.

The normal capacitance of the arterial wall serves to minimize the rise of pressure during systole as well as the fall of pressure during diastole. A decrease in arterial capacitance, therefore, would cause an increase in systolic pressure and a decrease in diastolic pressure, leading to a widening of the pulse pressure without changing the mean arterial pressure. If the decrease in distensibility also invokes the small arteries and other resistance vessels, then the resulting increase in flow resistance will raise the systolic, diastolic, and mean pressures. Tub is often seen in the elderly (Fig. 21–5).

The arteries become more rigid with age, so that the systolic pressure rise per volume increase (rP/rV = elastance) becomes greater and compliance decreases. This mainly increases systolic pressure (Fig. 21–5), without necessarily increasing the mean pressure (the shape of the pressure curve is changed). Thoughtless pharmacological lowering of an elevated systolic pressure in the elderly can thus result in dangerous under perfusion (e.g., of the brain).

Factors Influencing the Mean Arterial Pressure. The pressure drop between two points is equal to the product of flow and resistance. In the systemic circulation, the driving pressured that in the arterial system and the endpoint of the circuit is the right atrium, which has a pressure near zero (i.e., atmospheric). Therefore,

mean arterial pressure = cardiac output · total peripheral resistance,

where the total peripheral resistance (TPR) is the sum total of all resistances in the systemic circulation. The factors regulating cardiac output and TPR will be discussed later.

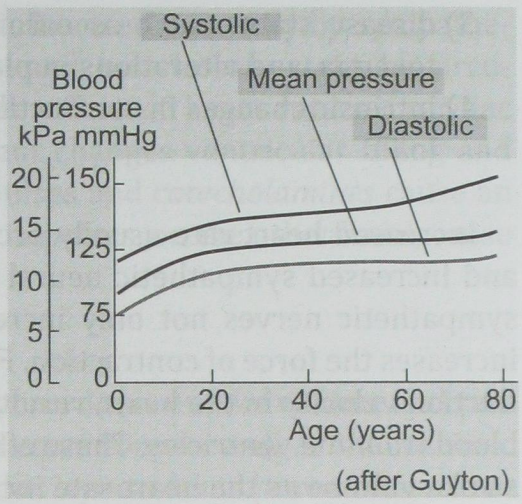


Fig. 21–5. Effects of aging on arterial pressure.

1.2. Factors Regulating Cardiac Output

Cardiac output equals to the product of *heart rate* and *stroke volume*, which in turn is equal to the difference between the end-diastolic volume (EDV) and the end-systolic volume (ESV) of the ventricle.

Cardiac output = Heart rate · Stroke volume

Heart rate · (EDV — ESV)

In each cardiac cycle, systolic ejection causes the ventricular volume to be reduced from EDV to ESV, and diastolic filling causes the ventricular volume to rise from ESV to EDV. Regulatory mechanisms exist for the control of both systolic ejection and diastolic filling. The control of these two processes is necessarily interrelated in order for the body to maintain a steady state.

In summary, cardiac output is regulated by changing stroke volume and heart rate. Stroke volume is influenced by the contractile force of the ventricular myocardium and by the force opposing ejection (the aortic pressure or afterload). Myocardial contractile force depends on ventricular end-diastolic fiber length (Starling's law) and myocardial contractility. Contractility is influenced by four major factors:

1) norepinephrine released from cardiac sympathetic nerves and, to a much lesser extent, circulating norepinephrine and epinephrine released from the adrenal medulla;
2) certain hormones and drugs, including glucagon, isoproterenol, and digitalis (which increase contractility) and anesthetics (which decrease contractility);

3) disease states, such as coronary artery disease, myocarditis, bacterial toxemia, and alterations in plasma electrolytes and acid-base balance;
4) intrinsic changes in contractility with changes in heart rate and/or afterload.

Increased heart rate usually occurs because of decreased parasympathetic and increased sympathetic neural activity. The release of norepinephrine by sympathetic nerves not only increases the heart rate but also dramatically increases the force of contraction. Furthermore, norepinephrine increases conduction velocity in the heart, resulting in a more efficient and rapid ejection of blood from the ventricles. These effects, summarized in Fig. 21–6, maintain the stroke volume as the heart rate increases.

When the heart rate increases physiologically as a result of an increase in sympathetic nervous system activity (as during exercise), cardiac output increases proportionately over a broad range.

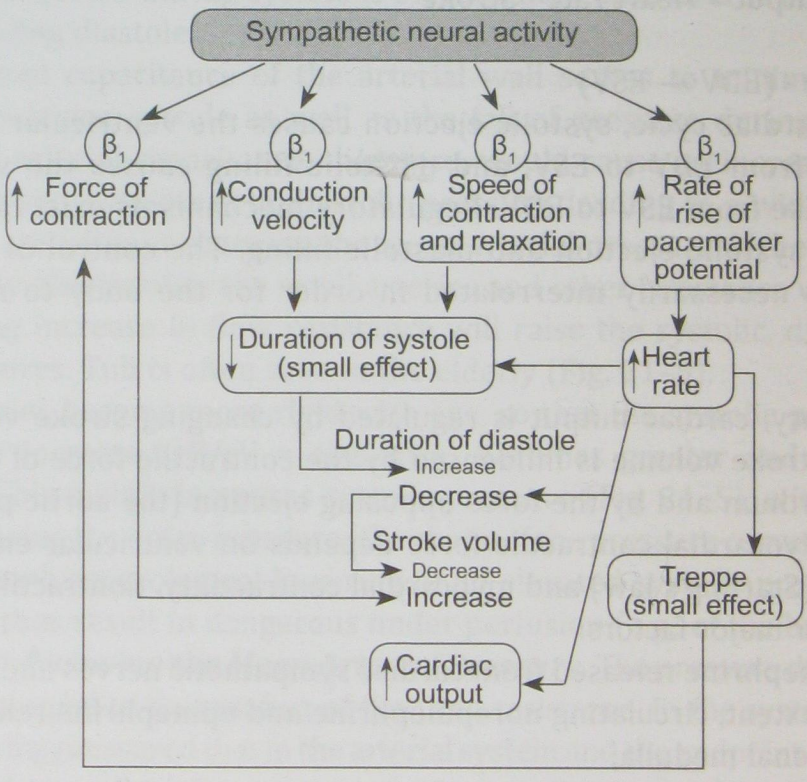


Fig. 21–6. Effects of increased sympathetic neural activity on heart rate, stroke volume, and cardiac output.

Regulation of Heart Rate. Heart rate is influenced primarily by sympathetic and parasympathetic nerves to the heart and, by a lesser extent, by circulating norepinephrine and epinephrine. The effect of heart rate on cardiac output depends on the extent of concomitant changes ventricular filling and contractility. *Cardiac sympathetic nerve* impulses and *catecholamines* cause an acceleration of heart rate (Fig. 21–6). *Vagal* impulses, on the other hand, cause cardiac slowing.

A. Regulation of Systolic Ejection

The force generated by the ventricular myocardium and the volume and rate of ejection depend on the *external working conditions* as well as the *intrinsic contractility*. The intrinsic contractility of the myocardium (force of contraction dependent on the intensity and duration of the active state) is enhanced by sympathetic nerve impulses and catecholamines. The influences of external working conditions serve to adjust the systolic ejection in response to alterations in peripheral circulatory dynamics. Thus, the increase in force of myocardial "contraction consequent to an increase in EDV (Starling's law) allows an adjustment of ventricular output in accordance with variations in return from the venous system. The reduction in the rate of systolic ejection following elevation of aortic pressure leads to a lowering of the cardiac output in response to a rise in peripheral resistance, thus maintaining a relatively constant arterial pressure.

B. Regulation of Diastolic Filling

The volume of blood filling the ventricle from end-systolic volume (ESV) to end-diastolic volume (EDV) in each cycle is a function of the *duration of filling* (diastolic interval) and the *rate of filling* (venous return).

B.1. Diastolic interval. The diastolic interval is curtailed by an increase in heart rate, especially when the heart rate is faster than 100 beats per min, and this leads to a decrease in diastolic filling (decrease of EDV). When cardioacceleration is brought about by sympathetic stimulation, however, there is generally a concomitant stimulation of myocardial contractility, which enhances systolic ejection (decrease of ESV). Therefore, when the cardiac sympathetic nerves are activated, e.g., during muscular exercise, the decreases of both EDV and ESV result in a stroke volume which is essentially unchanged or slightly increased, and the cardiac output increases mainly due to the acceleration in heart rate.

B.2. Relation between blood volume and vascular capacity. The rate of diastolic filling via venous return is determined by the *fullness of circulatory system*, the *right atrial pressure*, and the *resistance in the venous return path.* The fullness of the circulation refers to the relation between blood volume and vascular capacity, especially that of the veins, for a given level of vascular capacity, the fullness of the circulation and the rate of venous return vary directly with the volume of blood. For a given blood volume, the fullness of circulation varies inversely with vascular capacity. Thus an increase in venous capacity would raise the proportion of blood volume trapped in the peripheral veins and reduce the venous return.

Sympathetic nerve impulses and catecholamines cause a reduction in venous capacity by causing contraction of venous smooth muscle thus facilitating venous return. The degree of venous distention is affected not only by the tension in the venous smooth muscle, but also by mechanical factors originating extravascularly and intravascular.

The influence of the extravascular factor can be seen in the venous system when skeletal muscle undergoes exercise. Veins in the leg contain valves which permit only centripetal flow. When the leg muscles contract, the compression on the vein causes blood to flow out of the compressed segment towards the heart and therefore the venous return is increased (Fig. 21–7). When the leg muscles relax, the decompression of the vein allows blood to enter this segment from the more distal segments. Thus, rhythmic exercise of the leg muscle serves as an effective *muscle pump* to facilitate venous return to the right atrium.

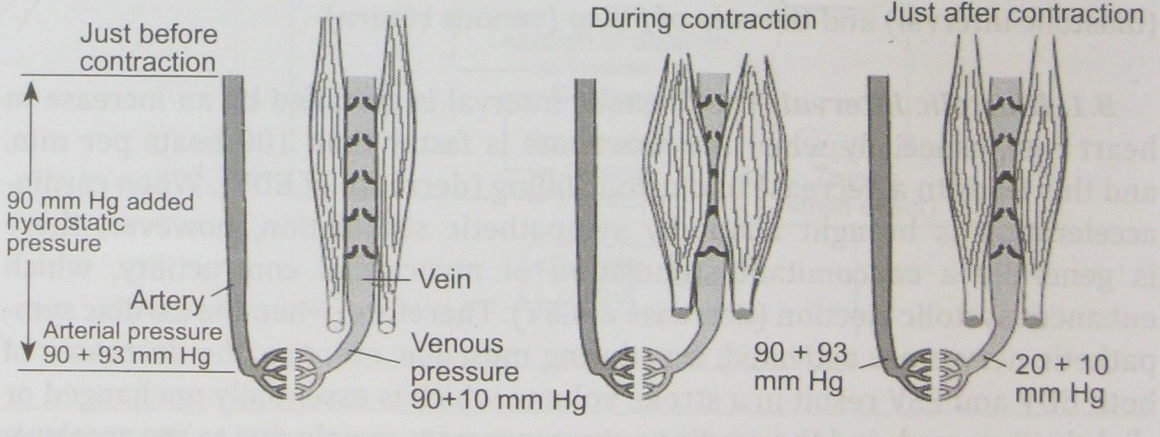


Fig. 21–7. Muscle pump.

This mechanism shifts blood volume from the legs to the central blood volume, and end-diastolic volume is increased. Even mild exercise, such as walking, returns the central blood volume and stroke volume to recumbent values (Fig. 21–8). The center section in the diagram shows the effects of a shift from the prone to the upright position with quiet standing. The right panel shows the effect of activating the muscle pump by contracting leg muscles. Note that the muscle pump restores central blood volume and cardiac output to the levels in the prone position. The fall in heart rate and rise in peripheral blood flow (forearm, splanchnic, and renal) associated with activation of the muscle pump reflect the reduction in baroreceptor reflex activity associated with increased cardiac output (Fig. 21–8).

Similarly, rhythmic change in the intrathoracic pressure with the respiratory cycle functions as a *respiratory pump,* enhancing venous return and restoring the central blood volume during inspiration (Fig. 21–9). Contraction of the diaphragm simultaneously raises intraabdominal pressure which compresses the abdominal veins. Because the venous valves prevent the back-

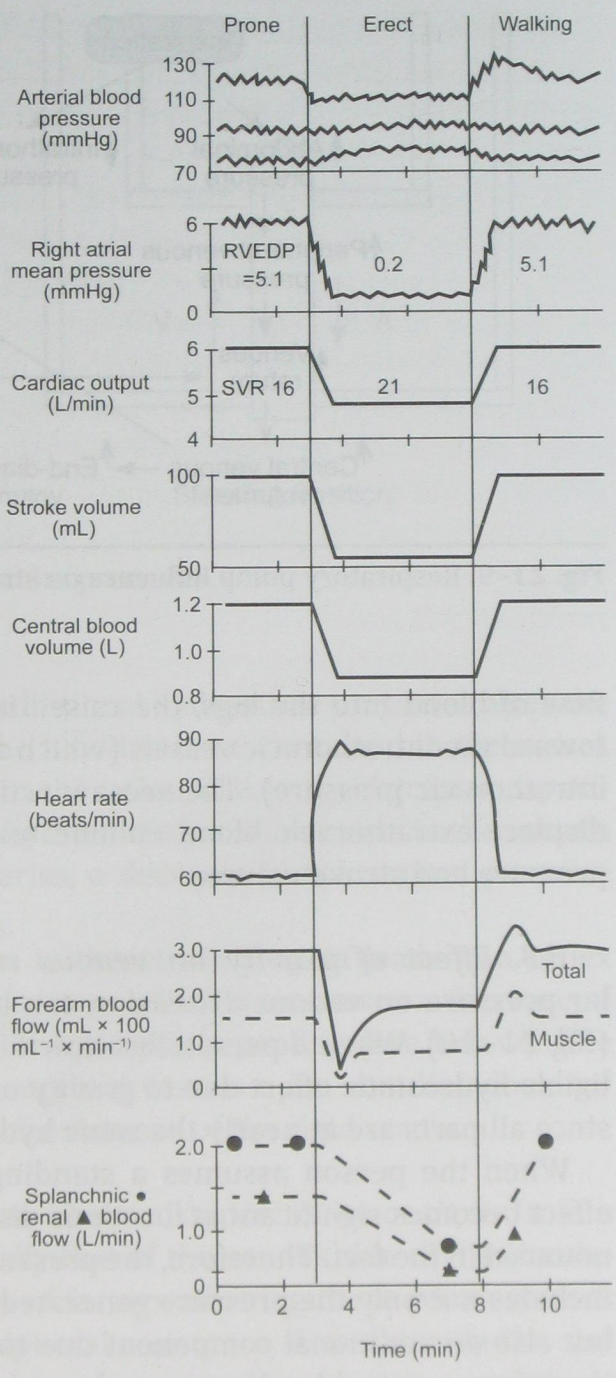


Fig. 21–8. Muscle pump influence on central blood volume and systemic hemodynamics. Abbreviation: RVEDP — right ventricular end-diastolic pressure; SVR — systemic vascular resistance.

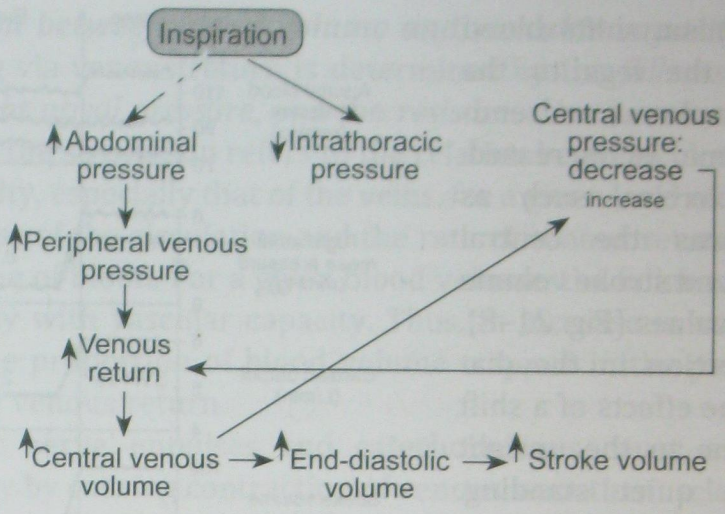


Fig. 21–9. Respiratory pump influence on stroke volume and venous return indexes.

flow of blood into the legs, the raised intraabdominal pressure forces blood toward the intrathoracic vessels (which are expanding because of the lowered intrathoracic pressure). The seesaw action of the respiratory pump tends to displace extrathoracic blood volume toward the chest and raise right atrial pressure and stroke volume.

B.3. Effect of gravity on venous return. The influence of intravascular pressure on venous distension can be illustrated by the effect of gravity (Fig. 21–10). When a person lies down in a horizontal position, there is negligible hydrostatic effect due to gravity on the various parts of the circulation, since all parts are at nearly the same hydrotatic level as the heart.

When the person assumes a standing position, however, the hydrostatic effect becomes significant at locations distant from the heart. This is most pronounced in the feet. Therefore, the pressure measured in each vessel of the foot includes not only the pressure generated by the pumping activity of the heart, but also an additional component due to the hydrostatic pressure exerted by the column of the blood between the right atrium and the foot This hydrostatic pressure varies with height and is approximately 100 mm Hg.

Therefore, upon assumption of the erect position, the mean arterial pressure in the foot may rise from 90 to 190 mm Hg and the venous pressure from 10 to 110 mm Hg. If the vessels were made of rigid tubes, such hydrostatic influences would have no effect on venous return, as in siphoning. Because the

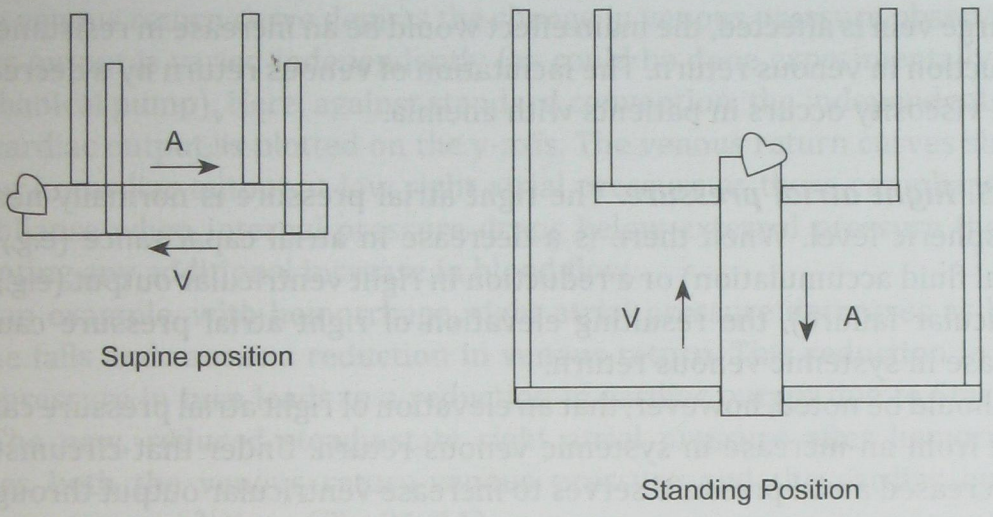


Fig. 21–10. Effect of gravity on cardiovascular pressures.

vessels are distensible, however, the additional hydrostatic pressure raises the transmural pressure and increases the blood volume contained in the vessel. This effect is especially pronounced in the veins, due to their greater capacitance than the arteries. As a negative hydrostatic pressure exists in the vessels above the heart, e.g., the carotid arteries, a decrease in transmural pressure occurs there upon standing.

A decrease of transmural pressure in the carotid region leads to a reflex stimulation of the sympathetic nerve system which then causes a decrease of venous capacitance and a reduction of hydrostatic pooling (see below). The negative hydrostatic pressure in the cerebral vessels upon standing does not lead to vascular collapse because of the existence of the rigid cranium.

B. 4. Resistance to venous return. The resistance to venous return is a function of *vascular geometry*, especially that of the venous side of the circulation, and *blood viscosity*. A discussion of factors affecting these parameters is given in the next section on total peripheral resistance. A decrease in the size of the veins may reduce the capacitance and increase the resistance, and these two effects have opposite influences on venous return.

As a rule, moderate increases in the tension of venous smooth muscle by sympathetic stimulation primarily decreases the capacitance without a marked increase in resistance, thus facilitating the venous return. On the other hand, with a severe constriction of the venous lumen, especially when a local segment

of a large vein is affected, the main effect would be an increase in resistance and a reduction in venous return. The facilitation of venous return by a decrease in blood viscosity occurs in patients with anemia.

B. 5. Right atrial pressure. The right atrial pressure is normally near the atmospheric level. When there is a decrease in atrial capacitance (e.g., pericardial fluid accumulation) or a reduction in right ventricular output (e.g., right ventricular failure), the resulting elevation of right atrial pressure causes a decrease in systemic venous return.

It should be noted, however, that an elevation of right atrial pressure can also result from an increase in systemic venous return. Under that circumstance, the increased atrial pressure serves to increase ventricular output through the increase in diastolic filling (Starling's law).

Thus, the atrial pressure reflects a balance between cardiac output and venous return, and hence it has a pivotal significance in circulatory dynamics. Through the influence on the right atrial pressure, an alteration in either the cardiac output or the venous return can effect a change in the other flow parameter thereby maintaining the balance between cardiac output and venous return and thus the continuity of circulation. In Fig. 21–11 the cardiac function curve illustrates Starling's law in that increases in right atrial pressure will increase ventricular volume and thereby increase stroke volume and cardiac output.

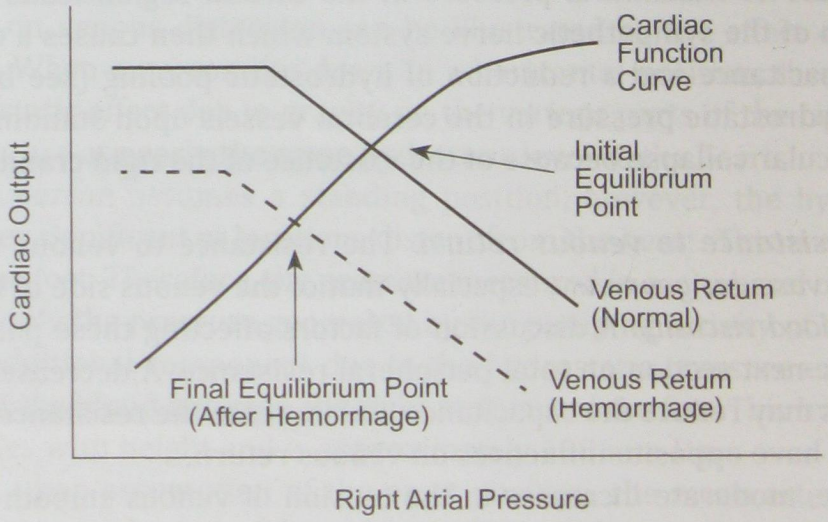


Fig. 21–11. Graphical illustration of right atrial pressure influence on cardiac output, stroke volume and venous return.

The venous return curve depicts the change in venous pressure observed as cardiac output is varied independently (as could be done experimentally with a mechanical pump). Here, against standard convention, the independent variable, cardiac output, is plotted on the y-axis. The venous return curves show a plateau in cardiac output at low right atrial pressure as these compliant vessels collapse when internal pressure drops below external pressure thereby preventing any additional increase in blood flow.

As an example, with hemorrhage, right atrial pressure decreases as blood volume falls and causes a reduction in venous return. This reduction in right atrial pressure in turn leads to a reduction in cardiac output due to Starling's law. The new, reduced steady-state right atrial pressure after hemorrhage satisfies both the venous return-venous pressure and the cardiac output-venous pressure relations (Fig. 21–11).

The factors regulating cardiac output are summarized in Fig. 21–12 together with those factors controlling peripheral resistance.

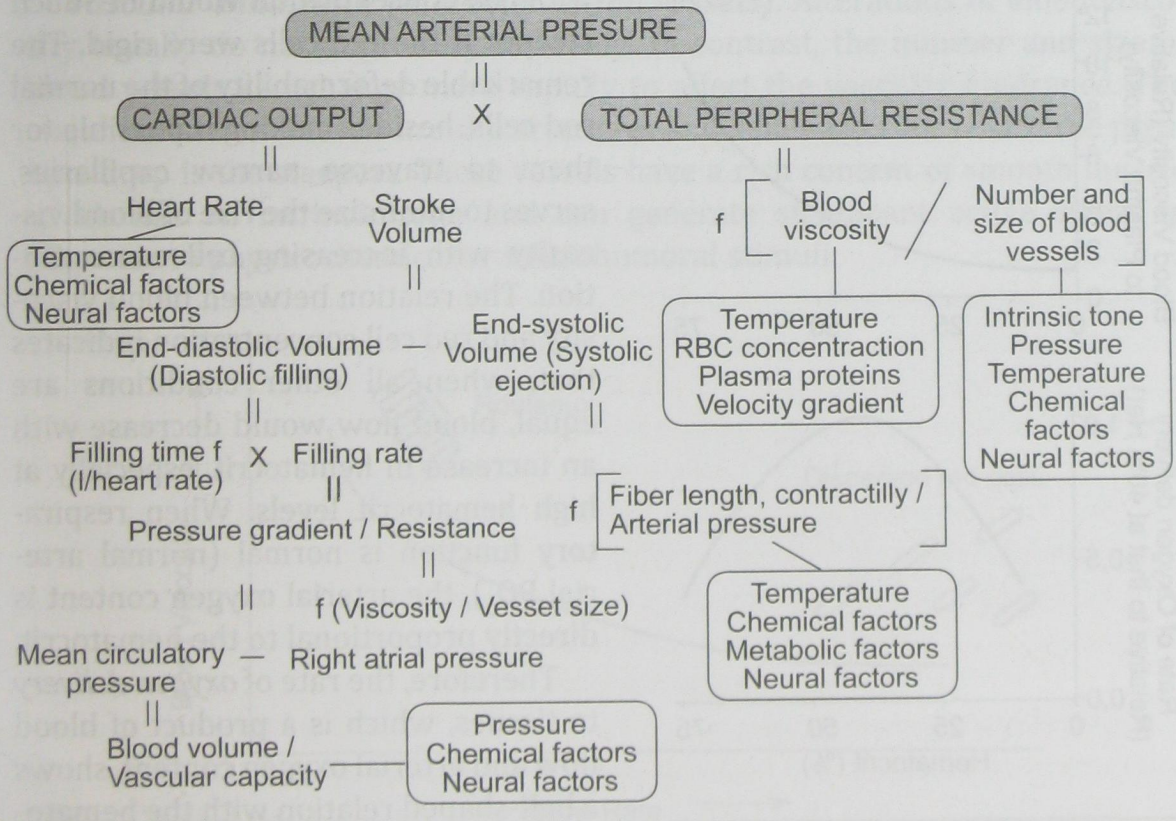


Fig. 21–12. The factors regulating cardiac output and the total peripheral resistance.

1.3. Factors Regulating Peripheral Resistance

Resistance to blood flow is determined by *blood viscosity* (h) and *vascular hindrance*, which varies inversely with the number and sizes of blood vessels. The length of the vessels do not change significantly except during growth. The factors regulating peripheral resistance are summarized in the Fig. 21–12 (above, together with factors controlling cardiac output) and discussed below.

Blood Viscosity. The viscosity of blood, as that of most liquids, varies inversely with *temperature*. Exposure to cold temperature causes an increase in blood viscosity, mainly during flow through the cutaneous circulation.

The *concentration of erythrocytes* in the blood has a strong influence on blood viscosity. At a hematocrit of 40-45 %, blood viscosity is approximately 3 times the value for plasma and approximately 5 times that of water. Blood viscosity shows a curvilinear relation with the hematocrit and it increases sharply when the hematocrit is raised much beyond the normal range (Fig. 22–13, upper panel). This increase in viscosity with cell concentration would be much steeper if the red cells were rigid. The remarkable deformability of the normal red cells, besides making it possible for them to traverse narrow capillaries, serves to minimize the rise of blood viscosity with increasing cell concentration. The relation between blood viscosity and red cell concentration indicates that, when all other conditions are equal, blood flow would decrease with an increase in hematocrit, especially at high hematocrit levels. When respiratory function is normal (normal arterial PO_2), the arterial oxygen content is directly proportional to the hematocrit.

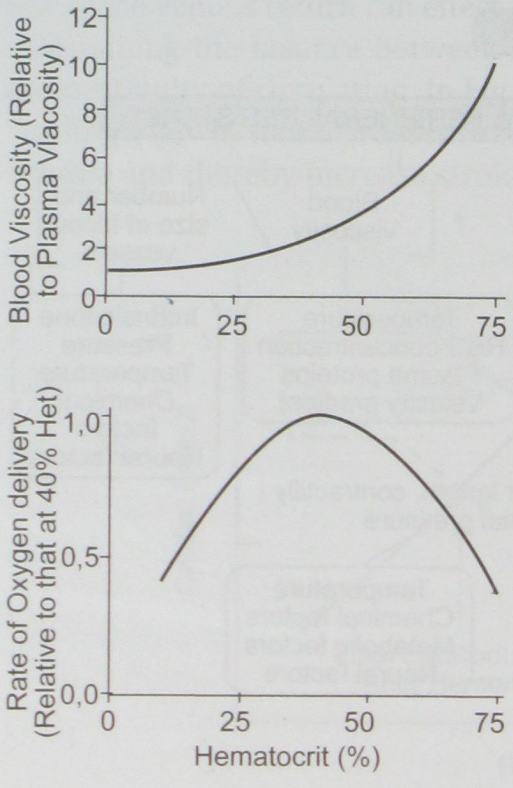


Fig. 21–13. Effect of hematocrit on blood viscosity and oxygen delivery.

Therefore, the rate of *oxygen delivery* to tissues, which is a product of blood flow and arterial oxygen content, shows a bell-shaped relation with the hematocrit value (lower panel of the diagram on the effect of hematocrit variation).

Oxygen delivery is reduced at tow hematocrit levels because of the low arterial oxygen content, and it is reduced at high hematocrit levels due to the increase in blood viscosity. The optimum hematocrit for maximum oxygen delivery is approximately 40 % red cells, i.e. about the normal value. (Fig. 21–13).

An elevation of the concentration of *plasma proteins*, especially fibrinogen and globulins, causes an increase in blood viscosity by two mechanisms. First, the plasma viscosity is raised. Secondly, when the flow rate is slow, fibrinogen and globulins can bridge red cell surfaces to form rouleaux, and such red cell aggregation also caries an increase in blood viscosity. High flow rates are associated as high mechanical shearing force which tends to disperse the aggregates and align the deformed cells, thus lowering the blood viscosity (Fig. 21–14).

Blood viscosity is independent of the size of the blood vessel when the diameter is above approximately 0.5 mm. A decrease of vessel diameter below this value causes a progressive reduction in blood viscosity, and the major reason for this observation is that the blood entering these small vessels tends to have a lower hematocrit than that in the large vessels (the Fahraeus effect).

Vascular Hindrance (diameter of the vessels). Alterations in blood viscosity usually do not occur very suddenly. In contrast, the number and size of blood vessels can change more rapidly to affect the vascular hindrance. The major resistance vessels are the arterioles and small arteries, where the pressure drop is the steepest. These vessels have a rich content of smooth muscle which has an intrinsic tone and can generate significant active stress in response to physicochemical or neurohumoral stimuli.

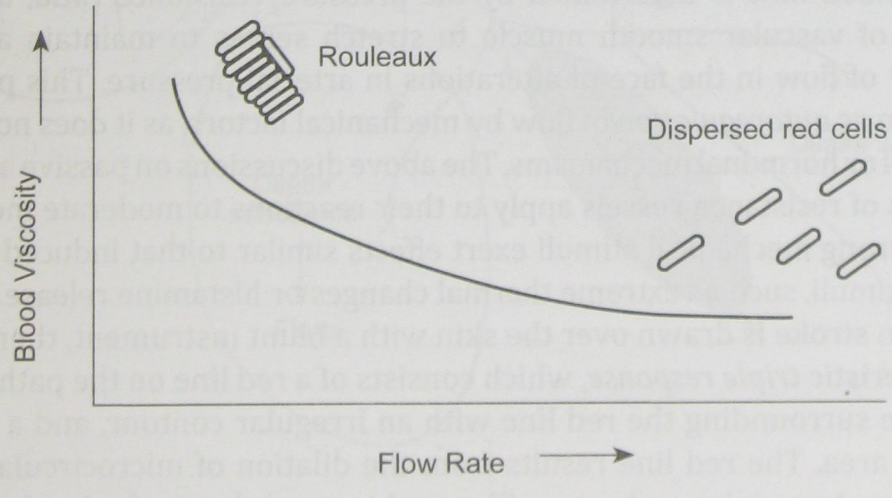


Fig. 21–14. Variation of blood viscosity with the flow rate.

Furthermore, the small ratio of the lumen radius to wall thickness in the resistance vessels can amplify the vasoconstriction due to smooth muscle contraction.

A. Physical Factors Affecting Vascular Hindrance.

A moderate decrease in ambient temperature (e.g. 10 °C) causes vasoconstriction and a moderate increase in temperature (e.g. 35 °C or 45 °C) causes vasodilatation, an effect similar to that induced by other noxious stimuli resulting from mechanical stretching or chemical irritation (see discussion below in relation to stretching). These direct *effects of temperature* on vessel size are important mainly in the cutaneous circulation, whereas the core temperature in the body is normally regulated within narrow limits.

The size of blood vessels may be altered in response to *mechanical factors* originating intravascularly (pressure changes) or extravascularly (compression or stretching). The circumferentially arranged smooth muscle fibers and other fibers in the vessel wall are deformed by an increase in the transmural pressure difference or by external stretching forces. The changes in vessel diameters resulting from the stretch vary in different circulatory beds. Some vessels show mainly a passive response to stretch, and an increase in transmural pressure which causes vasodilatation. Other vessels, e.g., the resistance vessels in the renal and cerebral circulations, however, react to mechanical stretch by an active contraction of the smooth muscle. As a result, an increase in arterial pressure can lead to a vasoconstriction and an increase in resistance.

Since blood flow is determined by the pressure/resistance ratio, the active response of vascular smooth muscle to stretch serves to maintain a relative constancy of flow in the face of alterations in arterial pressure. This process is referred to as *autoregulation* of flow by mechanical factors, as it does not involve any neural or hormonal mechanisms. The above discussions on passive and active responses of resistance vessels apply to their reactions to moderate mechanical stimuli. Strong mechanical stimuli exert effects similar to that induced by other noxious stimuli, such as extreme thermal changes or histamine release.

If a firm stroke is drawn over the skin with a blunt instrument, there results a characteristic *triple response*, which consists of a red line on the path stroked, a red flare surrounding the red line with an irregular contour, and a wheal in the same area. The red line results from the dilation of microcirculative vessels (e.g., metarterioles and precapillary sphincters) due to the local release of histamine and/or bradykinin. The mechanical irritations and/or the chemical

agents released cause a stimulation of some sensory endings located in the line of stroke. These sensory impulses are conducted not only along the main sensory nerve toward the dorsal root and the central nervous system, but also along any collateral branches arising from the main nerve path (Fig. 22–15). A vasodilator substance, probably adenosine triphosphate (ATP), is then released from the terminals of these collateral branches (dorsal root vasodilator fibers) and cause the flare.

Therefore, the flare is the result of a reflex-like neural mechanism called *axon reflex*, which involves only branches of the same axon. The irregular contour of the flare outlines the area of capillaries being filled with blood due to the dilation of feeding arterioles resulting from the axon reflex. Histamine and bradykinin cause an increase in capillary permeability and an escape of plasma proteins and fluid into the extravascular (interstitial) space. This process is facilitated by the increase in capillary pressure due to the dilation of precapillary vessels. The accumulation of such edema fluid in the interstitial space is the basis for wheal formation.

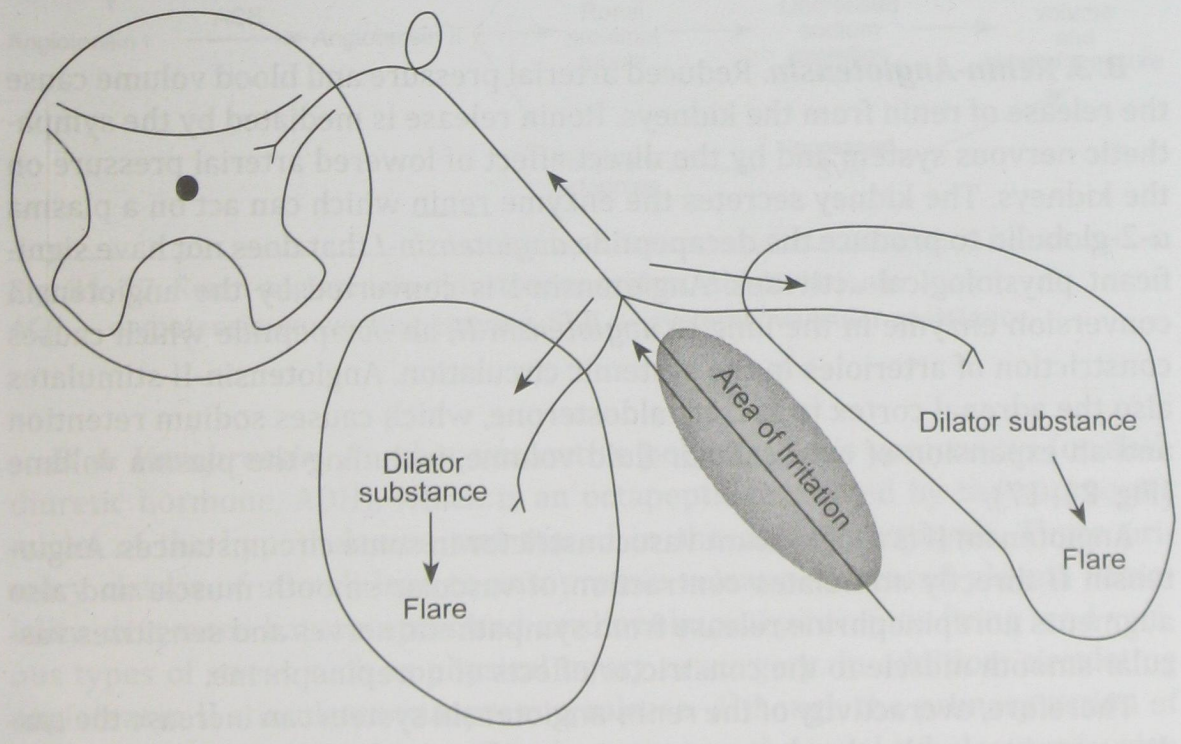


Fig. 21–15. Pathway of the axon reflex.

B. Chemical Factors Affecting Vascular Hindrance

B. 1. Histamine is present in the mast cells in various tissues and is released in inflammatory processes, allergic responses and antigen-antibody reactions. Localized release of histamine due to noxious stimuli leads to the *triple response* in the area stimulated that could be visually recognized as vascular dilation causing *increased redness (erythema), heat (calor)*, movement of fluid into the interstitial spaces (*swelling*) tohether with extravasation of plasma proteins into the extracellular spaces (exudate) and emigration of leukocytes. (Fig. 21–16). Generalized release of large amounts of histamine, e.g., in anaphylactic shock due to serum sickness, may cause severe vasodilatation and a precipitous drop in arterial pressure.

B. 2. Bradykinin. Sweat glands and salivary glands can secrete kallikrein, which is an enzyme acting on an a-2-globulin in plasma or interstitial fluid to produce a vasoactive polypeptide, bradykinin. Bradykinin causes vasodilatation and leads to an increase in capillary permeability. It could induce pain reaction.

B. 3. Renin-Angiotensin. Reduced arterial pressure and blood volume cause the release of renin from the kidneys. Renin release is mediated by the sympathetic nervous system and by the direct effect of lowered arterial pressure on the kidneys. The kidney secretes the enzyme renin which can act on a plasma α-2-globulin to produce the decapeptide *angiotensin-I* that does not have significant physiological activities. Angiotensin-I is converted by the angiotensin conversion enzyme in the lung to *angiotensin-II*, an octapeptide which causes constriction of arterioles in the systemic circulation. Angiotensin-II stimulates also the adrenal cortex to secrete aldosterone, which causes sodium retention and an expansion of extracellular fluid volume, including the plasma volume (Fig. 21–17).

Angiotensin II is a significant vasoconstrictor in some circumstances. Angiotensin II directly stimulates contraction of vascular smooth muscle and also augments norepinephrine release from sympathetic nerves and sensitizes vascular smooth muscle to the constrictor effects of norepinephrine.

Therefore, overactivity of the renin-angiotensin system can increase the cardiac output via blood volume expansion and raise the peripheral resistance by arteriolar constriction. These pathophysiological changes may have considerable significance in the pathogenesis of some forms of hypertension.

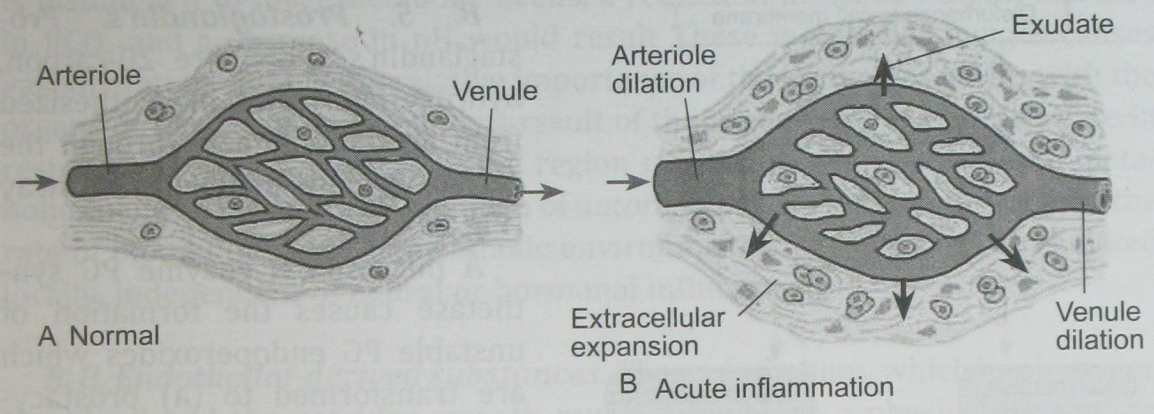


Fig. 21–16. Vascular phase of acute inflammation.
(A) Normal capillary bed. (B) Acute inflammation.

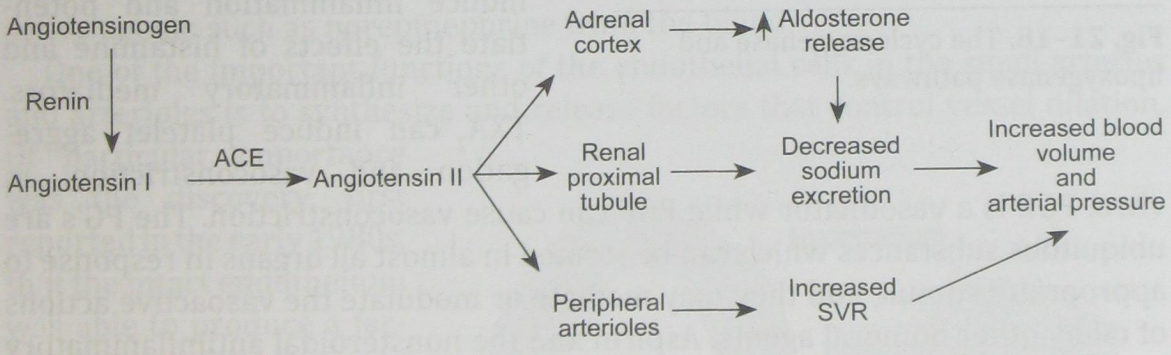


Fig. 21–17. Renin-angiotensin-aldosterone system. Abbreviations:
ACE — angiotensin-converting enzyme, SVR — systemic vascular resistance.

B. 4. Vasopressin. Another vasoactive polypeptide is *vasopressin* (or antidiuretic hormone, ADH), which is an octapeptide secreted by the supraoptic nuclei of the hypothalamus and stored in the neurohypophysis. Three primary classes of stimuli lead to vasopressin release: increased plasma osmolality, decreased baroreceptor and cardiopulmonary receptor firing and various types of stress such as physical injury or surgery. In addition, circulating angiotensin II stimulates vasopressin release. Although the primary action of vasopressin is to reduce urine flow by acting on the renal tubules it may also cause constriction of arterioles. It has been shown that vasopressin can act synergistically with catecholamines and enhance their vasoconstrictor effect.

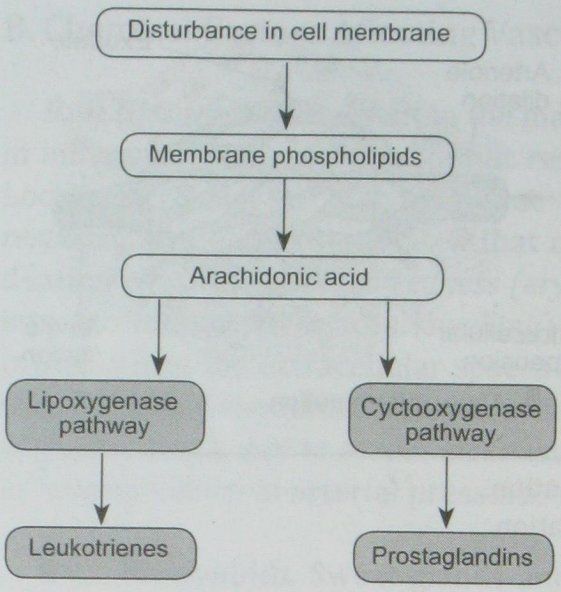


Fig. 21–18. The cyclooxygenase and lipoxygenase pathways.

B. 5. Prostaglandin's. Prostaglandin's (PG's) are 20-carbon, hydroxy fatty acids biosynthesized from arachidonic acid through the cyclooxygenase metabolic pathway (Fig. 21–18).

A microsomal enzyme PG synthetase causes the formation of unstable PG endoperoxides which are transformed to (a) prostacyclin (PGI_2) by prostacyclin synthetase, (b) thromboxane A_2 (TXA_2) by thromboxane synthetase or (c) PGE, PGF or PGD. Stable PG's induce inflammation and potentiate the effects of histamine and other inflammatory mediators. TXA_2 can induce platelet aggregation and vasoconstriction, in vitro. PGE is a vasodilator while PGF can cause vasoconstriction. The PG's are ubiquitous substances which can be formed in almost all organs in response to appropriate stimuli, and they may mediate or modulate the vasoactive actions of many other humoral agents. Aspirin and the nonsteroidal antiinflammatory drugs reduce inflammation by inactivating the first enzyme in the cyclooxygenase pathway for prostaglandin synthesis.

B. 6. Leukotrienes. Like the prostaglandins, the leukotrienes are formed from arachidonic acid, but through the lipoxygenase pathway (Fig. 21–18). Histamine and leukotrienes are complementary in action in that they have similar functions. Histamine is produced rapidly and transiently while the more potent leukotrienes are being synthesized. The leukotrienes also have been reported to affect the permeability of the postcapillary venules, the adhesion properties of endothelial cells, and the chemotaxis and extravascularization of neutrophils, eosinophils, and monocytes.

B. 7. PO_2, PCO_2, pH. The metabolic environment (the partial pressures of O_2 and CO_2 and the pH) in various tissues depends on the balance between circulatory transport and tissue metabolism. Whenever circulatory transport

is insufficient to meet metabolic needs, a reduction in tissue PO_2, an increase in PCO_2 and a decrease in pH would result These metabolic changes causes vasodilatation, though the relative importance of these changes varies with the organs or tissues in question. As a result of the vasodilatation and decrease in resistance, blood flow through the region rises to minimize the tissue metabolic changes. By such a mechanism of *autoregulation* by metabolic factors, the rate of blood flow and the metabolic environment of tissues can be regulated locally, independent of neural or hormonal influences.

B. 8. Endothelial-derived substances. The endothelium, which lies between the blood and the vascular smooth muscle, serves as a physical barrier for vasoactive substances that circulate in the blood. Once thought to be nothing more than a single layer of cells that line blood vessels, it is now known that the endothelium plays an active role in controlling vascular function. In addition to its function in capillary transport, the vascular endothelium removes vasoactive agents such as norepinephrine from the blood.

One of the important functions of the endothelial cells in the small arteries and arterioles is to synthesize and release factors that control vessel dilation. Of particular importance was the discovery, first reported in the early 1980s, that the intact endothelium was able to produce a factor that caused relaxation of vascular smooth muscle. This factor was originally named *endothelium-derived relaxing factor* and is now known to be *nitric oxide.* The normal endothelium maintains a continuous release of nitric oxide, which is formed from L-arginine through the action of an enzyme called nitric oxide synthase (Fig. 21–19). The production of nitric oxide can be

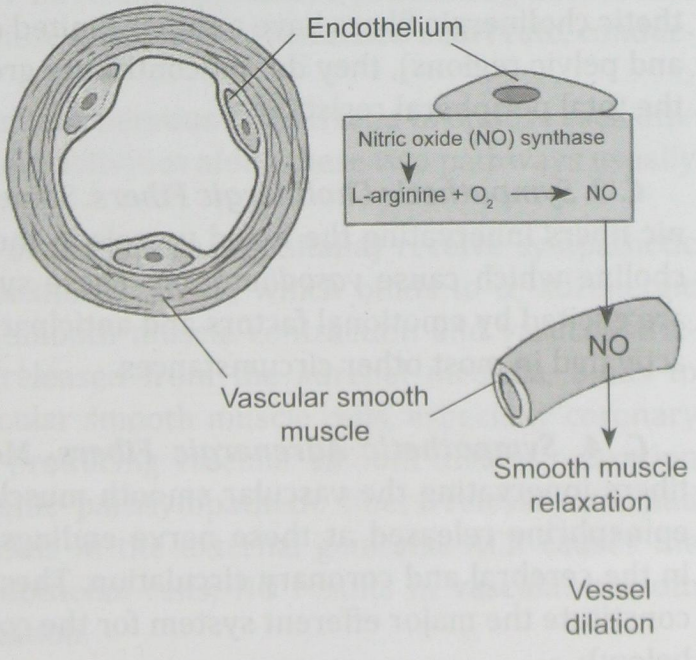


Fig. 21–19. Function of nitric oxide in smooth muscle relaxation.

stimulated by a variety of endothelial agonists, including acetylcholine, bradykinin, histamine, and thrombin.

Shear stress on the endothelium, resulting from an increase in bloodflow or blood pressure, also stimulates nitric oxide production and vessel relaxation. Nitric oxide also inhibits platelet aggregation and secretion of platelet contents, many of which cause vasoconstriction. The fact that nitric oxide is released into the vessel lumen (to inactivate platelets) and away from the lumen (to relax smooth muscle) suggests that it protects against both thrombosis and vasoconstriction. Nitroglycerin, which is used in treatment of angina, produces its effects by releasing nitric oxide in vascular smooth muscle of the target tissues.

C. Neural Regulation of Vascular Hindrance

C. 1. Dorsal Root Vasodilator Fibers. The axon reflex initiated by noxious stimuli leads to vasodilatation of the areas immediately surrounding the region affected. These changes are usually localized and exert no significant influence on the total peripheral resistance.

C. 2. Parasympathetic Fibers. The parasympathetic fibers release acetylcholine at their nerve endings to cause *vasodilatation*. Since these parasympathetic cholinergic fibers have a rather limited distribution (mainly to the facial and pelvic regions), they do not contribute greatly to the overall regulation of the total peripheral resistance.

C. 3. Sympathetic Cholinergic Fibers. Some of the sympathetic postganglionic fibers innervating the blood vessels in the skeletal muscle release acetylcholine which cause *vasodilatation*. These sympathetic cholinergic impulses are elicited by emotional factors and anticipation of exercise, and they are not activated in most other circumstances.

C. 4. Sympathetic Adrenergic Fibers. Most sympathetic postganglionic fibers innervating the vascular smooth muscles are adrenergic, and the norepinephrine released at these nerve endings causes *vasoconstriction* except in the cerebral and coronary circulation. These sympathetic adrenergic fibers constitute the major efferent system for the control of vascular hindrance (see below).

2. REFLECTORY REGULATION OF CARDIOVASCULAR FUNCTIONS

The above discussions of factors regulating cardiac output and peripheral resistance indicate that autonomic nervous activity, especially that of the sympathetic system, exerts important influences on cardiovascular function. Autonomic control of the heart and blood vessels was to a certain extent described in Chapter # 8.

Briefly, the heart is innervated by parasympathetic (vagus) and sympathetic (cardioaccelerator) nerve fibers. Parasympathetic fibers release acetylcholine (ACh), which binds to muscarinic receptors of the SA node, the atrioventricular node and specialized conducting tissues. Stimulation of parasympathetic fibers causes a slowing of the heart rate and conduction velocity. The ventricles are only sparsely innervated by parasympathetic nerve fibers, and stimulation of these fibers has little direct effect on cardiac contractility. Some cardiac parasympathetic fibers end on sympathetic nerves and inhibit the release of norepinephrine (NE) from sympathetic nerve fibers. Therefore, in the presence of sympathetic nervous system activity, parasympathetic activation reduces cardiac contractility.

Sympathetic fibers to the heart release NE, which binds to β_1-adrenergic receptors in the SA node, the AV node and specialized conducting tissues and cardiac muscle. Stimulation of these fibers causes increased heart rate, conduction velocity, and contractility.

The two divisions of the autonomic nervous system tend to oppose each other in their effects on the heart, and activities along these two pathways usually change in a reciprocal manner.

Blood vessels (except those of the external genitalia) receive sympathetic innervation only. The neurotransmitter is NE, which binds to α_1-adrenergic receptors and causes vascular smooth muscle contraction and vasoconstriction. Circulating epinephrine, released from the adrenal medulla, binds to β_2-adrenergic receptors of vascular smooth muscle cells, especially coronary and skeletal muscle arterioles, producing vascular smooth muscle relaxation and vasodilatation. Postganglionic parasympathetic fibers release ACh and nitric oxide (NO) to blood vessels in the external genitalia. ACh causes the further release of NO from endothelial cells; NO results in vascular smooth muscle relaxation and vasodilatation.

These neural control mechanisms are discussed below in terms of the three major components of a reflex arch: the efferent system, the integration center, and the afferent system.

2.1. The Efferent System

The efferent control of the cardiovascular system is exerted through the autonomic nerves. The main parasympathetic influence on the cardiovascular system is a slowing of the heart rate by vagal impulses. The sympathetic adrenergic system, on the other hand, has a much wider variety of actions. These include the acceleration of heart rate, increase of cardiac contractility, reduction of venous capacitance and constriction of resistance vessels. The first three effects cause an increase in cardiac output, whereas the last action raises the peripheral resistance. Therefore, sympathetic adrenergic impulses cause an elevation of arterial pressure, which causes an increase in blood flow through regions (the brain and the heart) not constricted by the adrenergic influence.

2.2. The Integrating Center

Circulatory reflexes are integrated at various levels of the central nervous system. A certain degree of integration of sympathetic reflexes can be achieved in the *spinal cord*.

For example, sensory impulses from thermal or pain receptors in the skin can directly excite the sympathetic neurons in the lateral horn of the thoraco-lumbar cord. The most important level of integration of autonomic efferent activities to the cardiovascular system, however, resides in the medulla oblongata. The cell bodies for the vagal efferents are located in the dorsal motor nuclei of the vagus in the dorsal aspects of the medulla. The neurons controlling the spinal sympathetic efferents are situated in the dorso-lateral reticular formation of the medulla (Fig. 21–20). Electrical stimulation of these areas causes an enhancement of sympathetic adrenergic activity and an AP elevation.

Therefore, these bilateral areas are designated as the pressor areas. More *medially and ventrally* located is the *depressor area*. Stimulation of neurons in the *depressor area* causes an inhibition of sympathetic adrenergic activity and a decrease in arterial pressure. There exists a reciprocal relationship between the pressor area and the depressor area in response to afferent impulses, such that whenever one is activated the other is reciprocally inhibited.

The pressor area, the depressor area and the dorsal motor nuclei of the vagus together constitute the *cardiovascular center*.

When we speak of an excitation of the cardiovascular center, however, we usually mean an excitation of the pressor area, while the other two areas may

be reciprocally inhibited. The pressor and depressor areas send descending fibers through multisynaptic pathways to regulate the activity of sympathetic neurons located in the thoraco-lumbar cord (Fig. 21–20).

Resuming, it should be stressed that *medulla oblongata has three* major cardiovascular functions:

- generating tonic excitatory signals to spinal sympathetic preganglionic fibers;
- integrating cardiovascular reflexes;
- integrating signals from supramedullary neural networks and from circulating hormones and drugs.

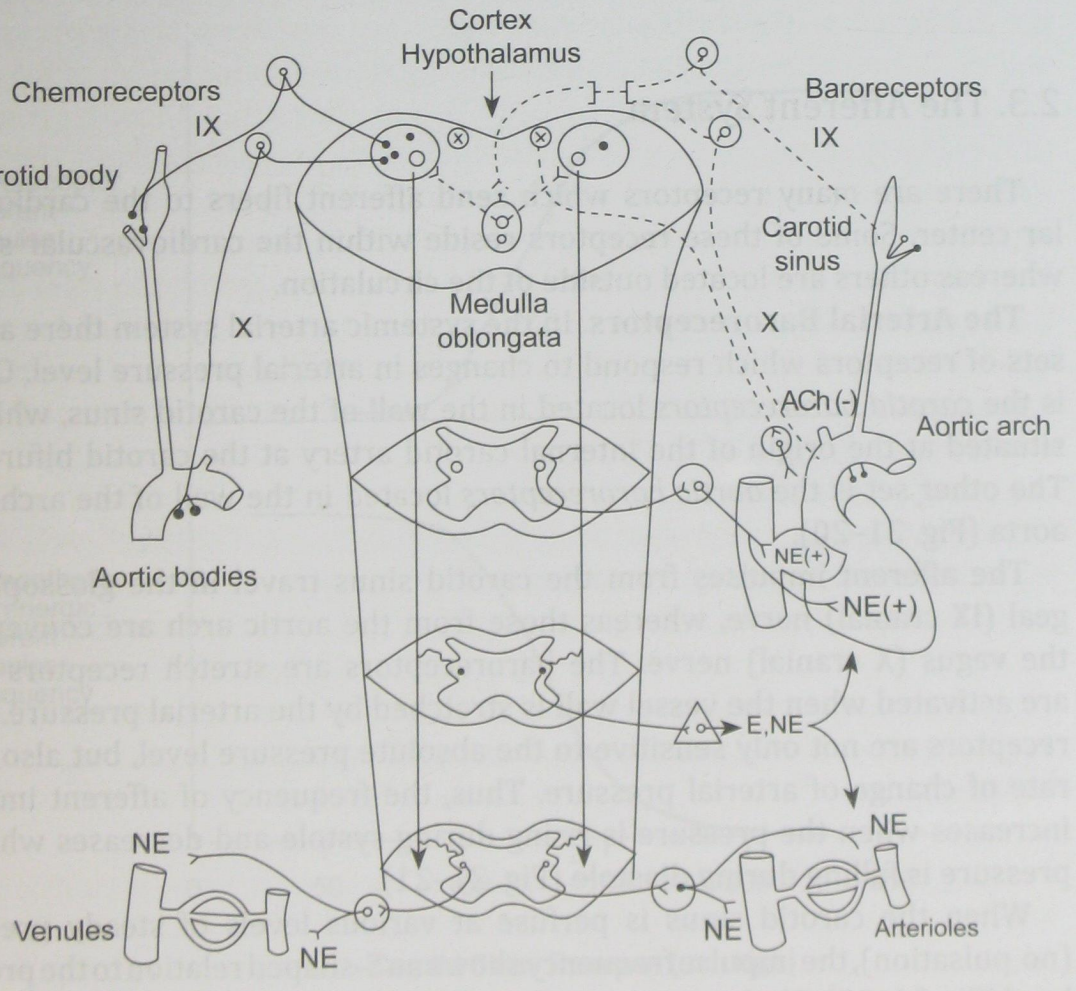


Fig. 21–20. Cardiovascular reflex pathways.

In addition to the medulla oblongata, there are also neurons in the higher parts of the central nervous system which can modify sympathetic influences on the cardiovascular system. These include the neurons in the *cerebral cortex* (especially the limbic system) and the *hypothalamus.* They are particularly important in the activation of sympathetic adrenergic efferents in response to emotional excitement and in anticipation of exercise.

The hypothalamus not only contains neurons which can modify sympathetic adrenergic efferents via their influence on the medullary cardiovascular center, but also has neurons which send fibers directly to the sympathetic cholinergic neurons in the spinal cord. This hypothalamus-sympathetic cholinergic pathway plays a significant role in inducing sudden vasodilatation in the skeletal muscle in response to certain stimuli. This is the pathway that causes some people to faint at the sight of blood.

2.3. The Afferent System

There are many receptors which send afferent fibers to the cardiovascular center. Some of these receptors reside within the cardiovascular system, whereas others are located outside of the circulation.

The Arterial Baroreceptors. In the systemic arterial system there are two sets of receptors which respond to changes in arterial pressure level. One set is the *carotid baroreceptors* located in the wall of the carotid sinus, which are situated at the origin of the internal carotid artery at the carotid bifurcation. The other set is the *aortic baroreceptors* located in the wall of the arch of the aorta (Fig. 21–20).

The afferent impulses from the carotid sinus travel in the glossopharyngeal (IX cranial) nerve, whereas those from the aortic arch are conveyed by the vagus (X cranial) nerve. The baroreceptors are stretch receptors which are activated when the vessel wall is stretched by the arterial pressure. These receptors are not only sensitive to the absolute pressure level, but also to the rate of change of arterial pressure. Thus, the frequency of afferent impulses increases when the pressure is rising during systole and decreases when the pressure is falling during diastole (Fig. 21–21).

When the carotid sinus is perfuse at various levels of steady pressures (no pulsation), the impulse frequency shows an S-shaped relation to the pressure level (Fig. 21–22). The baroreceptor impulses cease completely at pressure below 50 mm Hg, rise progressively between 60 and 180 mm Hg, and reach a maximum

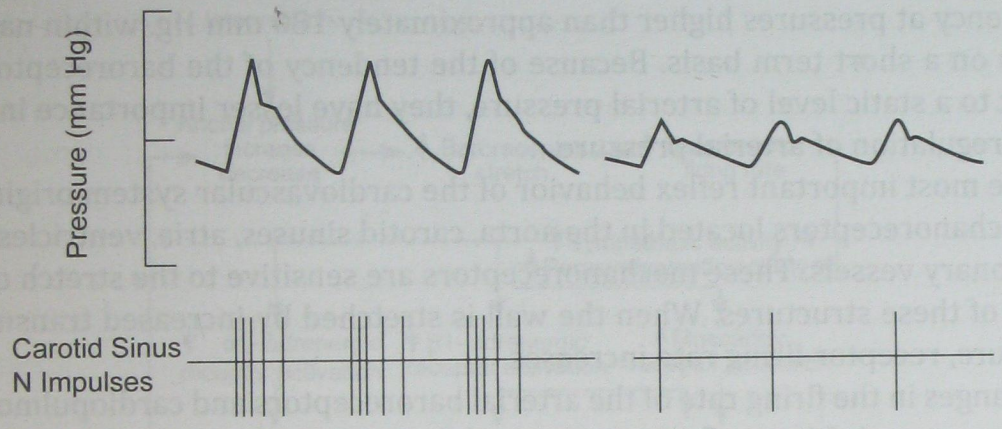


Fig. 21–21. Baroreceptor impulses.

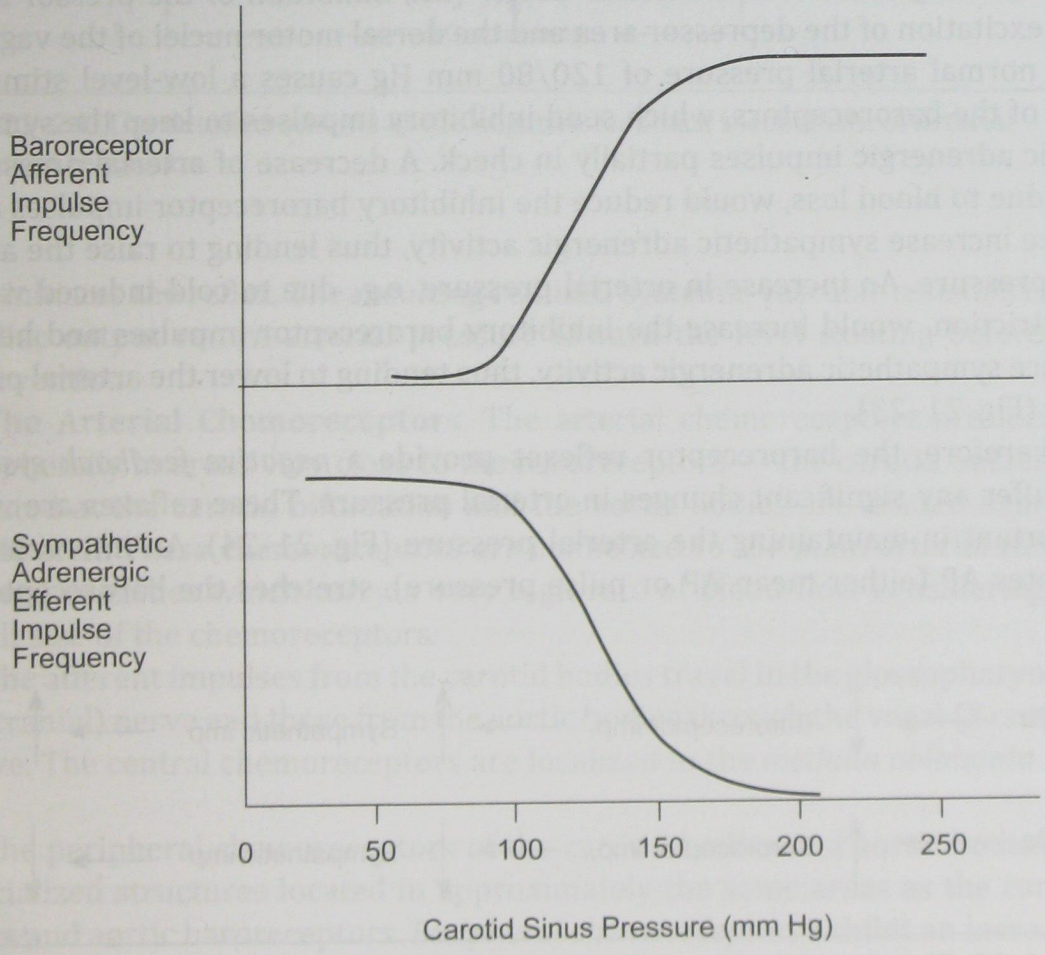


Fig. 21–22. Baroreceptors and sympathetic response.

frequency at pressures higher than approximately 180 mm Hg. within narrow limits on a short term basis. Because of the tendency of the baroreceptors to adapt to a static level of arterial pressure, they have lesser importance in long term regulation of arterial pressure.

The most important reflex behavior of the cardiovascular system originates in mechanoreceptors located in the aorta, carotid sinuses, atria, ventricles, and pulmonary vessels. These mechanoreceptors are sensitive to the stretch of the walls of these structures. When the wall is stretched by increased transmural pressure, receptor firing rate increases.

Changes in the firing rate of the arterial baroreceptors and cardiopulmonary baroreceptors initiate reflex responses of the autonomic nervous system that alter cardiac output and vascular resistance.

A very important feature of the *arterial baroreceptor impulses* is that they are *inhibitory to the cardiovascular center* (i.e., inhibition of the pressor area and excitation of the depressor area and the dorsal motor nuclei of the vagus). The normal arterial pressure of 120/80 mm Hg causes a low-level stimulation of the baroreceptors, which send inhibitory impulses to keep the sympathetic adrenergic impulses partially in check. A decrease of arterial pressure, e.g., due to blood loss, would reduce the inhibitory baroreceptor impulses and hence increase sympathetic adrenergic activity, thus lending to raise the arterial pressure. An increase in arterial pressure, e.g., due to cold-induced vasoconstriction, would increase the inhibitory baroreceptor impulses and hence reduce sympathetic adrenergic activity, thus tending to lower the arterial pressure (Fig. 21–23).

Therefore, the baroreceptor reflexes provide a *negative feedback system* to buffer any significant changes in arterial pressure. These reflexes are very important in maintaining the arterial pressure (Fig. 21–24). An intervention elevates AP (either mean AP or pulse pressure), stretches the baroreceptors,

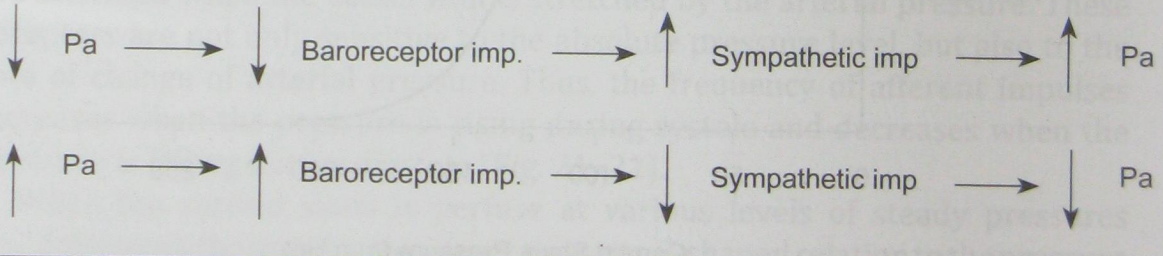


Fig. 21–23. The principal scheme of the negative feedback mechanism of arterial pressure regulation.

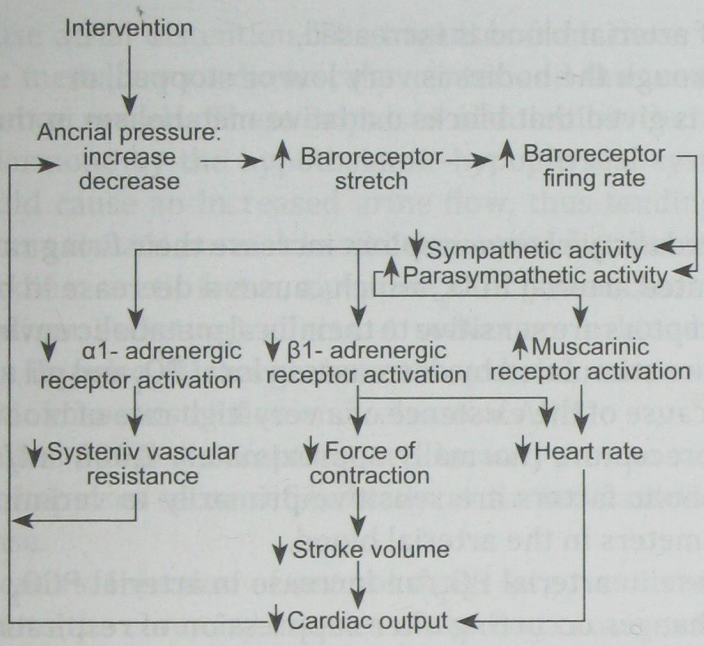


Fig. 21–24. The detailed scheme of the negative feedback mechanism of arterial pressure regulation.

and initiates the reflex. The resulting reduced systemic vascular resistance and cardiac output return arterial pressure toward the level existing before the intervention.

The Arterial Chemoreceptors. The arterial chemoreceptors are located peripherally in areas very close to the baroreceptors — the carotid bodies are situated at the carotid bifurcation and the aortic bodies are located near the aortic arch. These chemoreceptors are connected to the main arterial tree by arterial branches which carry a very high rate of blood flow considering the small size of the chemoreceptors.

The afferent impulses from the carotid bodies travel in the glossopharyngeal (IX cranial) nerve and those from the aortic bodies through the vagal (X cranial) nerve. The central chemoreceptors are localized in the *medulla oblongata*.

The peripheral chemoreceptors of the carotid bodies and aortic bodies are specialized structures located in approximately the same areas as the carotid sinus and aortic baroreceptors. *Peripheral chemoreceptors* exhibit an increased firing rate when:

1) the PO_2 or pH of the arterial blood is low,

2) the PCO_2 of arterial blood is increased,
3) the flow through the bodies is very low or stopped, or
4) a chemical is given that blocks oxidative metabolism in the chemoreceptor cells.

The *central medullary chemoreceptors* increase their firing rate primarily in response to elevated arterial PCO_2, which causes a decrease in brain pH.

The chemoreceptors are sensitive to their local metabolic environment. Thus, these receptors are stimulated by decreases in local PO_2 and pH and an increase in local PCO_2. Because of the existence of a very high rate of blood flow per unit weight of chemoreceptors (normally approximately 2,000 ml/ min/100 gm), these local metabolic factors are sensitive primarily to variations in the corresponding parameters in the arterial blood.

Thus, a decrease in arterial PO_2, an increase in arterial PCO_2 or a reduction in arterial pH (changes occurring after suppression of respiration) may cause *stimulation of the chemoreceptors.* The chemoreceptors can also be activated when there is a severe reduction in chemoreceptor blood flow, e.g., following a large volume of blood loss, where the marked reduction in blood flow will lead to cellular hypoxia and acidosis.

The *primary effect* of chemoreceptor impulses is to *stimulate the respiratory center* and increase ventilation but they also cause excitation of the cardiovascular center. The increase in sympathetic adrenergic activity resulting from chemoreceptor stimulation is mainly exerted on the resistance vessels and causes vasoconstriction. When circulatory and respiratory functions are normal, the existence of a high chemoreceptor blood flow and normal values for blood gases and pH provides a local metabolic environment which gives rise to the normal chemoreceptor firing rate. When the arterial PO_2 is reduced due to hypoventilation, the chemoreceptors become increasingly activated. The resulting chemoreceptor reflex causes a stimulation of respiratory movements to improve arterial PO_2 and a selective constriction of resistance vessels to provide the brain and the heart with sufficient blood flow and oxygen delivery.

Other Cardiovascular Receptors

1. *Mechanoreceptors.* There are receptors located in the atria and ventricles which are sensitive to mechanical stretch. The afferent fibers travel in the vagus nerves and in the sympathetic nerves. There are several types of atrial receptors. The type that has the clearest physiological function is the receptor which fires upon atrial distention. An expansion in blood volume

would cause atrial distention. The impulses from these atrial receptors inhibit the medullary cardiovascular center and thus cause dilatation of the resistance vessels. These impulses also inhibit the release of antidiuretic hormone by the hypothalamic-hypophyseal system. This latter effect would cause an increased urine flow, thus tending to reduce the blood volume toward normal. In addition, atrial distension causes release of atrial sodium-uretic factor, a peptide with potent diuretic and sodium-uretic (sodium excreting) effects. A reduction in blood volume would reduce the impulse discharge from these atrial receptors, thus causing sympathetic vasoconstriction and release of ADH.

The ventricular mechanoreceptors are located in the ventricular myocardium. These impulses cause reflex bradycardia and vasodilatation, leading to an AP decrease.

2. *Chemoreceptors.* Injection of several drugs (e.g., veratrum alkaloids and nicotine) into the coronary arteries of the dog causes reflex bradycardia and hypotension, together with some degree of respiratory inhibition. This is referred to as the Bezold-Jarisch reflex, the significance of which is not known in man.

Receptors in the Somatic System. In addition to the above receptors which are found in the cardiovascular system, there are also receptors in the skin, periosteum, skeletal muscles and joints which send afferent impulses to modify the activity of the cardiovascular center.

Thermoreceptors. Upon exposure to cold, the receptors in the skin send afferent impulses to stimulate the cardiovascular center and enhance sympathetic adrenergic impulses. Therefore, the AP rises in a cold environment. This can be demonstrated by the cold pressor test in which an individual immerses his forearm in a bucket of cold water (e.g., 10 °C) for several minutes. The vasoconstriction in the immersed hand is due to a combination of the direct effect of low temperature on the vascular smooth muscle and the reflex activation of the sympathetic adrenergic system. The vasoconstriction in the contralateral, non-immersed hand and other parts of the body is due to the reflex action alone. The afferent impulses resulting from exposure to warmth causes an inhibition of the cardiovascular center.

Pain Receptors. Superficial, acute pain, e.g. that inflicted by a cut in the skin, is associated with afferent impulses which stimulate the cardiovascular center and sympathetic adrenergic nerves.

Deep, dull pain, e.g., that elicited by the scraping of periosteum by an orthopedic surgeon, is associated with afferent impulses which inhibit the cardio-

vascular center. Therefore, this type of deep, dull pain is often accompanied by a precipitous fall in arterial pressure.

Proprioceptors in Muscles and Joints. During muscular exercise, the rhythmic activation of proprioceptors in the exercising limb leads to afferent impulses which are excitatory to the cardiovascular center and the respiratory center.

Cerebral Ischemic Response. The cardiovascular center is stimulated by local accumulation of PCO_2 and acid metabolites. This occurs when cerebral blood flow is reduced as the arterial pressure falls below 50 mm Hg. The cerebral ischemic response involves intense sympathetic adrenergic discharge, leading to a marked rise of arterial pressure as a last line of defense in an attempt to protect against cerebral ischemia.

Higher Centers. The activity of the medullary cardiovascular center can be modified by descending impulses from the cerebral cortex and hypothalamus. Emotional stimuli can activate these higher centers, which in turn excite the medullary cardiovascular center, giving rise to increases in heart rate and arterial pressure. These higher centers are also responsible for the activation of the sympathetic system that occurs in anticipation of exercise.

The overview of the reflectory regulation of the circulatory system is presented on Fig. 21–25.

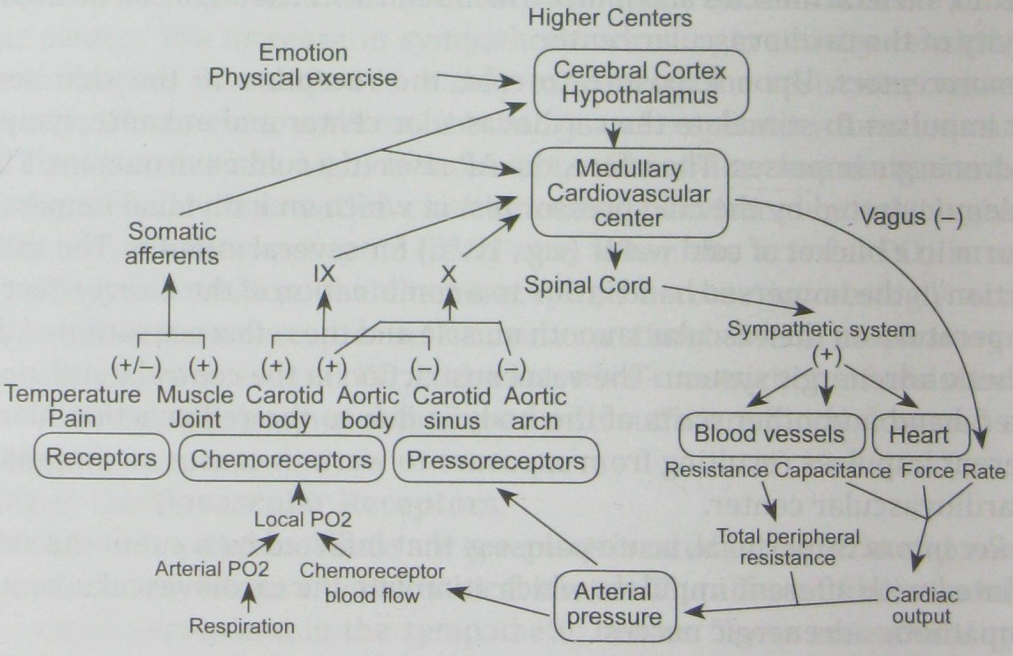


Fig. 21–25. Reflex control of the circulatory system.

Unit VIII RESPIRATORY PHYSIOLOGY

Chapter 22. Structural organization of the respiratory system. The mechanism of inspiration and expiration

Respiration through the lungs has two main functions: firstly, to supply O_2 to the blood and, secondly, to regulate the acid-base balance via the CO_2 concentration in the blood. When so defined, respiration includes the following major proper respirative processes:

1. Gas intake and exhalation (mechanics).
2. Alveolar gas exchange (convection and diffusion).
3. Transport of respiratory gases in blood (convective blood transport).
4. Tissue respiration (diffusion and metabolism).
5. Control of respiration (chemoreceptors and central nervous system).

With this the respiratory system operates to provide:

1. Intake of a large volume of fresh air into the alveolar space of the lungs (appr. 4 l/min at rest increasing to appr. 100 l/min at peak exercise).
2. Close contact between the large volume of alveolar gas and large volume of venous blood (appr. 5 l / min at rest and up to 30 l/min with exercise) which are separated by the thin alveolar epithelial and pulmonary capillary endothelial membranes. The total surface area of the alveolar-capillary membrane is appr. 90–140 m^2.
3. Production of fully oxygenated blood which then is transported via the systemic circulation to all cells of the body.
4. Exhalation of the large volume of alveolar gas (approximately 4 l/min) to excrete the metabolic product, CO_2.

It should be mentioned also that the respiratory system performs the other very important functions:

- liquid volume regulation;
- thermoregulation;
- bloodflow;
- organism defensive reaction performance;
- exchange and/or synthesis of the biologically active substances;
- speech-promoting function.

Respiration is a vital function, i.e., if stopped, life can not continue. The central nervous system is the most sensitive tissue to the lack of O_2 and irreversible losses in CNS function appear after 4–8 minutes of O_2 deprivation. Respiration is the process whereby oxygen from the air is transferred to the blood and carbon dioxide is eliminated from the body. The mechanics of breathing serve to ventilate the alveoli, through whose walls O_2 can diffuse into the blood and CO_2 can diffuse out. Respiratory gases in the blood are largely transported in bound form. The amount transported depends, among other factors, on the concentration in blood and on pulmonary blood flow (perfusion). It is the task of respiratory regulation to adapt ventilation to the specific requirements. The nervous system controls the movement of the respiratory muscles and adjusts the rate of breathing so that it matches the needs of the body during various levels of activity. The content in this chapter focuses on the structure and function of the respiratory system as it relates to these aspects of respiration.

1. STRUCTURAL ORGANIZATION OF THE RESPIRATORY SYSTEM

The respiratory system consists of the air passages and the lungs. Functionally, the respiratory system can be divided into two parts: the *conducting airways*, through which air moves as it passes between the atmosphere and the lungs, and the *respiratory tissues* of the lungs, where gas exchange takes place.

The conducting portion consists of the *nose, pharynx, larynx, trachea, bronchi, bronchioles* and *terminal bronchioles.*

The respiratory portion consists of *small respiratory bronchioles, alveolar ducts, alveolar sacs* and *alveoli.*

Air is propelled through the respiratory system by movements of the thoracic cage, intercostal muscles, diaphragm, and the elastic and collagen components

of the lungs. The cavities containing the lungs are lined by a serous membrane, the *pleura*, a thin layer of collagenous tissue containing fibroblasts, macrophages and several prominent layers of elastic fibers.

1.1. The Conducting Airways

The conducting airways consist of the nasal passages, mouth and pharynx, larynx, trachea, bronchi, and bronchioles (Fig. 22–1). The air we breathe is warmed, filtered, and moistened as it moves through these structures. Additionally, one could smell the air while breathing by nose. Heat is transferred to the air from the blood flowing through the walls of the respiratory passages; the mucociliary blanket removes foreign materials; and water from the mucous membranes is used to moisten the air.

The conducting airways are lined with a pseudostratified columnar epithelium that contains a mosaic of mucus-secreting goblet cells and cells that secreting goblet cells and cells that contain hair like projections called *cilia* (Fig. 22–2).

The structural support of the trachea is provided by 16 to 20 C-shaped hyaline cartilage plates that are embedded in the fibroelastic connective tissue of the ventral and lateral walls of the trachea. The posterior wall of the trachea that is devoid of cartilage contains a thick band of smooth muscle bundles, the *tracheal muscle* and a *fibroelastic ligament.* The ligament prevents overdistention of the lumen. Contraction and relaxation of the tracheal muscle in response to parasympathetic and sympathetic stimulation narrows and expands the lumen of the trachea.

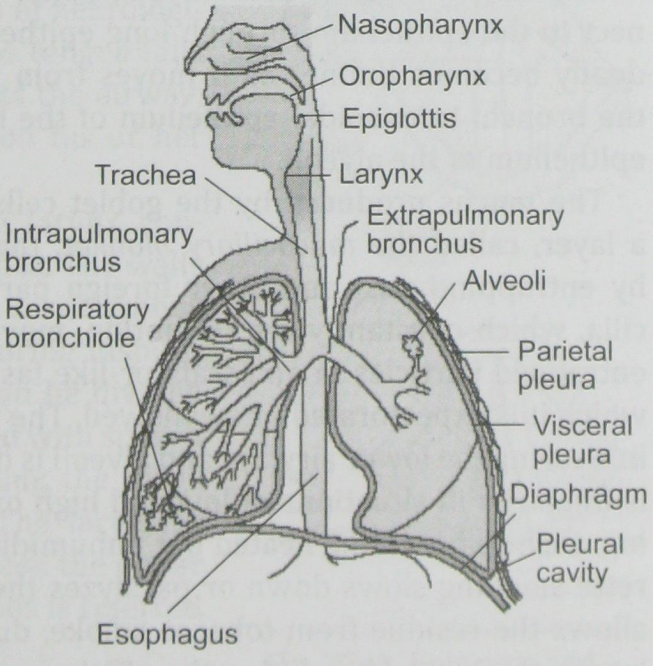


Fig. 22–1. The structure of the respiratory system.

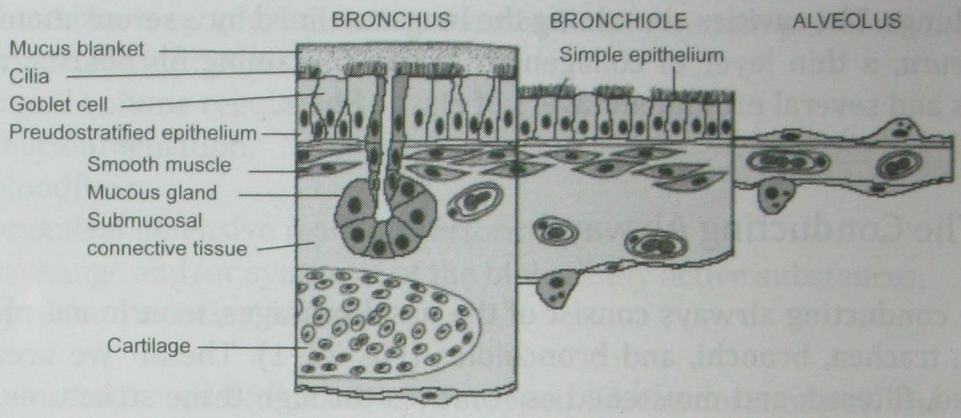


Fig. 22–2. The structure of the airway wall

The airway wall structure presented on this figure has bronchus, bronchiole, and alveolus. The bronchial wall contains pseudostratified epithelium, smooth muscle cells, mucous glands, connective tissue, and cartilage. In smaller bronchioles, a simple epithelium is found, cartilage is absent, and the wall is thinner. The alveolar wall is designed for gas exchange, rather than structural support (Fig. 22–2).

The larger bronchi contain additional mucus-secreting glands that connect to the epithelium through long epithelial ducts. The epithelial layer gradually becomes thinner as it moves from the pseudostratified epithelium of the bronchi to cuboidal epithelium of the bronchioles and then to squamous epithelium of the alveoli.

The mucus produced by the goblet cells in the conducting airways forms a layer, called the *mucociliary blanket*, that protects the respiratory system by entrapping dust and other foreign particles that enter the airways. The cilia, which constantly are in motion, move the mucociliary blanket with its entrapped particles in an escalator-like fashion toward the oropharynx, from which it is expectorated or swallowed. The function of the mucociliary blanket in clearing the lower airways and alveoli is optimal at normal oxygen levels and is impaired in situations of low and high oxygen levels. It is impaired by drying, such as breathing heated but unhumidified indoor air during winter. Cigarette smoking slows down or paralyzes the motility of the cilia. This slowing allows the residue from tobacco smoke, dust, and other particles to accumulate in the lungs, decreasing the efficiency of this pulmonary defense system. The airways are kept moist by water contained in the mucous layer. Moisture is added to the air as it moves through the conducting airways. The capacity

of the air to contain moisture or water vapor without condensation increases as the temperature rises. Thus, the air in the alveoli, which is maintained at body temperature, usually contains considerably more water vapor than the atmospheric-temperature air that we breathe.

Nasopharyngeal Airways. The nose is the preferred route for the entrance of air into the respiratory tract during normal breathing. As air passes through the nasal passages, it is filtered, warmed, and humidified. The outer nasal passages are lined with coarse hairs, which filter and trap dust and other large particles from the air. The upper portion of the nasal cavity is lined with mucous membrane that contains a rich network of small blood vessels; this portion of the nasal cavity supplies warmth and moisture to the air we breathe.

The mouth serves as an alternative airway when the nasal passages are plugged or when there is a need for the exchange of large amounts of air, as occurs during exercise. The oropharynx extends posteriorly from the soft palate to the epiglottis. The oropharynx is the only opening between the nose and mouth and the lungs. Both swallowed food on its way to the esophagus and air on its way to the larynx pass through it. Obstruction of the oropharynx leads to immediate cessation of ventilation. Neural control of the tongue and pharyngeal muscles may be impaired in coma and certain types of neurological disease. In these conditions, the tongue falls back into the pharynx and obstructs the airway, particularly if the person is lying on his or her back.

Laryngotracheal Airways. The larynx connects the oropharynx with the trachea. The walls of the larynx are supported by firm cartilaginous structures that prevent collapse during inspiration. The functions of the larynx can be divided into two categories: those associated with speech and those associated with protecting the lungs from substances other than air. The larynx is located in a strategic position between the upper airways and the lungs and sometimes is referred to as the "watchdog of the lungs".

The cavity of the larynx is divided into two pairs of two-by-two-folds of mucous membrane stretching from front to back with an opening

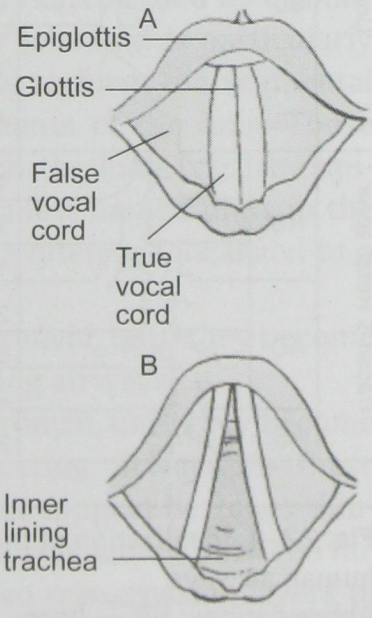


Fig. 22–3. Epiglottis and vocal cords viewed above with glottis closed (part A) and glottis opened (part B).

in the midline (Fig. 22–3). The upper pair of folds, called the *vestibular folds*, has a protective function. The lower pair of folds has cord-like margins; they are termed the *vocal folds* because their vibrations are required for making vocal sounds. The true vocal folds and the elongated opening between them are called the *glottis*. A complex set of muscles controls the opening and closing of the glottis. Speech involves the intermittent release of expired air and opening and closing of the glottis. The epiglottis, which is located above the vocal folds, is a large, leaf-shaped piece of cartilage that is covered with epithelium. During swallowing, the free edges of the epiglottis move downward to cover the larynx, thus routing liquids and foods into the esophagus.

In addition to opening and closing the glottis for speech, the vocal folds of the larynx can perform a sphincter function in closing off the airways. When confronted with substances other than air, the laryngeal muscles contract and close off the airway.

Tracheobronchial Tree. The tracheobronchial tree, which consists of the trachea, bronchi, and bronchioles, can be viewed as a system of branching tubes (Fig. 22–4). It is similar to a tree whose branches become smaller and more numerous as they divide. There are approximately 23 levels of branching, beginning with the conducting airways and ending with the respiratory airways, where gas exchange takes place. The first 16 generations of branching (Z) make up the conducting airways, and the last seven constitute the respiratory zone (or transitional and respiratory zone) (Fig. 22–4).

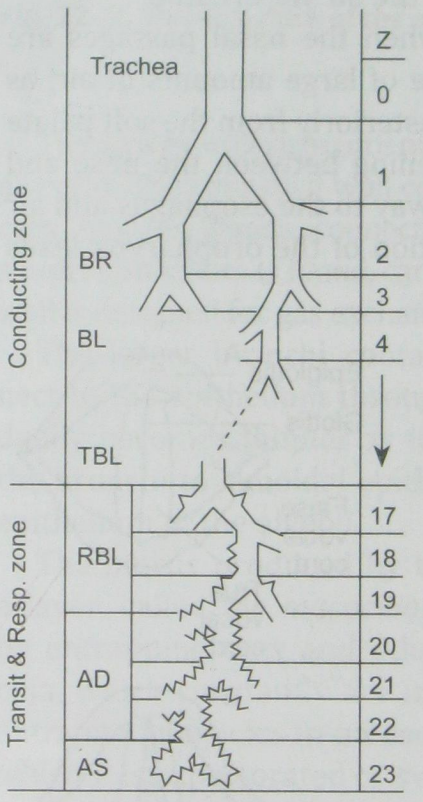


Fig. 22–4. Idealization of the human airways.
Abbreviations: BR — bronchus; BL — bronchiole; TBL — terminal bronchiole; RBL — respiratory bronchiole; AD — alveolar ducts; AS — alveolar sacs.

The trachea, or windpipe, is a continuous tube that connects the larynx and the major bronchi of the lungs (Fig. 22–5). The walls of the trachea are supported by horseshoe-shaped cartilages, which pre vent it from collapsing when the pressure in the thorax becomes negative. The trachea divides to form

the right and the left primary bronchi. Each bronchus enters the lung through a slit called the hilus. The structure of the primary bronchi is similar to that of the trachea in that these airways are supported by cartilaginous rings.

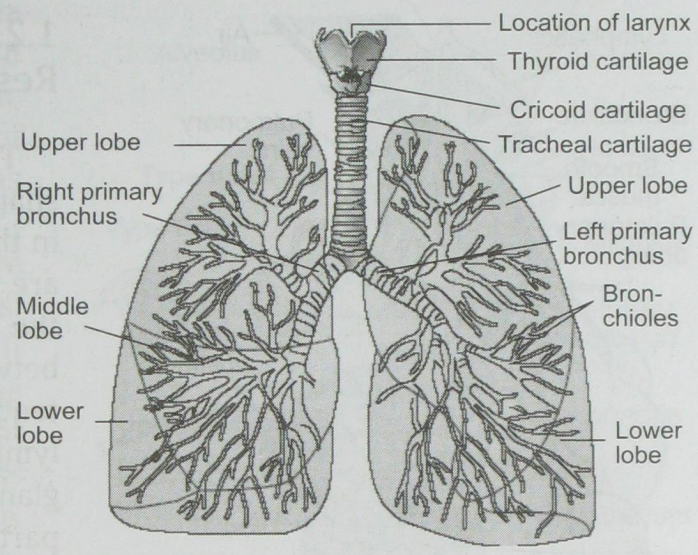


Fig. 22–5. The frontal view of the larynx, trachea and bronchial tree.

Each primary bronchus divides into *secondary*, or *lobular, bronchi*, which supply each of the lobes of the lungs — three in the right lung and two in the left. The right middle lobe bronchus is of relatively small diameter and length and sometimes bends sharply near its bifurcation. It is surrounded by a collar of lymph nodes that drain the middle and the lower lobe and is particularly subject to obstruction. The secondary bronchi divide to form the segmental bronchi, which supply the bronchopulmonary segments of the lung. These segments are identified according to their location in the lung (*e.g.*, the apical segment of the right upper lobe) and are the smallest named units in the lung. Lung lesions such as atelectasis and pneumonia often are localized to a particular bronchopulmonary segment.

The bronchi continue to branch, forming smaller bronchi, until they become the terminal bronchioles, the smallest of the conducting airways.

The bronchi continue to branch, forming smaller bronchi, until they become the terminal bronchioles, the smallest of the conducting airways. As these bronchi branch and become smaller, this cartilaginous support becomes thinner and then disappears at the level of the respiratory bronchioles. Between the cartilaginous support and the mucosal surface are two crisscrossing layers of smooth muscle that wind in opposite directions (Fig. 22–6). Bronchospasm, or contraction of these muscles, causes narrowing of the bronchioles and impairs air flow. The epithelial lining of conducting airways gradually becomes reduced from pseudo-stratified epithelium in the bronchi to a thin layer of tightly joined epithelial cells in the alveoli.

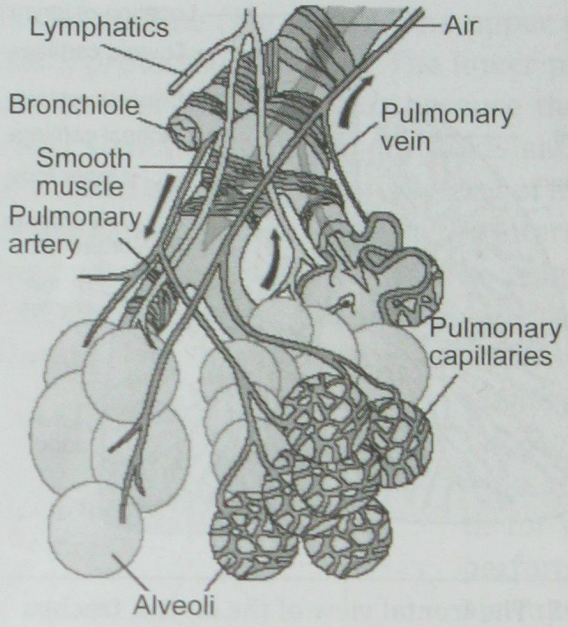


Fig. 22–6. Lobule of the lung, showing the bronchial smooth muscle fibers, pulmonary blood vessels, and lymphatics.

1.2. The Lungs and Respiratory Airways

The lungs are soft, spongy, cone-shaped organs located side by side in the chest cavity (Fig. 22–1). They are separated from each other by the *mediastinum* (*i.e.*, the space between the lungs) and its contents — the heart, blood vessels, lymph nodes, nerve fibers, thymus gland, and esophagus. The upper part of the lung, which lies against the top of the thoracic cavity, is called the *apex*, and the lower part, which lies against the diaphragm, is called the *base*. The lungs are divided into lobes: three in the right lung and two in the left (Fig. 22–5).

The lungs are the functional structures of the respiratory system. In addition to their gas exchange function, they inactivate vasoactive substances such as bradykinin; they convert angiotensin I to angiotensin II; and they serve as a reservoir for blood storage. Heparin-producing cells are particularly abundant in the capillaries of the lung, where small clots may be trapped.

Respiratory Lobules. The gas exchange function of the lung takes place in the lobules of the lungs. Each lobule, which is the smallest functional unit of the lung, is supplied by a branch of a terminal bronchiole, an arteriole, the pulmonary capillaries, and a venule (Fig. 22–6). Gas exchange takes place in the terminal respiratory bronchioles and the alveolar ducts and sacs. Blood enters the lobules through a pulmonary artery and exits through a pulmonary vein. Lymphatic structures surround the lobule and aid in the removal of plasma proteins and other particles from the interstitial spaces.

One feature that distinguishes respiratory bronchioles from other bronchioles is the presence of thin saccular outpockets, alveoli. Respiratory bronchioles give rise to alveolar ducts. Here the number of alveolar openings is so great that there is no detectable alveoli-free wall. Physiologically, the most important components of the lung are the thin-walled, terminal sac-

cular components, alveoli. The number of alveoli in human lung is estimated at 300 million, which produces a surface of approximately 143 m^2 for gas exchange.

The alveolar structures are composed of two types of cells: type I alveolar cells and type II alveolar cells (Fig. 22–7). The type I alveolar cells are flat squamous epithelial cells across which gas exchange takes place. The type II alveolar cells produce surfactant, a lipoprotein substance that decreases the surface tension in the alveoli. The alveoli also contain alveolar macrophages, which are responsible for the removal of offending substances from the alveolar epithelium. Therefore, alveolar type I cells comprise most of the alveolar surface. Alveolar type II cells are located in the corner between two adjacent alveoli (Fig. 22–7).

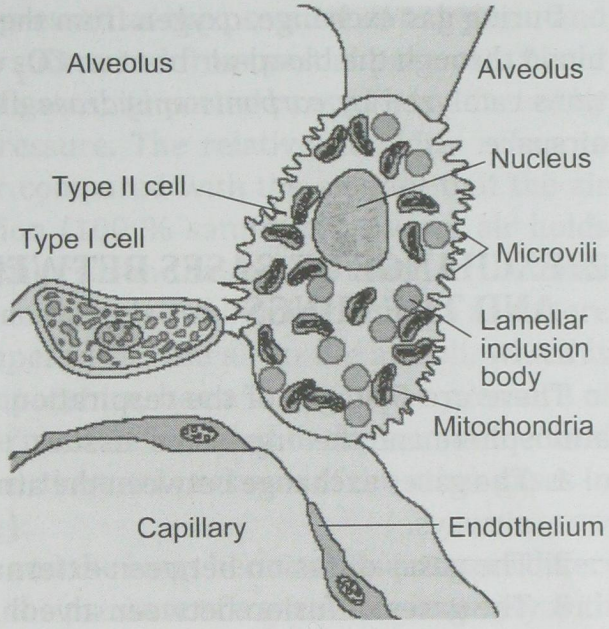


Fig. 22–7. Schematic drawing of the two types of alveolar cells and their relation to alveoli and capillaries.

The *interalveolar wall* (also know as the *interalveolar septum*) consists of two thin squamous epithelial layers. Between these two epithelial layers is the *interstitium* which consists of capillaries, fibroblasts, elastic and collagen fibers and macrophages.

At certain points, a common basement membrane forms between the epithelium of the alveoli and the endothelium of the capillary. The entire width of these three cell layers varies from 0.1 to 1.5 μm. The interalveolar wall may contain one or more 10–15 μm diameter pores, which connect neighboring alveoli and may be important for equalizing pressure in alveoli, or controlling air flow during bronchiolar obstruction.

Air in the alveoli is separated from capillary blood by 1) the surface lining and cytoplasm of the alveolar epithelial cell, 2) the fused basement membranes of closely apposed alveolar and endothelial cells, and 3) the cytoplasm of the endothelial cells. These three components are collectively referred to as the *blood-air (aero-hematic) barrier.*

During gas exchange, oxygen from the alveolar air passes into the capillary blood through the blood-air barrier. CO_2 which is liberated from H_2CO_3 in reactions catalyzed by *carbonic anhydrase* diffuses from the blood to the alveolar airspace.

2. EXCHANGE OF GASES BETWEEN THE ATMOSPHERE AND THE LUNGS

There are 5 stages of the respiration, i.e. the gases exchange between the atmosphere and the organisms' tissues.

1. The gases exchange between the atmosphere and the external conductive airways.
2. The gases diffusion between external airways and alveoli.
3. The gases diffusion between alveoli and blood (the air movement across the so-called aero-hematic barrier).
4. The gases transport by blood.
5. The gases diffusion between the blood and tissues (internal respiration).

2.1. Basic Properties of Gases

The air we breathe is made up of a mixture of gases, mainly *nitrogen* and *oxygen*. These gases exert a combined pressure called the *atmospheric pressure*. The pressure at sea level is defined as 1 atmosphere, which is equal to 760 mm Hg. When measuring respiratory pressures, atmospheric pressure is assigned a value of 0. A respiratory pressure of +15 mm Hg means that the pressure is 15 mm Hg above atmospheric pressure, and a respiratory pressure of −15 mm Hg is 15 mm Hg less than atmospheric pressure. Respiratory pressures often are expressed in centimeters of water (cm H_2O) because of the small pressures involved (1 mm Hg = 1.35 cm H_2O pressure).

The pressure exerted by a single gas in a mixture is called the *partial pressure* (The Law of Dalton). The law of partial pressures states that the total pressure of a mixture of gases, as in the atmosphere, is equal to the sum of the partial pressures of the different gases in the mixture. If the concentration of oxygen at 760 mm Hg (1 atmosphere) is 20 %, its partial pressure is 152 mm Hg (760 · 0.20).

The pressure of a gas is always expressed as the pressure of a dry gas, and the pressure of water vapor must be subtracted from partial pressure. Water vapor is different from other types of gases; its partial pressure is affected by temperature but not atmospheric pressure. The relative humidity refers to the percentage of moisture in the air compared with the amount that the air can hold without causing condensation (100 % saturation). Warm air holds more moisture than cold air. This is the reason that precipitation in the form of rain or snow commonly occurs when the relative humidity is high and there is a sudden drop in atmospheric temperature. The air in the alveoli, which is 100 % saturated at normal body temperature, has a water vapor pressure of 47 mm Hg. The water vapor pressure must be included in the sum of the total pressure of the gases in the alveoli (*i.e.*, the total pressure of the other gases in the alveoli is 760 — 47 = 713 mm Hg).

Air moves between the atmosphere and the lungs because of a pressure difference. According to the laws of physics, the pressure of a gas varies inversely with the volume of its container, provided the temperature remains constant. If equal amounts of a gas are placed in two different-size containers, the pressure of the gas in the smaller container is greater than the pressure in the larger container. The movement of gases is always from the container with the greater pressure to the one with the lesser pressure. The chest cavity can be viewed as a volume container. During inspiration, the size of the chest cavity increases and air moves into the lungs; during expiration, air moves out as the size of the chest cavity decreases.

2.2. Partial Pressures Of Gases In Solution

The amount of a gas which dissolves in a liquid is: 1) proportional to the pressure of that gas in contact with the surface of the liquid, 2) proportional to the solubility characteristics of the gas in that liquid and 3) inversely proportional to the temperature of the liquid.

When a gas is brought into contact with a liquid, many of the gas particles which collide with the surface of the liquid will enter into the liquid and become associated with the water molecules (solubilized). The kinetic energy of these dissolved gas molecules in turn produce a bombardment under the surface layer of the liquid, and many gas particles will escape into the gas phase. With time, an equilibrium is reached in which the rate of escape will equal the rate of entry, and the liquid is said to be saturated with the gas at the existing temperature and pressure. The amount of a gas dissolved in a liquid is directly pro-

portional to the pressure of that gas over the liquid; this principle is referred to as *Henry's law.*

Henry's law strictly applies only to a gas *in simple solution*, i.e., the gas must not undergo any chemical reaction with the solvent or with any solutes present. Henry's law applies fairly well to CO_2 in aqueous solutions at low CO_2 pressures, because the hydration of CO_2 does not proceed to any great extent in the absence of carbonic anhydrase.

For our purposes, the solubility coefficients for CO_2 and O_2 in plasma (0.51, 0.024) are also applied to whole blood on the assumption that the gases are similarly dissolved throughout the intracellular water of the red cell.

It is only the gas in simple solution which exerts the partial pressure, for it is only the gas in simple solution which has an escaping tendency and can equilibrate with the gas above the liquid. The dissolved gas which has undergone chemical reaction (H_2CO_3, $HbCO_2$) does not exert a partial pressure or escaping tendency. The partial pressure of a gas in solution is sometimes referred to as the *tension* of the gas, e.g., arterial 0_2 tension means P_aO_2.

3. MECHANICS OF BREATHING

Air movement occurs as a result of a pressure difference (gradient) between the alveolar air pressure, P_A, and the pressure of ambient air at the airway opening. In this setting, the pressure at the airway opening is P_B, the atmospheric or barometric pressure.

During *normal inspiration* the primary inspiratory muscles (diaphragm and external intercostals) contract generating a subatmospheric pleural pressure. This subatmospheric pleural pressure is transmitted to the alveolar space. Air at the airway opening is transported to the alveoli as a result of the pressure gradient and continues throughout inspiration until the pressure in the alveoli becomes equal to the pressure at the airway opening.

During *normal expiration* the inspiratory muscles relax, the relaxed diaphragm is moved superiorly due to the "negative" Ppl. The loss of inspiratory muscle tension causes the pleural pressure to increase and the pressure in the alveoli becomes greater than the pressure at the airway opening. Accordingly, expired gas flows out through the airway opening until pressure in the alveolus equals pressure at the airway opening.

During *active expiration* the inspiratory muscles relax, and the primary expiratory muscles (abdominal muscles, internal intercostals) contract, greatly

increasing the pleural and alveolar pressures. The P_A will greatly exceed pressure at the airway opening and a large volume of gas is rapidly exhaled.

Ventilation is concerned with the movement of gases into and out of the lungs. It relies on a system of open airways and the respiratory pressures created as the movements of the respiratory muscles change the size of the chest cage. The degree to which the lungs inflate and deflate depends on the respiratory pressures inflating the lung, compliance of the lungs, and airway resistance.

To explain this material one should know well the absolute indexes of the external respiration, which divided on two subgroups — the static and dynamic. The lung volumes and capacities are static indexes of the external respiration.

3.1. Lung Volumes

The volume of air which enters the respiratory system will be equal to the increase in the volume of the lung-chest wall unit. Many of the lung volume subdivisions can be measured by use of a *spirometer*, a device that allows accurate measurement of gas volumes. The following diagram shows a typical spirometer tracing of the various volumes and their definitions (Fig. 22–8).

Lung volumes, or the amount of air exchanged during ventilation, can be subdivided into four components:

1) the *tidal volume* (TV) — volume of air inspired and then expired with each breath. The T is usually about 500 ml.
2) the *inspiratory reserve volume* (IRV) — volume of air that we can additionally inspired till maximal lungs volume after the end of the normal inspiration. The average IRV constitutes 2.5–3.0 l of air.
3) the *expiratory reserve volume* (ERV) — volume of air that we can additionally expired till minimal lungs volume after the end of the normal expiration. The average ERV constitutes 1200–1500 ml of air.
4) Approximately from 800 ml till 1200 ml of air always remains in the lungs after forced expiration. This air is the *residual volume* (RV). The RV increases with age because there is more trapping of air in the lungs at the end of expiration. The RV cannot be measured with the spirometer because this air cannot be expressed from the lungs. It is measured by indirect methods, such as the helium dilution methods, the nitrogen washout methods or body plethysmography.

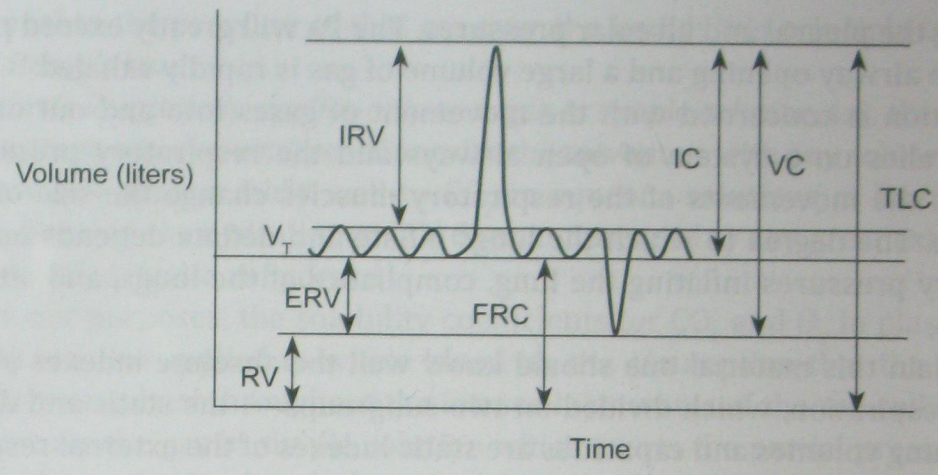


Fig. 22–8. Respiratory volumes and capacities matching made with the help of spirome ter. Explanations see in the text.

3.2. Lung Capacities

Lung capacities include two or more lung volumes.

1. The *total lung capacity* (TLC) is the sum of all the volumes in the lungs. TLC is the volume of air that locates inside the lungs on the top of the deep forced inspiration. It is equal to appr. 6–8 liters of air.
2. The *vital capacity* (VC) equals the IRV plus the TV plus the ERV and is the amount of air that can be exhaled from the point of maximal inspiration. It is equal to appr. 4–5 liters of air.
3. The *functional residual capacity* (FRC) is the sum of the RV and ERV; it is the volume of air that remains in the lungs at the end of normal expiration. It is equal to appr. 2–3 liters of air.
4. The *inspiratory capacity* (IC) equals the TV plus the IRV. It is the amount of air a person can breathe in beginning at the normal expiratory level and distending the lungs to the maximal amount. It is equal to appr. 3.0-3.5 liters of air.

3.3. Lungs Dynamic Indexes

The previously described lung volumes and capacities are anatomic or static measures, determined by lung volumes and measured without relation to time.

The spirometer also is used to measure dynamic lung function (*i.e.*, ventilation with respect to time); these tests often are used in assessing pulmonary function.

1. The *rate of respiration*. Normally it should be within 13±1 per minute.
2. The *minute volume of respiration* or total ventilation, is the amount of air that is exchanged in 1 minute. It is determined by the metabolic needs of the body. The minute volume is equal to the TV multiplied by the respiratory rate. During normal activity it is about 6-7 liters (TV = 500 ml × respiratory rate of 12–14 breaths per minute). The efficiency of breathing is determined by matching the TV and respiratory rate in a manner that provides an optimal minute volume while minimizing the work of breathing.
3. Alveolar ventilation (AV) per minute is the amount of air that moves in and out all the alveoli during 1 minute. It could be measured by the following equation: AV = (TV — dead space volume) · the rate of respiration.
4. The *maximum voluntary ventilation* measures the volume of air that a person can move into and out of the lungs during maximum effort lasting for a specific period of time. This measurement usually is converted to liters per minute.
5. The *forced vital capacity* (FVC) which involves full inspiration to total lung capacity followed by forceful maximal expiration. Obstruction of airways produces a FVC that is lower than that observed with more slowly performed vital capacity measurements.
6. The *forced expiratory volume* (FEV) is the expiratory volume achieved in a given time period. The $FEV_{1.0}$ is the forced expiratory volume that can be exhaled in 1 second. The $FEV_{1.0}$ frequently is expressed as a percentage of the FVC. The $FEV_{1.0}$ and FVC are used in the diagnosis of obstructive lung disorders.
7. The *work of breathing* is determined by the amount of effort required to move air through the conducting airways and by the ease of lung expansion or compliance. Expansion of the lungs is difficult for persons with stiff and noncompliant lungs; they usually find it easier to breathe if they keep their TV low and breathe at a more rapid rate (*e.g.*, 300 · 20 = 6000 ml) to achieve their minute volume and meet their oxygen needs. In contrast, persons with obstructive airway disease usually find it less difficult to inflate their lungs but expend more energy in moving air through the airways. As a result, these persons take deeper breaths and breathe at a slower rate (*e.g.*, 600 · 10 = 6 000 ml) to achieve their oxygen needs.

3.4. Respiratory Pressures

The pressure inside the airways and alveoli of the lungs is called the *intrapulmonary pressure* or *alveolar pressure*. The gases in this area of the lungs are in communication with atmospheric pressure (Fig. 22–9). When the glottis is open and air is not moving into or out of the lungs, as occurs just before inspiration or expiration, the intrapulmonary pressure is zero or equal to atmospheric pressure.

The pressure in the pleural cavity is called the *intrapleural pressure*. The intrapleural pressure is always negative in relation to alveolar pressure, approximately — 4 mm Hg between breaths when the glottis is open and the alveolar spaces are open to the atmosphere. The lungs and the chest wall have elastic properties, each pulling in the opposite direction. If removed from the chest, the lungs would contract to a smaller size, and the chest wall, if freed from the lungs, would expand. The opposing forces of the chest wall and lungs create a pull against the visceral and parietal layers of the pleura, causing the pressure in the pleural cavity to become negative. During inspiration, the elastic recoil of the lungs increases, causing intrapleural pressure to become more negative than during expiration. Without the negative intrapleural pres-

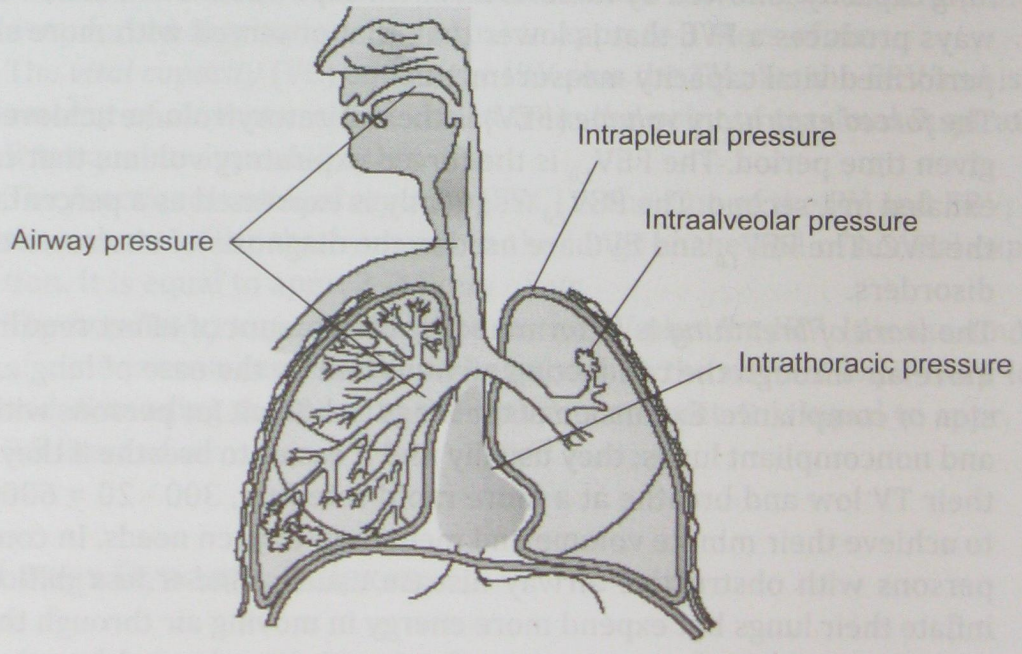


Fig. 22–9. Different types of respiratory pressure.

sure holding the lungs against the chest wall, their elastic recoil properties would cause them to collapse. Although the intrapleural pressure of the inflated lung is always negative in relation to alveolar pressure, it may become positive in relation to atmospheric pressure (*e.g.*, during forced expiration and coughing).

The *intrathoracic pressure* is the pressure in the thoracic cavity. It is essentially equal to intrapleural pressure and is the pressure to which the lungs, heart, and great vessels are exposed. Forced expiration against a closed glottis (Valsalva maneuver) compresses the air in the thoracic cavity and produces marked increases in intrathoracic and intrapleural pressures.

3.5. The Mechanism of Inspiration and Expiration

The Chest Cage and Respiratory Muscles

Mechanically, the act of breathing depends on the fact that the chest cavity is a closed compartment whose only opening to the exterior is the trachea.

Ventilation consists of inspiration and expiration. During *inspiration*, the size of the chest cavity increases, the intrapleural pressure becomes more negative, and air is drawn into the lungs. *Expiration* occurs as the elastic components of the chest wall and lung structures that were stretched during inspiration recoil, causing the size of the chest cavity to decrease and the pressure in the chest cavity to increase (Fig. 22–10).

Inspiration. The diaphragm is the principal muscle of inspiration. When the diaphragm contracts, the abdominal contents are forced downward and the chest expands from top to bottom (Fig. 22–10). The external intercostal muscles, which also aid in inspiration, connect to the adjacent ribs and slope downward and forward. When they contract, they raise the ribs and rotate them slightly so that the sternum is pushed forward; this enlarges the chest from side to side and from front to back. Because of the muscle active work inspiration is considered to be an active process.

At the beginning of inspiration the following processes occur. The glottis is opened and all respiratory muscles are relaxed. This is the neutral or equilibrium point of the respiratory system (Fig. 22–11). Intrapleural pressure in this moment is negative because the inward elastic recoil of the lungs is opposed by the outward-directed recoil of the chest wall. In this moment the intrapleural pressure is appr.3 mm Hg less if compare with the atmosphere air. The intrapleural force and recoil force are equal and opposite and because no air is flow-

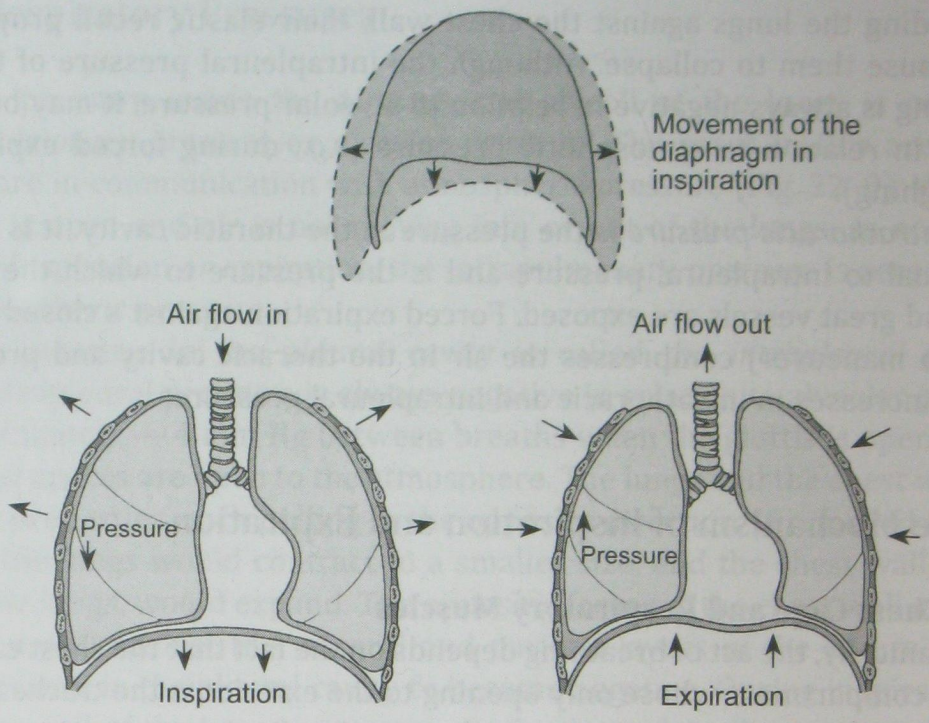


Fig. 22–10. Movement of the diaphragm and changes in chest volume and pressure during inspiration and expiration.

Top fragment – schematical illustration of the diaphragm movement and expansion of the chest cavity during inspiration.

Left fragment – schematical illustration of the intrapleural pressure decreasing and lungs size increasing during inspiration. As the result air moves into the lungs.

Right fragment – schematical illustration of the intrapleural pressure increasing and lungs size decreasing during inspiration. As the result air moves out of the lungs.

ing through the open glottis alveolar pressure must be zero. By convention the atmospheric pressure is set to equal zero.

During the top of inspiration that induced by the diaphragm and some additional muscles contraction the size of the thoracic cage is increased. The net result is to make intrapleural pressure more negative if compare with the atmospheric pressure. On the top of the silent inspiration the intrapleural pressure is 9–12 mm Hg less comparing with the surrounding atmospheric air pressure.

The greater the contraction, the greater the change in intrapleural pressure and the larger the force trying to expand the lungs (Fig. 22–12). On the top of the forced inspiration the intrapleural pressure is 15–21 mm Hg less com-

paring with the surrounding atmospheric air pressure. The lungs expansion causes the gases in the alveoli to expand creating a slightly negative alveolar pressure. This causes air to flow into the lung.

During the end of inspiration the lung expands until the recoil force increases to equal the intrapleural pressure (Fig. 22–13).

Once the forces are again equal and opposite the lungs are at their new larger volume. The inflowing air returns alveolar pressure toward zero and when it reaches zero airflow stops. Under resting conditions about 500 ml of air flows into the lung system in order to return alveolar pressure back to zero.

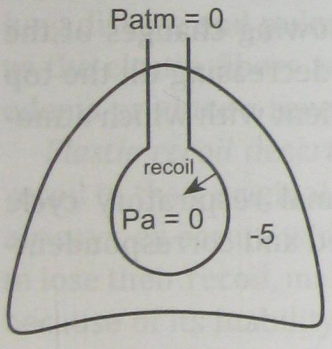


Fig. 22–11. Schematical draw of the differences of pressures before inspiration.

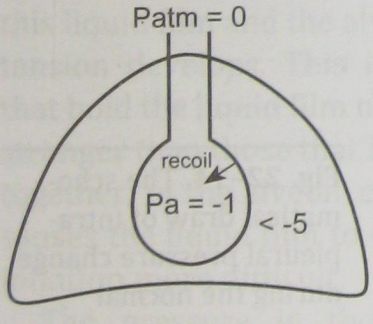


Fig. 22–12. Schematical draw of the differences of pressures during inspiration.

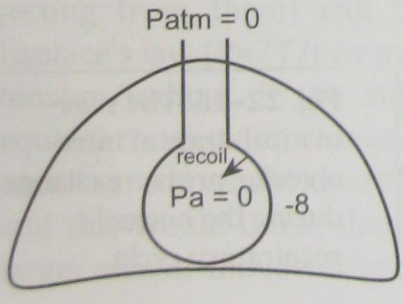


Fig. 22–13. Schematical draw of the differences of pressures at the end of inspiration.

Expiration. Expiration is largely passive. It occurs as the elastic components of the chest wall and lung structures that were stretched during inspiration recoil, causing air to leave the lungs as the intrapleural pressure increases (Fig. 22–11). Deflation of the lung compresses the gases in the alveoli creating a slightly positive alveolar pressure. This is also causes air to flow out of the lungs.

When needed, the abdominal and the internal intercostal muscles can be used to increase expiratory effort. The increase in intraabdominal pressure that accompanies the forceful contraction of the abdominal muscles pushes the diaphragm upward and results in an increase in intrapleural pressure. The internal intercostals muscles move inward, which pulls the chest downward, increasing expiratory effort.

Therefore, resuming it should be mentioned the following changes of the intrapleural pressure that indicate on the Fig. 22–14. It decreasing till the top of the silent inspiration makes greater the pressure gradient with which atmospheric air flows into the lungs.

The intraalveolar pressure changes during the normal respiratory cycle given on the Fig. 22–15 shows that inspiration terminated and correspondently expiration starts when this index equals to zero.

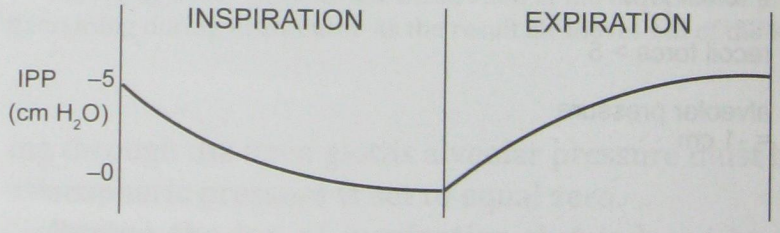


Fig. 22–14. The schematical draw of intrapleural pressure change during the normal respiratory cycle.

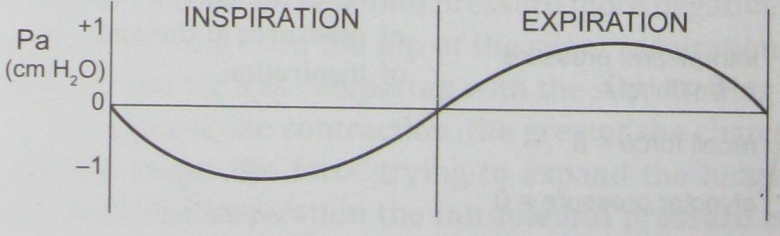


Fig. 22–15. The schematical draw of intraalveolar pressure change during the normal respiratory cycle.

3.6. The Forces Providing the Respiration

Lung Compliance

Lung compliance refers to the ease with which the lungs can be inflated. It is determined by the elastin and collagen fibers of the lung, its water content, and surface tension. Compliance can be appreciated by comparing the ease of blowing up a new balloon that is stiff and noncompliant with one that has been previously blown up and stretched.

Changes in Elastin/Collagen Composition of Lung Tissue. Lung tissue is made up of elastin and collagen fibers. The elastin fibers are easily stretched and increase the ease of lung inflation, whereas the collagen fibers resist stretching and make lung inflation more difficult. In lung diseases such as interstitial lung disease and pulmonary fibrosis, the lungs become stiff and noncompliant as the elastin fibers are replaced with scar tissue. Pulmonary congestion and edema produce a reversible decrease in pulmonary compliance.

Elastic recoil describes the ability of the elastic components of the lung to recoil to their original position after having been stretched. Overstretching the airways, as occurs with emphysema, causes the elastic components of the lung to lose their recoil, making the lung easier to inflate but more difficult to deflate because of its inability to recoil.

Surface Tension in the alveoli — an important factor in lung compliance. The alveoli are lined with a thin film of liquid, and it is at the interface between this liquid film and the alveolar air that surface tension develops. This is because the forces that hold the liquid film molecules together are stronger than those that hold the air molecules together. In the alveoli, excess surface tension causes the liquid film to contract, making lung inflation more difficult.

The pressure in the alveoli (which are modeled as spheres with open airways projecting from them) can be predicted using Laplace's law ($P=2T/r$ or pressure = 2 · surface tension/radius). If the surface tension were equal throughout the lungs, the alveoli with the smallest radii would have the greatest pressure, and this would cause them to empty into the larger alveoli and lower pressure (Fig. 22–16).

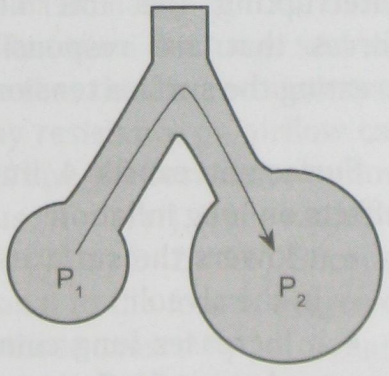


Fig. 22–16. The law of Laplace illustrating the effect of the radius on the pressure and movement of gases in the alveolar structures.

On the draw air moves from P_1 with a small radius and higher pressure to P_2 with its larger radius and lower pressure.

Assuming the lungs was lined by a simple electrolyte solution estimates indicate that lungs recoil should be so great that lungs theoretically should not be able to inflate. The reason this does not occur is because of special surface tension-lowering molecules, called *surfactant*, that line the inner surface of the alveoli.

Surfactant is a complex mixture of lipoproteins (largely phospholipids) and small amounts of carbohydrates that is synthesized in the type II alveolar cells. The surfactant molecule has two ends: a hydrophobic (water-insoluble) tail and a hydrophilic (water-soluble) head (Fig. 22–17). The hydrophilic head of the surfactant molecule attaches to the liquid molecules and the hydrophobic tail to the gas molecules, interrupting the intermolecular forces that are responsible for creating the surface tension.

Surfactant exerts 4 important effects on lung inflation:

- it lowers the surface tension in the alveoli;
- it increases lung compliance and ease of inflation;
- it provides for stability and more even inflation of the alveoli. This promotes stability among alveoli of different sizes by decreasing the

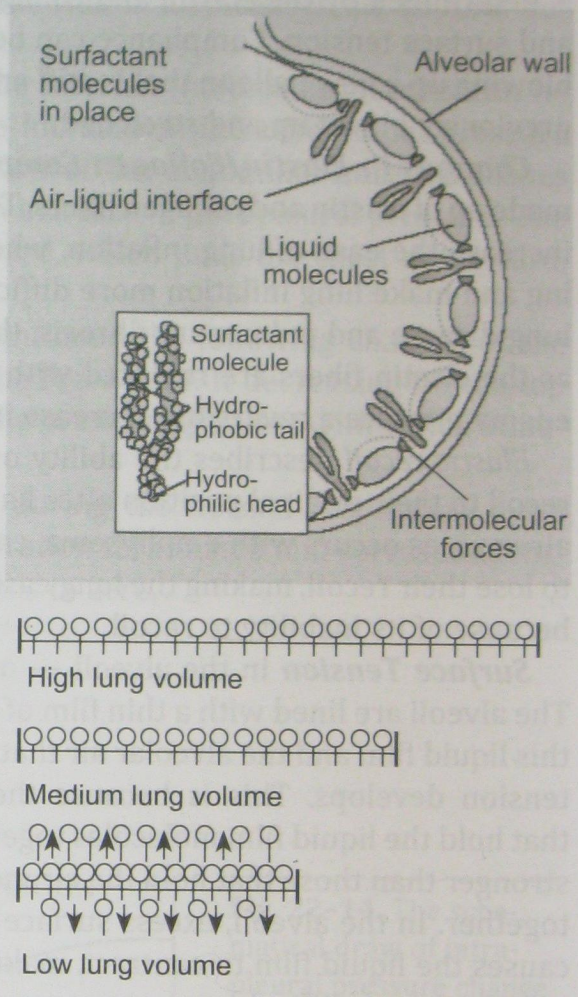


Fig. 22–17. Alveolar wall depicting surface tension resulting from the intramolecular forces in the air–liquid film interface; the surfactant molecule with its hydrophobic tail and hydrophilic head; and its function in reducing surface tension by disrupting the intermolecular forces (on the top). The surface concentration of surfactant molecules at high, medium, and low lung volumes (on the bottom).

tendency of small alveoli to collapse (very important during the expiration) and decreases the tendency to develop atelectasis (very important during the inspiration);
- it assists in preventing pulmonary edema by keeping the alveoli dry.

Without surfactant, lung inflation would be extremely difficult, requiring an intrapleural pressure of –20 to –30 mm Hg, compared with the –3 to –5 mm Hg pressure that normally is needed. The surfactant molecules are more densely packed in the small alveoli than in larger alveoli, where the density of the molecules is less. Therefore, surfactant reduces the surface tension more effectively in the small alveoli, which have the greatest tendency to collapse, providing for stability and more even distribution of ventilation. Surfactant also helps to keep the alveoli dry and prevent pulmonary edema. This is because water is pulled out of the pulmonary capillaries into the alveoli when increased surface tension causes the alveoli to contract.

The type II alveolar cells that produce surfactant do not begin to mature until the 26th to 28th week of gestation; consequently, many premature infants have difficulty producing sufficient amounts of surfactant. This can lead to alveolar collapse and severe respiratory distress. This condition, called *infant respiratory distress syndrome*, is the single most common cause of respiratory disease in premature infants. Surfactant dysfunction also is possible in the adult.

Airway Resistance

The volume of air that moves into and out of the air exchange portion of the lungs is directly related to the pressure difference between the lungs and the atmosphere and inversely related to the resistance that the air encounters as it moves through the airways. The effects of airway resistance on airflow can be illustrated using *Poiseuille's law*. According to this law, the resistance to flow is inversely related to the fourth power of the radius ($R = 1/r^4$). If the radius is reduced by one half, the resistance increases 16-fold ($2 \cdot 2 \cdot 2 \cdot 2 = 16$). Because the resistance of the airways is inversely proportional to the fourth degree of the radius, small changes in airway caliber, such as those caused by pulmonary secretions or bronchospasm, can produce a marked increase in airway resistance (Fig. 22–18).

Airway resistance is also affected by lung volumes, being less during inspiration than during expiration. This is because elastic-type fibers connect the outside of the airways to the surrounding lung tissues. As a result, these airways are pulled open as the lungs expand during inspiration, and they become

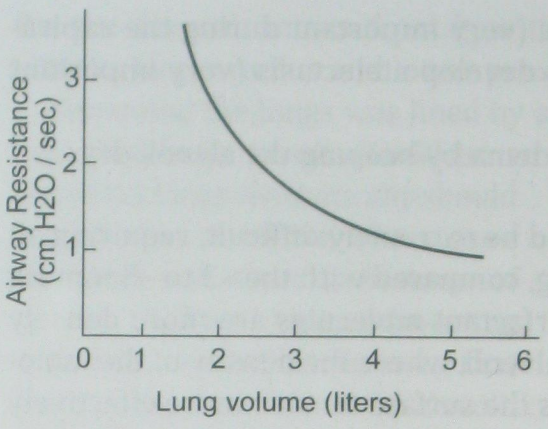


Fig. 22–18. Dependence of airway resistance upon the value of airway diameter.

narrower as the lungs deflate during expiration. This is one of the reasons that persons with conditions that increase airway resistance, such as bronchial asthma, usually have less difficulty during inspiration than during expiration.

Airway Compression. Airflow through the collapsible airways in the lungs depends on the distending airway (intrapulmonary) pressures that hold the airways open and the external (intrapleural or intrathoracic) pressures that surround and compress the airways. The difference between these two pressures (airway pressure minus intrathoracic pressure) is called the *transpulmonary pressure*. For airflow to occur, the distending pressure inside the airways must be greater than the compressing pressure outside the airways (Fig. 22–19).

Airway potency and airflow in the nonrigid airways of the lungs rely on a transpulmonary pressure gradient in which airway pressure is greater than intrapleural pressure. (Part B) Airway resistance normally produces a drop in

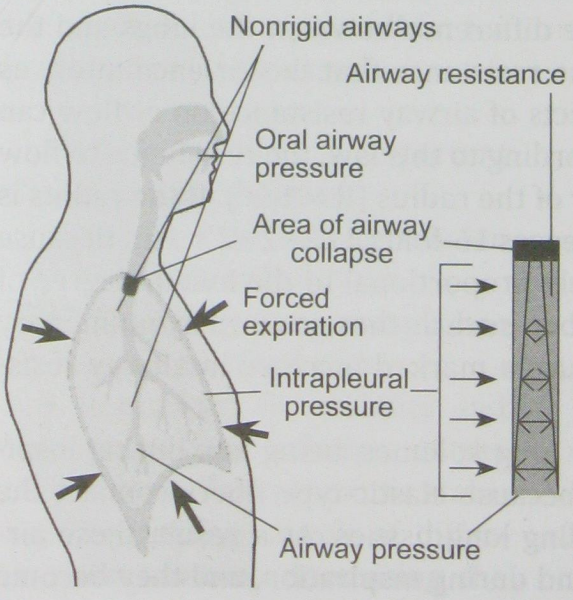


Fig. 22–19. Mechanism that limits maximal expiratory flow rate. (Part A). Airway potency and airflow in the nonrigid airways of the lungs rely on a transpulmonary pressure gradient in which airway pressure is greater than intrapleural pressure. (Part B) Airway resistance normally produces a drop in airway pressure as air moves out of the lungs. The increased intrapleural pressure that occurs with forced expiration produces airway collapse in the nonrigid airways at the point where intrapleural pressure exceeds airway pressure.

airway pressure as air moves out of the lungs. The increased intrapleural pressure that occurs with forced expiration produces airway collapse in the nonrigid airways at the point where intrapleural pressure exceeds airway pressure.

During forced expiration, the transpulmonary pressure is decreased because of a disproportionate increase in the intrathoracic pressure compared with airway pressure. The resistance that air encounters as it moves out of the lungs causes a further drop in airway pressure. If this drop in airway pressure is sufficiently great, the surrounding intrathoracic pressure will compress the collapsible airways (*i.e.*, those that lack cartilaginous support), causing airflow to be interrupted and air to be trapped in the alveoli (Fig. 22–19). Although this type of airway compression usually is seen only during forced expiration in persons with normal respiratory function, it may occur during normal breathing in persons with lung disease.

Chapter 23. Gas exchange and gases transportation by blood

The primary functions of the lungs are oxygenation of the blood and removal of carbon dioxide. Pulmonary gas exchange is conventionally divided into three processes:

1) *ventilation* or the flow of gases into and out of the alveoli of the lungs,
2) *perfusion* or flow of blood in the adjacent pulmonary capillaries,
3) *diffusion* or transfer of gases between the alveoli and the pulmonary capillaries.

The efficiency of gas exchange requires that alveolar ventilation occur adjacent to perfused pulmonary capillaries.

1. VENTILATION

Ventilation refers to the exchange of gases in the respiratory system. There are two types of ventilation: pulmonary and alveolar.

Pulmonary ventilation refers to the total exchange of gases between the atmosphere and the lungs.

The minute ventilation (or pulmonary ventilation) is the volume of gas expired per minute. By convention, the dot above the symbol here and for other terms means rate or quantity per unit time

$$\dot{V}_E = (volume\ per\ breath)(breaths\ per\ \min) = V_T(f)$$

$$\dot{V}_E = (500\,ml)(12\ to\ 14/\min) = 6\ to\ 7\,L/\min$$

where V_T = tidal volume and f = respiratory frequency in breaths/min.

$\dot{V}_E$ is determined by collecting the gas expired in one minute and measuring its volume (corrected to body temperature and atmospheric pressure with the gases fully saturated with water vapor).

Alveolar ventilation is the exchange of gases within the gas exchange portion of the lungs. Ventilation requires a system of open airways and a pressure difference that moves air into and out of the lungs. It is affected by body posi-

tion and lung volume as well as by disease conditions that affect the heart and respiratory system.

The distribution of ventilation between the base (bottom) and apex (top) of the lung varies with *body position* and the *effects of gravity* on intrapleural pressure. Compliance reflects the change in volume that occurs with a change in intrapleural pressure. It is less in fully expanded alveoli, which have difficulty accommodating more air, and greater in alveoli that are less inflated. In the seated or standing position, gravity exerts a downward pull on the lung, causing intrapleural pressure at the apex of the lung to become more negative. As a result, the alveoli at the apex of the lung are more fully expanded and less compliant than those at the base of the lung. The same holds true for lung expansion in the dependent portions of the lung in the supine or lateral position. In the supine position, ventilation in the lowermost (posterior) parts of the lung exceeds that in the uppermost (anterior) parts. In the lateral position (*i.e.*, lying on the side), the dependent lung is better ventilated.

The distribution of ventilation also is affected by *lung volumes*. During full inspiration (high lung volumes) in the seated or standing position, the airways are pulled open and air moves into the more compliant portions of the lower lung. At low lung volumes, the opposite occurs. At functional residual capacity (Fig. 23–8), the pleural pressure at the base of the lung exceeds airway pressure, compressing the airways so that ventilation is greatly reduced. In contrast, the airways in the apex of the lung remain open, and this area of the lung is well ventilated.

2. PERFUSION

The term *perfusion* is used to describe the flow of blood through the pulmonary capillary bed. The primary functions of the pulmonary circulation are to perfuse or provide blood flow to the gas exchange portion of the lung and to facilitate gas exchange. The pulmonary circulation serves several important functions in addition to gas exchange. It filters all the blood that moves from the right to the left side of the circulation; it removes most of the thromboemboli that might form; and it serves as a reservoir of blood for the left side of the heart.

The gas exchange function of the lungs requires a continuous flow of blood through the respiratory portion of the lungs. Deoxygenated blood enters the lung through the pulmonary artery, which has its origin in the right side

of the heart and enters the lung at the hilus, along with the primary bronchus. The pulmonary arteries branch in a manner similar to that of the airways. The small pulmonary arteries accompany the bronchi as they move down the lobules and branch to supply the capillary network that surrounds the alveoli (Fig. 22-8). The oxygenated capillary blood is collected in the small pulmonary veins of the lobules, and then it moves to the larger veins to be collected in the four large pulmonary veins that empty into the left atrium.

Distribution of Blood Flow. As with ventilation, the distribution of pulmonary blood flow is affected by *body position* and *gravity*. In the upright position, the distance of the upper apices of the lung above the level of the heart may exceed the perfusion capabilities of the mean pulmonary arterial pressure (appr. 12 mm Hg). Therefore, blood flow in the upper part of the lungs is less than that in the base or bottom part of the lungs. In the supine position, the lungs and the heart are at the same level, and blood flow to the apices and base of the lungs becomes more uniform.

3. MATCHING OF VENTILATION AND PERFUSION

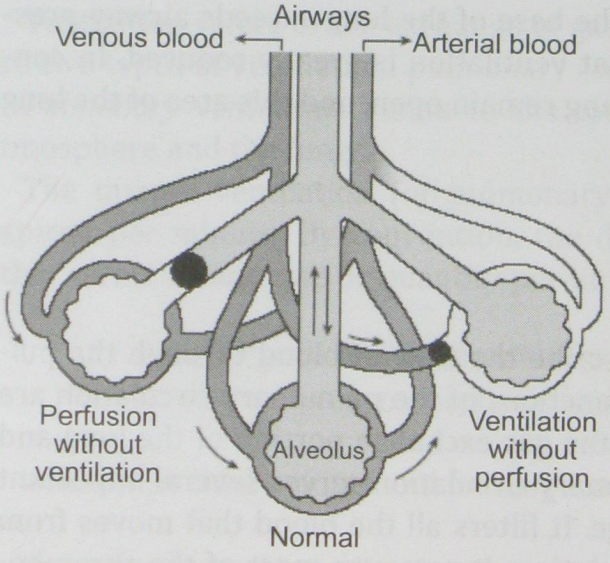


Fig. 23–1. Matching of ventilation and perfusion. Normal matching of ventilation and perfusion (center); perfusion without ventilation (i.e., shunt; left); ventilation without perfusion (i.e., dead air space; right).

The gas exchange properties of the lung depend on matching ventilation and perfusion, ensuring that equal amounts of air and blood are entering the respiratory portion of the lungs (Fig. 23–1).

There are two factors that may interfere with the matching of ventilation and perfusion: (1) dead air space and (2) shunt.

3.1. Dead Air Space

Dead space refers to the air that must be moved with each breath but does not participate

in gas exchange. The movement of air through dead space contributes to the work of breathing but not to gas exchange.

There are two types of dead space: that contained in the conducting airways, called the *anatomic dead space*, and that contained in the respiratory portion of the lung, called the *alveolar dead space*.

The volume of *anatomic airway dead space* is fixed at appr. 150 to 200 ml depending on body size. The anatomic dead space is the volume of gas occupying the conducting airways (the upper airways including the nasopharynx and oropharynx through the terminal bronchioles) in which no gas exchange occurs (Fig. 23–2). This is in contrast to the volume of gas located in the respiratory bronchioles through the alveoli where rapid gas exchange occurs. A good estimation of the anatomic airway dead space obtained from autopsy studies is *2.2 ml* per *1 kg* of the subject's *ideal body weight*. Thus, the normal average adult male weighing 70 kg would have an estimated anatomic airway dead space of (70 kg) · (2.2 ml/kg) = 154 ml.

Alveolar dead space, normally about 5 to 10 ml, constitutes alveolar air that does not participate in gas exchange. When alveoli are ventilated but deprived of blood flow, they do not contribute to gas exchange and thereby constitute alveolar dead space. In other words, alveolar dead space is wasted ventilation in the sense that with each tidal breath inspired air reaches the alveolar space but *contributes only minimally* to O_2 uptake or CO_2 elimination by the lung.

Panel A of the Fig. 23–3 below depicts the normal situation. The balloons represent alveoli in contact with pulmonary blood with high PCO_2 (systemic

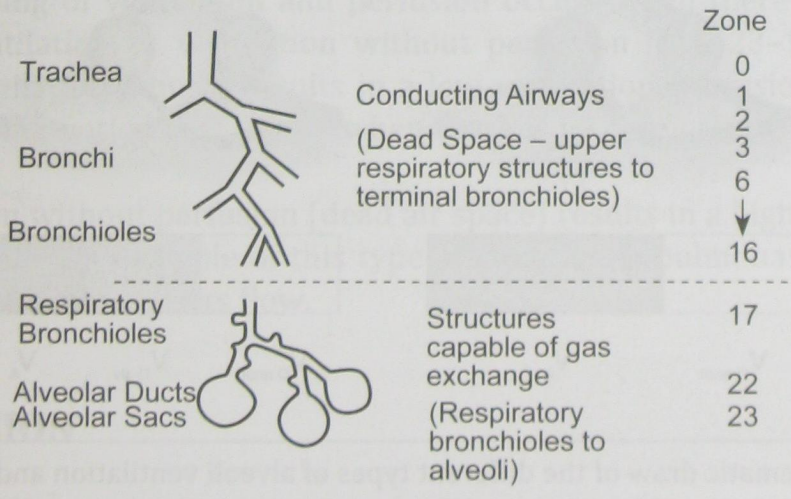


Fig. 23–2. Schematical draw of the dead space.

venous blood). The intensity of shading corresponds to the CO_2 concentration. CO_2 diffuses into the alveoli and the expired tidal volume is composed of anatomic dead space with no CO_2 and alveolar ventilation with CO_2 concentration equal to that of the alveoli.

In *Panel* B there is relative *over-ventilation* of the alveolus on the right (or, more correctly as drawn, under-perfusion of pulmonary blood flow to the alveolus on the right). Here, less CO_2 diffuses into the alveolus and the CO_2 concentration is decreased in relation to the alveolus on the right. The expired tidal volume is composed of anatomic dead space with no CO_2, alveolar ventilation with CO_2 concentration equal to that of the normal alveolus, and the *alveolar dead space*. Dead alveolar ventilation with a CO_2 concentration that is less than that present in the normal alveolus.

Only when alveolar ventilation is closely matched to pulmonary capillary blood flow does *efficient* gas exchange occur. Thus, ventilation to perfusion ratios in *excess* of the normal value results in *wasted* ventilation. The subject must do work to transport inspired air to anatomic dead space yet only small quantities or no PO_2 is cleared from the systemic venous blood in contact with these alveoli. When gas exchange is efficient for the whole lung, the arterial PO_2 and PCO_2 values *closely approximate* the PO_2 and PCO_2 of the mixed alveolar air; the latter is obtained by collecting the "expired" alveolar air (air exhaled near end of expiration).

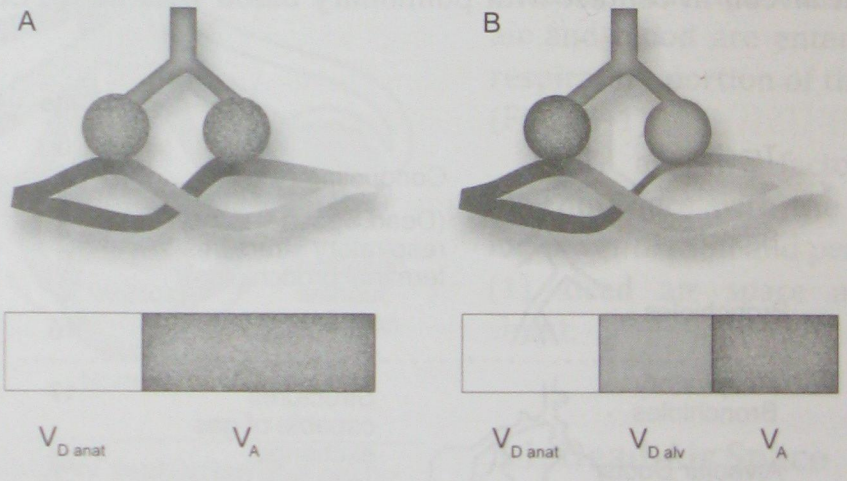


Fig. 23–3. Schematic draw of the different types of alveoli ventilation and composition of the expired volume. Abbreviations: VDanat — anatomic dead space, VA — alveolar ventilation, VDalv — alveolar dead space.

Logically, the *total dead space* is equal to the sum of the anatomical dead space and the alveolar dead space:

The *physiologic dead space* includes the anatomic dead space plus alveolar dead space. In persons with normal respiratory function, physiologic dead space is about the same as anatomic dead space. Only in lung disease does physiologic dead space increase.

3.2. Shunt

Shunt refers to blood that moves from the right to the left side of the circulation without being oxygenated. There are two types of shunts: physiologic and anatomic. In a *physiologic shunt*, there is mismatching of ventilation and perfusion, resulting in insufficient ventilation to provide the oxygen needed to oxygenate the blood flowing through the alveolar capillaries.

Physiologic shunting of blood usually results from destructive lung disease that impairs ventilation or from heart failure that interferes with movement of blood through sections of the lungs. In an *anatomic shunt*, blood moves from the venous to the arterial side of the circulation without moving through the lungs.

3.3. Mismatching of Ventilation and Perfusion

Mismatching of ventilation and perfusion occurs when there is perfusion without ventilation or ventilation without perfusion (Fig. 23–1). Perfusion without ventilation (shunt) results in a low ventilation–perfusion ratio. This is the type of situation that occurs when there is incomplete expansion of the lung.

Ventilation without perfusion (dead air space) results in a high ventilation–perfusion ratio. An example of this type of situation is pulmonary embolism, when a blood clot obstructs flow.

4. DIFFUSION

Diffusion refers to the movement of gases in the alveoli and across the alveolar capillary membrane. Diffusion of gases in the lung is affected by

the *difference in the pressure* of gas across the membrane, the surface area that is available for diffusion, and the thickness of the alveolar capillary membrane through which the gas must pass.

Recall that partial pressure or gas tension can be determined by measuring barometric pressure and the fractional concentration (F) of the gas (Dalton's law). At sea level, PO_2 is 160 mm Hg (760 mm Hg · 0.21). FO_2 does not change with altitude, which means that the percentage of O_2 in the atmosphere is essentially the same at 30,000 feet (about 9,000 m) as it is as sea level. Therefore, the decreased PO_2 at an altitude that makes it difficult to breathe is due to a decrease in the PB, not to a decrease in FO_2 (Fig. 23–4).

Abbreviations: P_B — barometric pressure, F_{IO2} — the fractional concentration of inspired O_2.

Oxygen is taken up by blood in the lungs and is transported to the tissues. *Oxygen uptake* is the transfer of oxygen from the alveolar spaces to the blood in the pulmonary capillaries. Gas uptake is determined *by three factors*:

- the diffusion properties of the alveolar-capillary membrane,
- the partial pressure gradient,
- pulmonary capillary blood flow.

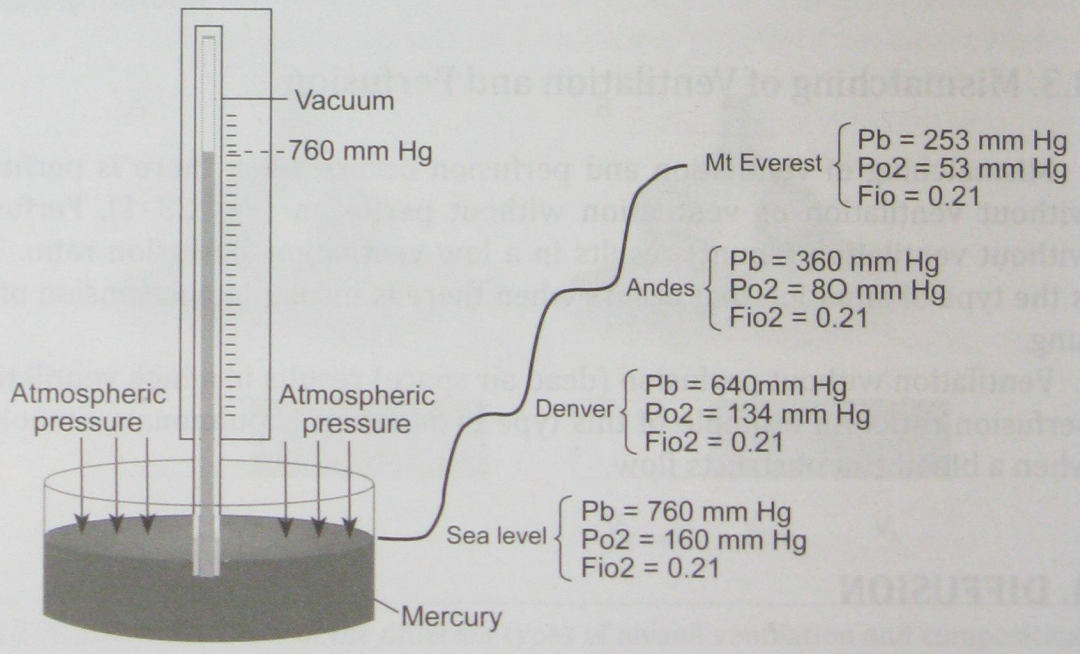


Fig. 23–4. Changes in O_2 tension with altitude.

Very important is that gases diffusion is a function of the partial pressure difference of the individual gases. For example, oxygen diffuses across the alveolar-capillary membrane because of the difference in PO_2 between the alveoli and pulmonary capillaries (Fig. 23–5). The partial pressure difference for oxygen is referred to as the *oxygen diffusion gradient*; in the normal lung, the *initial* oxygen diffusion gradient, P_AO_2 (102 mm Hg) minus PvO_2 (40 mm Hg), *is 62 mm Hg*. The *initial* diffusion gradient across the alveolar-capillary membrane for carbon dioxide ($PvCO_2$ — P_ACO_2) *is about 6 mm Hg*, which is much smaller than that of oxygen (Fig. 23–5).

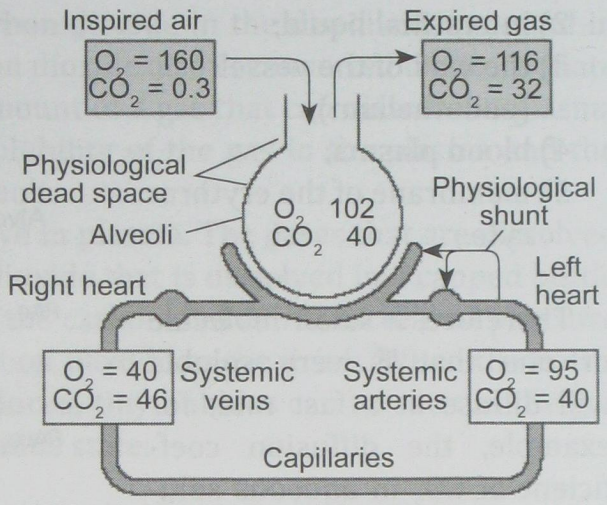


Fig. 23–5. Partial pressures of O_2 and CO_2 in the lungs and systemic circulation.

The following important issue that gases exposed to a liquid such as blood plasma should move into the liquid and exist in a dissolved state. The dissolved gases also exert a partial pressure. A gas will continue to dissolve in the liquid until the partial pressure of the dissolved gas equals the partial pressure above the liquid. *Henry's law* states that at equilibrium, the amount of gas dissolved in a liquid at a given temperature is directly proportional to the partial pressure and the solubility of the gas. Henry's law only accounts for the gas that is physically dissolved and not for chemically combined gases (e.g., oxygen bound to hemoglobin).

Gas diffusion in the lungs can be described by *Fick's law*, which states that the volume of gas diffusing per minute across a membrane is directly proportional to the membrane surface area, the diffusion coefficient of the gas and the partial pressure difference of the gas and inversely proportional to membrane thickness (Fig. 23–6).

The composition of the aero-hematic barrier if gases have to move through it in both directions is important. It consists if to move from the lungs towards the blood from the following functional components (Fig. 23–6):

1) alveolar epithelium;

2) interstitial liquid;
3) the wall of the vessels (endothelium);
4) blood plasma;
5) membrane of the erythrocyte.

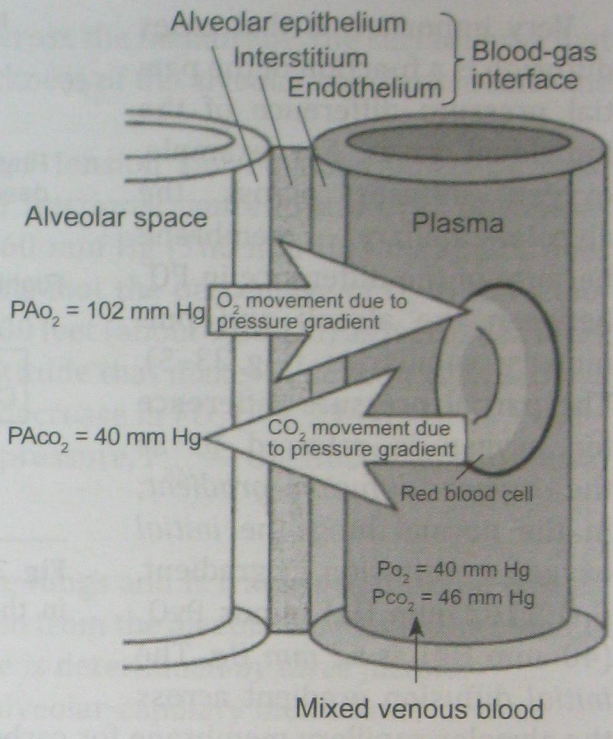


Fig. 23–6. The composition aero-hematic barrier and the routes of diffusion for O2 and CO_2.

Therefore, a small molecule or one that is very soluble will diffuse at a fast rate; for example, the diffusion coefficient of CO_2 in aqueous solutions is about 20 times greater than that of O_2 because of its higher solubility, even though it is a larger molecule than O_2.

Fick's law states that the rate of gas diffusion is inversely related to membrane thickness. This means that the diffusion of a gas will be halved if membrane thickness is doubled. Fick's law also states that the rate of diffusion is directly proportional to surface area. If two lungs have the same oxygen diffusion gradient and membrane thickness but one has twice the alveolar-capillary surface area, the rate of diffusion will differ by 2-fold.

Under steady state conditions, appr. 250 ml of O_2 per minute are transferred to the pulmonary circulation (VO_2) while 200 ml of CO_2 per minute are removed (VCO_2). The ratio VCO_2/VO_2 is the *respiratory exchange ratio* and, in this case, is 0.8.

5. GAS TRANSPORT BY BLOOD

The lungs enable inhaled air to come in contact with blood flowing through the pulmonary capillaries so that exchange of gases between the external environment and the internal environment of the body can occur. The lungs restore the oxygen content of the arterial blood and remove carbon dioxide from the venous blood.

The blood carries oxygen and carbon dioxide in the dissolved state and in combination with hemoglobin. Carbon dioxide also is converted to bicarbonate and transported in that form. The amount of a gas that can dissolve in plasma is determined by two factors: the solubility of the gas in the plasma and the partial pressure of the gas in the alveoli.

Oxygen and carbon dioxide dissolve in plasma. The gases that are dissolved in plasma are similar to the carbon dioxide that is dissolved in a capped bottle of a carbonated drink. In the case of the carbonated drink, increased pressure is used to increase the dissolved carbon dioxide to the drink. When the bottle cap is removed and the pressure reduced, tiny bubbles can be seen as the gas moves from the dissolved to the gaseous state.

5.1. Oxygen Transport

Oxygen is transported in two forms: (1) in chemical combination with hemoglobin and (2) in the dissolved state. The hemoglobin in red blood cells serves as a transport vehicle for oxygen. It binds oxygen in the pulmonary capillaries and releases it in the tissue capillaries. As oxygen moves into or out of the red blood cells, it dissolves in the plasma. It is the dissolved form of oxygen that leaves the capillary, crosses cell membranes, and participates in cell metabolism. Only approximately 1 % of the oxygen in the blood is carried in the dissolved state; the remainder is carried in combination with hemoglobin. The oxygen content of the blood includes the oxygen carried by hemoglobin and dissolved oxygen.

Hemoglobin Transport. Hemoglobin is a highly efficient carrier of oxygen, and approximately 98 % to 99 % of the oxygen used by body tissues is carried in this manner. Hemoglobin with bound oxygen is called *oxyhemoglobin*, and when oxygen is removed, it is called *deoxygenated* or *reduced hemoglobin*.

Each gram of hemoglobin carries approximately 1.34 ml of O_2 when it is fully saturated. This means that a person with a hemoglobin of 150 g/l carries 201 ml of O_2 per 1 llter of blood ($1.34 \cdot 150 = 201$). In the lungs, oxygen moves across the alveolar-capillary membrane, through the plasma, and into the red blood cell, where it forms a loose and reversible bond with the hemoglobin molecule. In normal lungs, this process is rapid, so that even with a fast heart rate, the hemoglobin is almost completely saturated with O_2 during the short time it spends in the pulmonary capillaries.

The oxygenated hemoglobin is transported in the arterial blood to the peripheral capillaries, where the oxygen is released and made available to the tissues

for use in cell metabolism. As the oxygen moves out of the capillaries in response to the needs of the tissues, the hemoglobin saturation, which usually is approximately 95 % to 97 % as the blood leaves the left side of the heart, drops to approximately 75 % as the mixed venous blood returns to the right side of the heart.

Dissolved Oxygen. The partial pressure of oxygen (PO_2) represents the level of dissolved oxygen in plasma. The amount of gas that can be dissolved in a liquid depends on the solubility of the gas and its pressure. Plasma O_2 solubility is fixed and very small. For every 1 mm Hg of PO_2 present in the alveoli, 0.003 ml of O_2 becomes dissolved in 100 ml of plasma. This means that at a normal alveolar PO_2 of 100 mm Hg, the blood carries only 0.3 ml of dissolved O_2 in each 100 ml of plasma. This amount is very small compared with the amount that can be carried in an equal amount of blood when O_2 is attached to hemoglobin.

Although the amount of O_2 carried in plasma under normal conditions is small, it can become a life-saving mode of transport in carbon monoxide poisoning, when most of the hemoglobin sites are occupied by carbon monoxide and are unavailable for transport of oxygen. The use of a hyperbaric chamber, in which 100 % oxygen can be administered at high atmospheric pressures, increases the amount of O_2 that can be carried in the dissolved state.

5.2. Oxygen-Hemoglobin Dissociation

Oxygen that remains bound to hemoglobin cannot participate in tissue metabolism. The efficiency of the oxygen dissociation transport system depends on the ability of the hemoglobin molecule to bind oxygen in the lungs and release it, as it is needed in the tissues. The *affinity* of hemoglobin refers to its capacity to bind oxygen. Hemoglobin binds oxygen more readily when its affinity is increased and releases it more readily when its affinity is decreased.

As described in Chapter # 13, the hemoglobin molecule is composed of four polypeptide chains bound to an iron-containing heme group. Because oxygen binds to the iron atom, each hemoglobin molecule can attach and carry up to four oxygen molecules. Binding sites on Hb have different affinities for O_2. Also, the affinity of a site can and does change as oxygen is loaded or unloaded from the Hb molecule and as the chemical composition of the plasma changes.

When oxygen is bound to all four of the heme groups, the hemoglobin molecule is said to be fully *saturated*. Hemoglobin is partially saturated when it contains only one, two, or three molecules of oxygen. This could be explained with the help of the Fig. 23–7.

Site 1 — O_2 usually remains attached under physiologic conditions. Only sites 2, 3 and 4 need to be considered under physiologic conditions.

Site 2 has the greatest affinity of the three sites and thus requires the lowest PO_2 for an O_2 to remain attached. The minimum O_2 to keep an O_2 attached at this site is about 26 mm Hg. As long as the PO_2 of the plasma is 26 mm Hg or above, oxygen remains on this site. This site represents point P_{50} for the arterial blood that means that half of the hemoglodin is saturated with O_2.

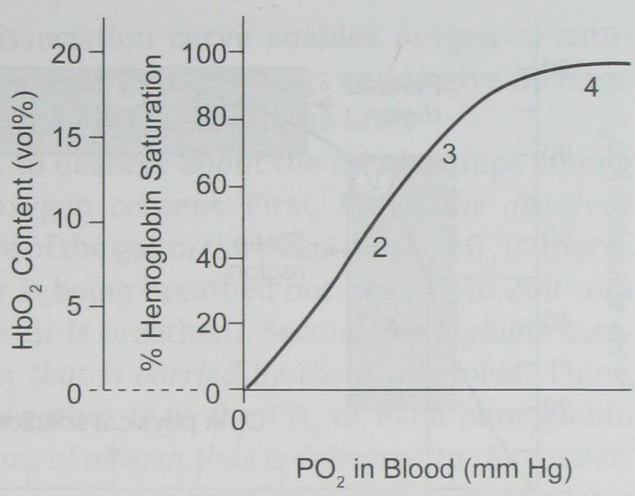


Fig. 23–7. Schematic drawing of the oxyhemoglobin dissociation curve with the plotted numbers that corresponded to the sites of oxygen attachment to Hb molecule.

Site 3. When O_2 attaches to the second site all sites gain affinity but the ranking does not change. The third site has less affinity than the second. It thus requires a greater PO_2 to maintain the attachment of oxygen, about 40 mm Hg. At this point the systemic arterial blood has hemoglodin that saturated with O_2 on 75 % (resting state).

Site 4 has the least affinity for oxygen and thus requires the greatest PO_2 to maintain attachment. This is just above 100 mm Hg. At a PO_2 of 100 mm Hg, the hemoglobin is appr. 97 % saturated (Fig. 23–7).

The relation between the oxygen carried in combination with hemoglobin and the PO_2 of the blood is described by the *oxyhemoglobin dissociation curve.* Fig. 23–7 illustrates this curve with the numbered above sites for oxygen binding to hemoglobin molecule. This graph also defines as *oxyhemoglobin equilibrium curve* and has S-shaped curve over a range of arterial oxygen tensions from 0 to 100 mm Hg (Fig. 23–8). The shape of the curve results because the hemoglobin affinity for oxygen increases progressively as blood PO_2 increases.

The x axis depicts the PO_2; the left y axis, hemoglobin saturation, and the right y axis, the oxygen content or total amount of oxygen in the blood, including the dissolved oxygen and that carried by the hemoglobin.

The shape of the *oxyhemoglobin dissociation curve* reflects several important physiological advantages. This curve has a flat top portion representing

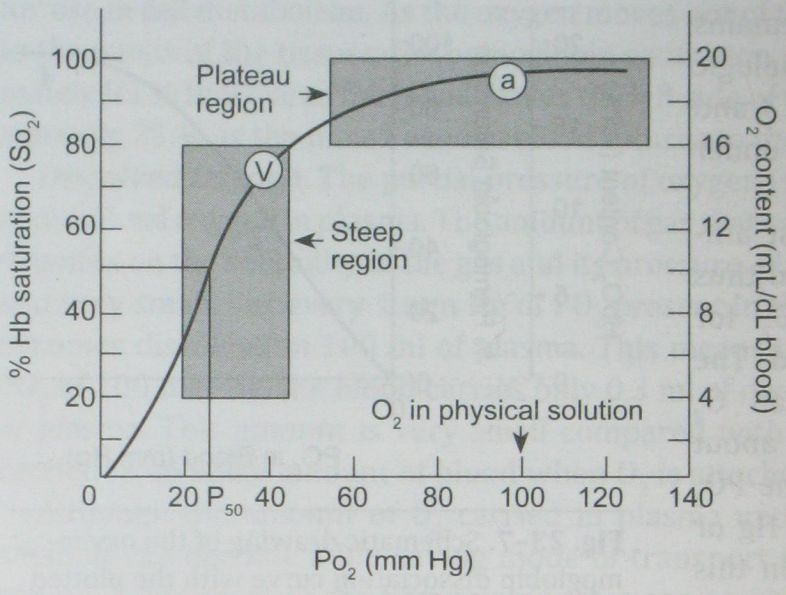


Fig. 23–8. An oxyhemoglobin dissociation curve. The dashed line indicates amount of oxygen dissolved in the plasma.
Abbreviations: a — arterial, v — venous; SO_2 — oxygen saturation; and P_{50} — partial pressure of O_2 required to saturate 50 % of the hemoglobin with oxygen.

binding of oxygen by the hemoglobin in the lungs and a steep portion representing its release into the tissue capillaries.

The *plateau region* of the curve is the *loading phase* in which oxygen is loaded onto hemoglobin to form oxyhemoglobin in the pulmonary capillaries. The plateau region illustrates how oxygen saturation and content remain fairly constant despite wide fluctuations in alveolar PO_2. For example, if P_AO_2 were to rise from 100 to 120 mm Hg, hemoglobin would become only slightly more saturated (97 to 98 %).

For this reason, O_2 content cannot be raised significantly by hyperventilation. Increasing the alveolar PO_2 above this level does not increase the hemoglo-bin saturation. Even at high altitudes, when the partial pressure of O_2 is considerably decreased, the hemo-globin remains relati-vely well saturated. For example, at 60 mm Hg PO_2, the hemoglobin is still approximately 89 % saturated.

The steep *unloading phase* of the curve — between 60 and 40 mm Hg — allows large quantities of oxygen to be released or unloaded from hemoglobin in the tissue capillaries where a lower capillary PO_2 prevails. This portion of the curve shows that there is considerable transfer of oxygen from hemoglobin to the tissues with only a small drop in PO_2, thereby ensuring a gradient for oxygen to move into body cells. The tissues normally remove approximately 5 ml of oxygen per 100 ml of blood, and the hemoglobin of mixed venous blood is approximately 75 % saturated as it returns to the right side of the heart.

The S-shaped oxyhemoglobin dissociation curve enables oxygen to saturate hemoglobin under high partial pressures in the lungs and to give up large amounts of oxygen with small changes in PO_2 at the tissue level.

There are three important things to observe about the relationships among PO_2, hemoglobin saturation, and oxygen content. First, *PO_2 is the dissolved oxygen.* It reflects the partial pressure of the gas in the lung (*i.e.*, the PO_2 is approximately 100 mm Hg when room air is being breathed but can rise to 200 mm Hg or higher when oxygen-enriched air is breathed). Second, *hemoglobin saturation reflects the amount of oxygen that is carried by the hemoglobin.* Third, it is *the oxygen content of the blood* rather than the PO_2 or even hemoglobin saturation that *determines the amount of oxygen that is delivered to the tissues.*

Hemoglobin can be regarded as an oxygen buffer system that regulates oxygen pressure in the tissues. Hemoglobin affinity for oxygen must change with the metabolic needs of the tissues. This change is represented by a shift to the right or left in the dissociation curve, as shown in Fig. 23–9.

Hemoglobin's affinity for O_2 is influenced by pH, CO_2 concentration, and temperature. Hemoglobin binds O_2 more strongly under conditions of increased pH (alkalosis), decreased CO_2 concentration, and decreased body temperature and releases it more readily under conditions of decreased pH (acidosis), increased CO_2 concentration, and fever. Conditions that decrease affinity and favor unloading of oxygen reflect the level of tissue metabolism and need for oxygen.

The P_{50} represents the PO_2 at a hemoglobin of 50 %. It can be used to determine whether the dissociation curve has shifted to the right or to the left. A *shift to the right* indicates that the tissue PO_2 is greater for any given level of hemoglobin saturation and repre-

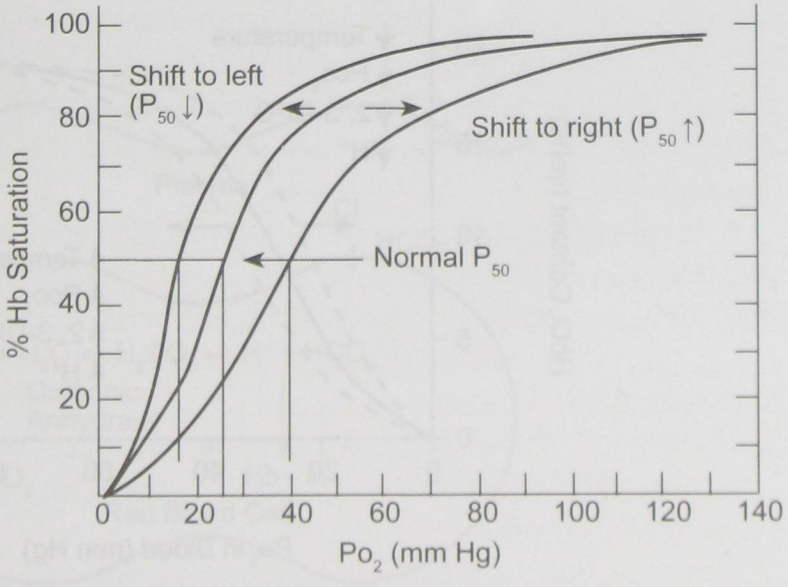


Fig. 23–9. The oxyhemoglobin dissociation curve shifts to the left and right sides affect the P_{50}.

sents reduced affinity of the hemoglobin for oxygen at any given PO_2. It usually is caused by conditions such as fever or acidosis or by an increase in PCO_2, which reflects increased tissue metabolism. High altitude and conditions such as pulmonary insufficiency, heart failure, and severe anemia also cause the oxygen dissociation curve to shift to the right.

A *shift to the left* in the oxygen dissociation curve represents enhanced affinity of hemoglobin for oxygen and occurs in situations associated with a decrease in tissue metabolism, such as alkalosis, decreased body temperature, and decreased carbon dioxide levels.

All the above mentioned factors influencing on the O_2 dissociation curve shifts are summarized on Fig. 23–10 with only one additional substance — 2,3-diphosphoglycerate (2,3-DPG) — metabolic intermediate locates inside the red blood cells that also affects the affinity of hemoglobin for oxygen.

5.3. Carbon Dioxide Transport

Carbon dioxide is transported in the blood in three forms: as dissolved plasma carbon dioxide (10 %), as bicarbonate ions in the plasma and in the erythrocytes (60 %), and as carbamino proteins (30 %). It should be stressed

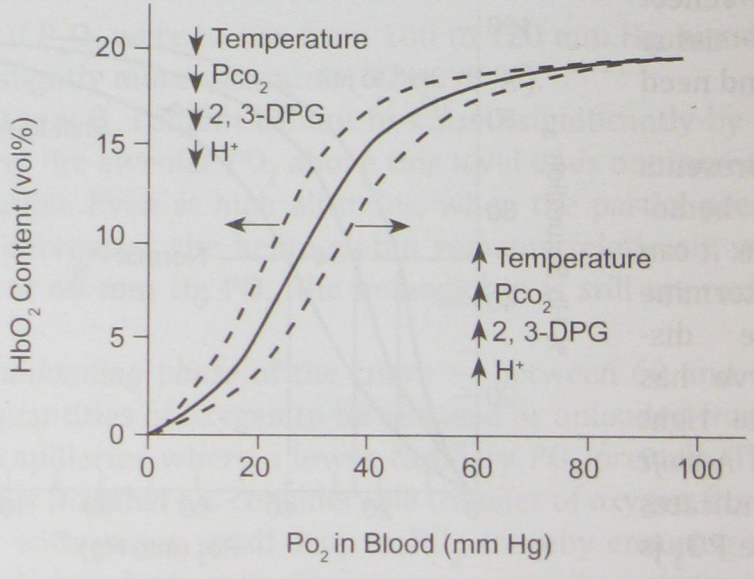


Fig. 23–10. Factors that could induce the oxyhemoglobin dissociation curve shifts to the left and right sides.

also that acid-base balance is influenced by the amount of dissolved CO_2 and the bicarbonate level in the blood.

As CO_2 is formed during the metabolic process, it diffuses out of cells into the tissue spaces and then into the capillaries (Fig. 23–11). The high PCO_2 in the tissues drives CO_2 into the blood, but only a small amount stays as dissolved CO_2 in the plasma. Most of the CO_2 diffuses into the red blood cell where it forms either carbonic acid (H_2CO_3) or *carbaminohemoglobin*. In the red cell, carbonic acid is formed in the following reaction:

$$\mathbf{CO_2 + H_2O \rightarrow H_2CO_3 \rightarrow H^+ + HCO_3^-}$$

The process is catalyzed by an enzyme called *carbonic anhydrase* which greatly increases the rate of the reaction. Carbonic anhydrase concentration 1000-fold higher inside the erythrocytes than in blood plasma. This is the main reason why CO_2 moves directly inside the BRC but doesn't stay in blood plasma.

Carbonic acid readily ionizes (HCO_3^-) and a hydrogen (H^+) ion (Fig. 23–11). The hydrogen ion that is generated combines with the hemoglobin, which is a powerful acid-base buffer.

HCO_3 leaves the red blood cells, and chloride diffuses in from the plasma to maintain electrical neutrality (Fig. 23–11). The chloride movement is known as the *chloride shift* and is facilitated by a *chloride-bicarbonate exchanger (anion exchanger)* in the red blood cell membrane.

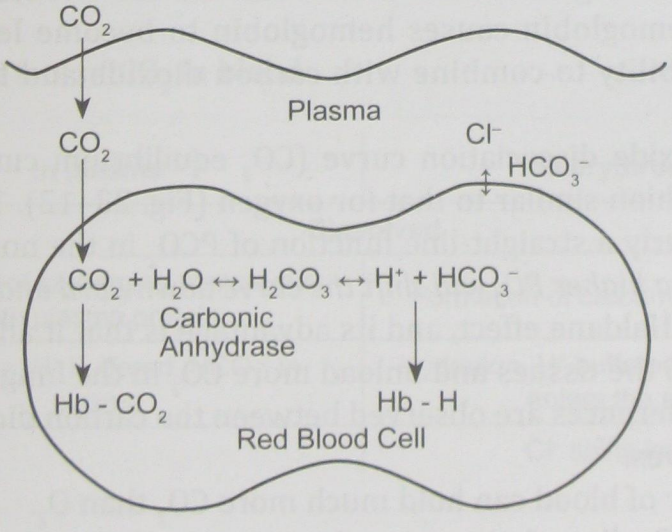


Fig. 23–11. The steps of the CO_2 conversion into bicarbonate in the systemic capillary.

The H^+ cannot readily move out because of the low permeability of the membrane to H^+. Most of the H+ is buffered by hemoglobin:

$$H^+ + HbO_2^- \rightarrow HHb + O_2.$$

As H^+ binds to hemoglobin, it decreases oxygen binding and shifts the oxyhemoglobin dissociation curve to the right. This promotes the unloading of oxygen from hemoglobin in the tissues and favors the carrying of carbon dioxide. In the pulmonary capillaries, the oxygenation of hemoglobin favors the unloading of carbon dioxide.

Carbaminohemoglobin is formed in red cells from the reaction of carbon dioxide with free amine groups (NH_2) on the hemoglobin molecule:

$$CO_2 + HbNH_2 \rightarrow HbNHCOOH$$

Deoxygenated hemoglobin can bind much more CO_2 in this way than oxygenated hemoglobin. Although major reactions related to CO_2 transport occur in the red cells, the bulk of the CO_2 is actually carried in the plasma in the form of bicarbonate.

The binding of carbon dioxide to hemoglobin is determined by the acidic nature of hemoglobin. Binding with carbon dioxide causes the hemoglobin to become a stronger acid. In the lungs, the highly acidic hemoglobin has a lesser tendency to form carbaminohemoglobin, and carbon dioxide is released from hemoglobin into the alveoli. In the tissues, the release of oxygen from hemoglobin causes hemoglobin to become less acid, thereby increasing its ability to combine with carbon dioxide and form carbaminohemoglobin.

A carbon dioxide dissociation curve (CO_2 equilibrium curve) can be constructed in a fashion similar to that for oxygen (Fig. 23–12). The CO_2 dissociation curve is nearly a straight-line function of PCO_2 in the normal arterial CO_2 range. Note that *a higher PO_2 will shift the curve downward and to the right.* This is known as the Haldane effect, and its advantage is that it allows the blood to load more CO_2 in the tissues and unload more CO_2 in the lungs.

Important differences are observed between the carbon dioxide and oxygen dissociation curves.

First, one liter of blood can hold much more CO_2 than O_2.

Second, the CO_2 dissociation curve is steeper and more linear and because of its shape large amounts of CO_2 can be loaded and unloaded from the blood

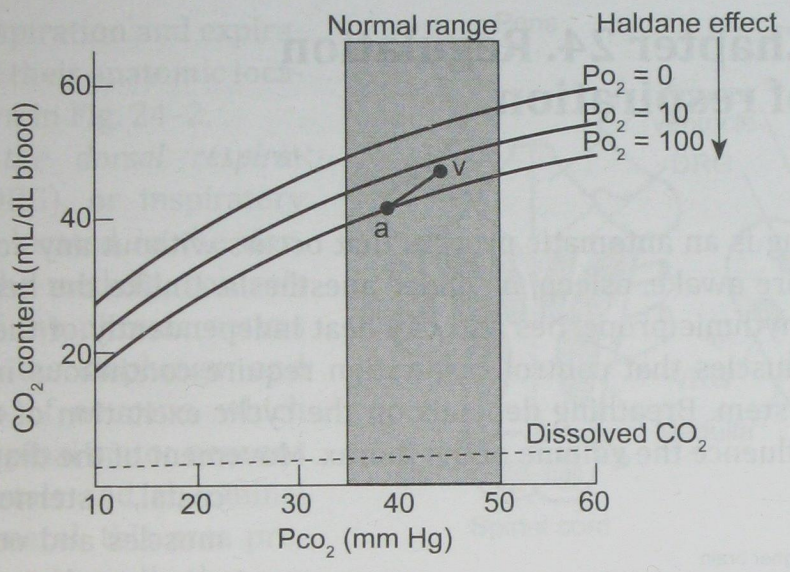


Fig. 23–12. Influence of oxygen on the carbon dioxide dissociation curve.
The dashed line indicates the amount dissolved in plasma: a — CO_2 content in arterial blood and v — CO_2 content in mixed venous blood.

with a small change in PCO_2. This is important not only in gas exchange and transport but also in the regulation of acid-base balance.

For convenience, the various fates of CO_2 in the plasma and red cells are summarized in table 23–1.

Table 23–1. The fate of CO_2 in blood

	In plasma	In erythrocytes
	Dissolved	
1	Formation of carbamino compounds with plasma protein	Formation of carbamino-hemoglobin
2	Hydration, H^+ buffered, HCO_3^- in plasma	Hydration, H^+ buffered, 70 % of HCO_3^- enters the plasma
3		Cl^- shifts into cells

Chapter 24. Regulation of respiration

Breathing is an automatic process that occurs without any conscious effort while we are awake, asleep, or under anesthesia. Unlike the heart, which has inherent rhythmic properties and can beat independently of the nervous system, the muscles that control respiration require continuous input from the nervous system. Breathing depends on the cyclic excitation of many muscles that can influence the volume of the thorax. Movement of the diaphragm, intercostal, sternocleidomastoid muscles and other accessory muscles that control ventilation is integrated by neurons located in the pons and medulla. These neurons are collectively referred to as the *respiratory center* (Fig. 24–1).

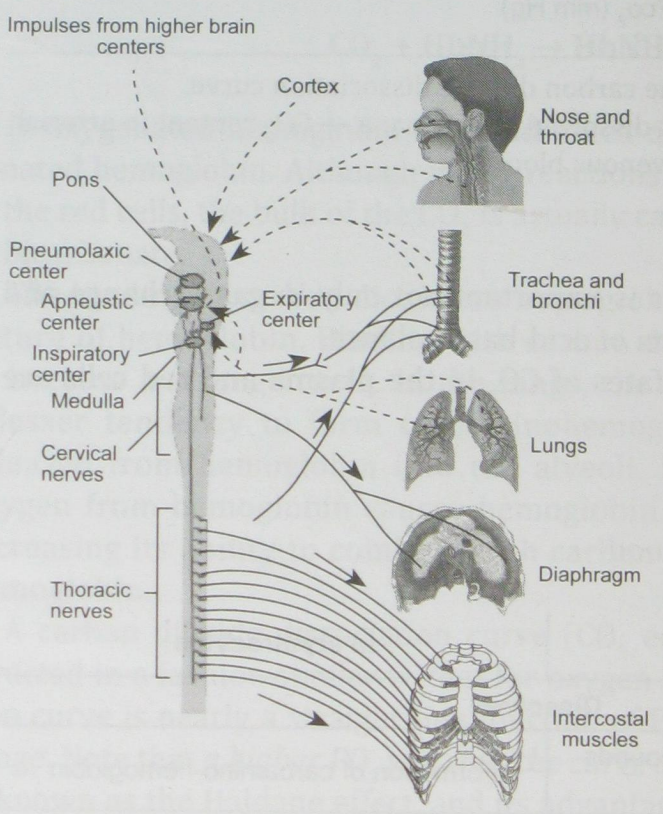


Fig. 24–1. Schematic representation of activity in the respiratory center. Impulses traveling over afferent neurons (red dashed lines) communicate with central neurons, which activate efferent neurons that supply the muscles of respiration.

1. RESPIRATORY CENTER

The central pattern for the basic breathing rhythm has been localized to fairly discrete areas in the medulla oblongata that discharge action potentials in a phasic pattern with respiration. The medullary respiratory center consists of two that discharge action potentials in a phasic pattern with respiration. The medullary respiratory center consists of two groups of neurons involved

in initiating inspiration and expiration found and their anatomic locations are shown in Fig. 24–2.

The first, *the dorsal respiratory group* (DRG), or inspiratory area which is located in the dorsal region of the medulla, controls the activity of the phrenic nerves that innervate the diaphragm and another group of neurons, which innervate the spinal motor neurons of the intercostals and abdominal muscles. In general, this area predominantly contains cells that are active during inspiration.

The second group of neurons, the *ventral respiratory group* (VRG), is concerned predominantly with the expiratory phase of respiration. It should be stressed that VRG contains both inspiration- and expiration-related neurons.

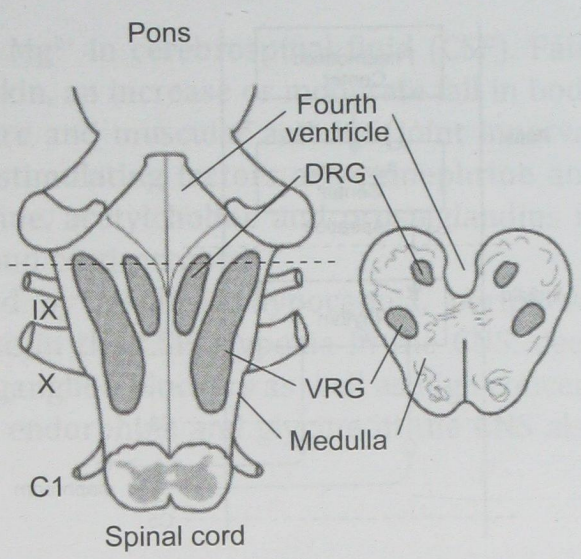


Fig. 24–2. The general locations of the dorsal respiratory group (DRG) and ventral respiratory group (VRG). These drawings show the dorsal aspect of the medulla oblongata and a cross section in the region of the fourth ventricle. Abbreviations: C1 — first cervical nerve, X — vagus nerve, IX — glossopharyngeal nerve.

Both groups contain cells projecting ultimately to the bulbospinal motor neuron pools. The DRG and VRG are bilaterally paired, but cross-communication enables them to respond in synchrony; as a consequence, respiratory movements are symmetric.

The pacemaker properties of the respiratory center result from the cycling of the two groups of respiratory neurons in the pons: the *pneumotaxic center* in the upper pons and the *apneustic center* in the lower pons (Fig. 24–3). These two groups of neurons contribute to the function of the respiratory center in the medulla. The apneustic center has an excitatory effect on inspiration, tending to prolong inspiration.

The pneumotaxic center switches inspiration off, assisting in the control of respiratory rate and inspiratory volume. Brain injury that damages the connection between the pneumotaxic and apneustic centers results in an irregular breathing pattern that consists of prolonged inspiratory gasps interrupted by expiratory efforts.

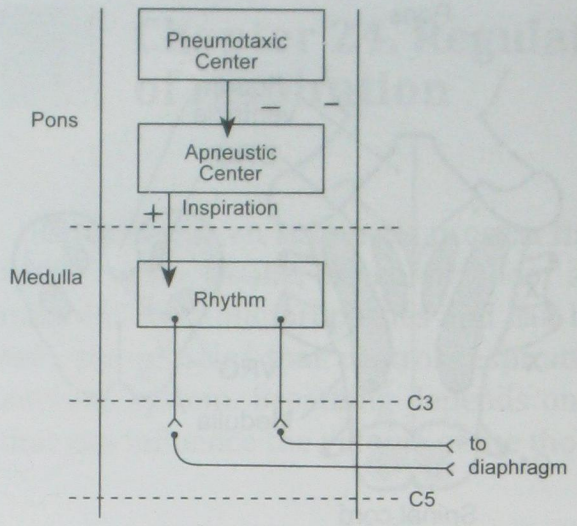


Fig. 24–3. The main features involved into the central control of ventilation. Pattern of interaction between the pneumotaxic and apneustic center is indicated also.

2. NEURONAL PATTERN OF RESPIRATION

As it was stressed earlier, DRG neurons are excitated during inspiration. Additionally, it was established that DRG neurons contain two subtypes of neurons — inspiratory-alpha (Iα-) and inspiratory-beta (Iβ-) neurons.

Iα-neurons excitate and generate activity only during the inspiration.

Iβ-neurons excitate together with the Iα-neurons excitation but generate activity after the arrest of the Iα-neurons activity.

Iα-neurons are typically inspiratory neurons which excitation transmits also towards to Iβ- neurons. Iβ-neurons activation in turn results to Iα-neurons activity inhibition according to the mechanism of the feedback path.

After Iα-neurons are activated and started to generate their activity they stimulate expiratory neurons work. These neurons generate activity only during the expiration and thus responsible for the neuronal pattern of expiration.

Numerous factors influence the respiratory neurons in the medulla oblongata (Fig. 24–4). Hence, ventilation is increased by acidosis, hypercapnia,

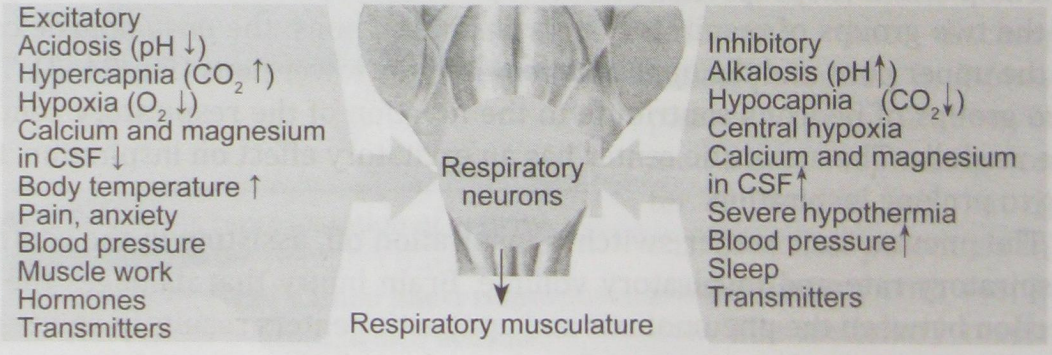


Fig. 24–4. Modulators of respiratory neurons.

hypoxia, and a decrease of Ca^{2+} and Mg^{2+} in cerebrospinal fluid (CSF). Pain, intensive cold or heat stimuli to the skin, an increase or moderate fall in body temperature, a drop in blood pressure and muscular activity (joint innervation) all increase ventilation. Other stimulating factors are epinephrine and norepinephrine in the blood, histamine, acetylcholine and prostaglandins in the CNS, progesterone, testosterone and corticotropin.

Conversely, ventilation is reduced by alkalosis, hypocapnia, peripheral hyperoxia and Ca^{2+} and Mg^{2+} increase in the CSF. Hypoxia in the CNS, deep hypothermia, rise in blood pressure, ganglion blockers as well as high concentrations of atropine, catecholamines, endorphins and glycine in the CNS also diminish ventilation.

3. CONTROL OVER RESPIRATION

The control of breathing has automatic and voluntary components. The automatic regulation of ventilation is controlled by input from two types of sensors or receptors: chemoreceptors and lung receptors.

Chemoreceptors monitor blood levels of oxygen, carbon dioxide, and pH and adjust ventilation to meet the changing metabolic needs of the body. The level of alveolar ventilation is driven mainly from the input of specific chemoreceptors to the CNS. The stronger the stimulation of these receptors, the greater the level of alveolar ventilation. Chemoreceptors monitor the chemical composition of body fluids. In this system there are receptors that respond to pH, PCO_2 and PO_2.

Lung receptors monitor breathing patterns and lung function.

Voluntary regulation of ventilation integrates breathing with voluntary acts, such as speaking, blowing, and singing. These voluntary acts, initiated by the motor and premotor cortex, cause temporary suspension of automatic breathing. The automatic and voluntary components of respiration are regulated by afferent impulses that come to the respiratory center from a number of sources. Afferent input from higher brain centers is evidenced by the fact that a person can consciously alter the depth and rate of respiration. Fever, pain, and emotion exert their influence through lower brain centers. Vagal afferents from sensory receptors in the lungs and airways are integrated in the dorsal area of the respiratory center.

3.1. Chemoreceptors

There are two types of chemoreceptors: central and peripheral.

The most important chemoreceptors for sensing changes in blood carbon dioxide content are the *central chemoreceptors.* These receptors are located in chemosensitive regions near the respiratory center in the medulla and are bathed in cerebrospinal fluid. Although the central chemoreceptors monitor carbon dioxide levels, the actual stimulus for these receptors is provided by hydrogen ions in the CSF.

The CSF is separated from the blood by the blood-brain barrier, which permits free diffusion of CO_2 but not bicarbonate or hydrogen ions (Fig. 24–5). The CO_2 combines rapidly with water to form carbonic acid (H_2CO_3), which dissociates into hydrogen and bicarbonate ions:

$$CO_2 + H_2O \rightarrow H_2CO_3 \rightarrow H^+ + HCO_3^-$$

The carbon dioxide content in the blood regulates ventilation through its effect on the pH of the extracellular fluid of the brain. The central chemoreceptors are extremely sensitive to short-term changes in carbon dioxide (Fig. 24–5).

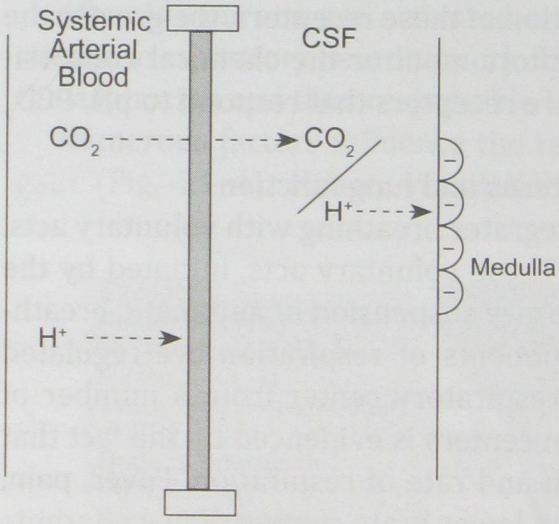


Fig. 24–5. Schematic interrelation between the central chemoreceptors and the systemic arterial blood.

An increase in plasma CO_2 levels increases ventilation, reaching its peak within a minute or so and then declining if the CO_2 level remains elevated.

Thus, persons with chronically elevated levels of carbon dioxide no longer have a response to this stimulus for increased ventilation but rely on the stimulus provided by a decrease in blood oxygen levels.

Normally the pH around the respiratory neurons or the pH in the CSF has a decisive influence on ventilation (Fig. 24–6, part 1). A shift in pH in the brain following rapid changes in PCO_2 is accentuated by the low buffering power of CSF (low protein concentration).

Because CO_2, but not HCO_3^- or H^+ quickly passes through the blood–CSF and blood–brain barriers, changes in CO_2 concentration in the blood result in very rapid adaptation of ventilation, while adaptation after changes in blood pH or blood HCO_3 — occurs only after a delay of several days.

If sudden metabolic acidosis occurs (Fig. 24–6, part 2), respiratory compensation will thus occur only slowly, usually within 12–24 hours (Fig. 24–6, parts 3, 4). The mechanism of adaptation may be the pumping of HCO_3^- into and out of the CSF. There are no central PO_2 receptors. Thus, treatment of a partly compensated respiratory acidosis, for example, by HCO_3^- infusion often leaves behind respiratory alkalosis (Fig. 24–6, part 5). Also, with a sudden fall of O_2 partial pressure in inspiratory air (at high altitude) ventilation is not immediately and adequately raised.

The onset of hyperventilation leads to hypocapnia, and the resulting intracerebral alkalosis will then transiently inhibit any further rise in ventilation. Complete adaptation of breathing to a reduced O_2 supply requires an increase in renal HCO_3^- excretion with subsequent decrease in HCO_3^- concentration in plasma and (after a delay) in CSF (Fig. 24–6, part 6).

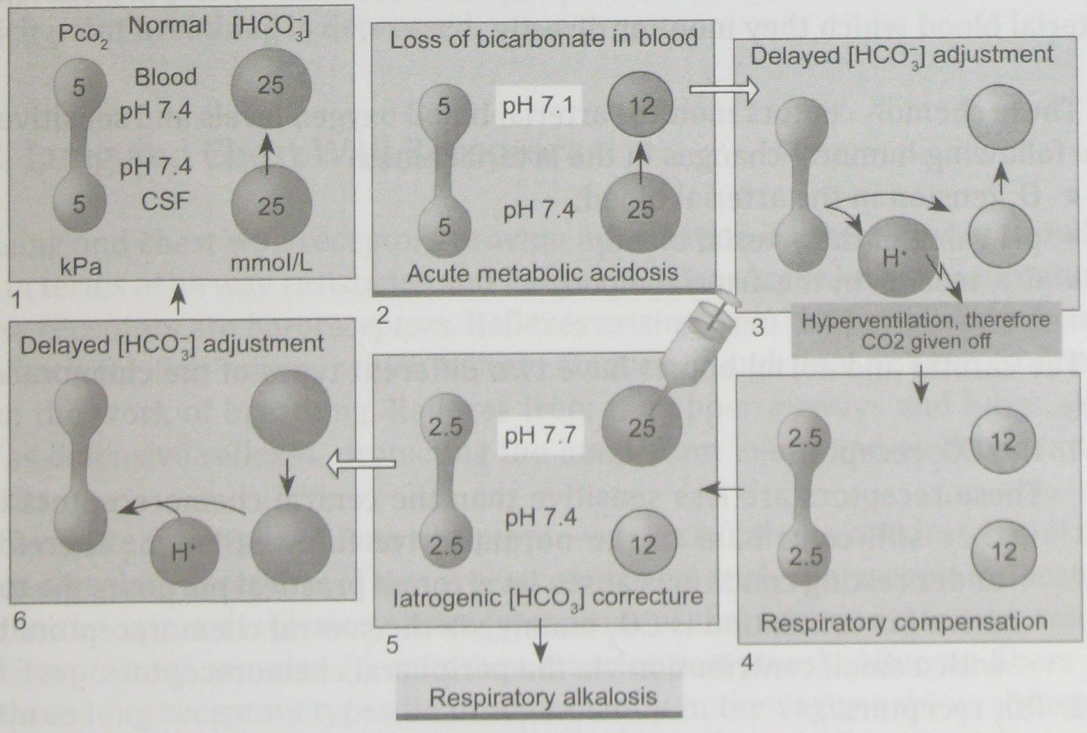


Fig. 24–6. Effects of blood-brain barrier penetration on respiratory rate.

Barbiturates and chronic respiratory failure decrease the sensitivity of the respiratory neurons to pH or CO_2 in CSF. Lack of O_2 thus becomes the most important stimulus to breathing. In both cases the supply of O_2 enriched air leads to hypoventilation and respiratory acidosis. This response is increased by, for example, uremia or sleep. Because O_2 uptake varies within a wide range independently of alveolar ventilation (Fig. 24–4) breathing is stimulated only when there is a marked diminution in alveolar O_2 partial pressure and a fall in O_2 saturation in the blood. The resulting increase in ventilation will again cease as soon as O_2 saturation in the blood is normal; breathing is therefore irregular.

The *peripheral chemoreceptors* are located in the carotid and aortic bodies. *Carotid bodies* are small (appr. 2 mm wide) sensory organs located bilaterally near the bifurcations of the common carotid arteries near the base of the skull. Afferent nerves travel to the CNS from the carotid bodies in the glossopharyngeal nerves. *Aortic bodies* are located along the ascending aorta and are innervated by vagal afferents.

The carotid body is the most important peripheral chemoreceptor in humans. Because it receives the most blood per gram of tissue in the body and is so small it can meet al of its metabolic requirements for O_2 by utilizing the O_2 that is dissolved in the blood. The peripheral chemoreceptors are bathed in arterial blood which they monitor directly.

These chemoreceptors monitor arterial blood oxygen levels and sensitive to the following humoral changes in the arterial blood:

- O_2 tension in the arterial blood;
- pH value of the arterial blood;
- CO_2 tension in the arterial blood.

The carotid and aortal bodies have two different types of the chemoreceptors:

1. $H+/CO_2$ receptors.
   These receptors are less sensitive than the central chemoreceptors but they still contribute to the normal drive for ventilation. Therefore, under resting conditions at sea level for all practical purposes the total drive for ventilation is CO_2 mainly via the central chemoreceptors but with a small contribution via the peripheral chemoreceptors.
2. PO_2 receptors.
   The factor monitored by these receptors is PO_2 but not oxygen content. Because they respond to PO_2 they are actually monitoring dissolved

oxygen and not oxygen on hemoglobin. When systemic arterial PO_2 is close to normal (appr. equal to 100 mm Hg) or above normal, there is little if any stimulation of these receptors. Thus, they do not contribute to our normal drive for ventilation. They are strongly stimulated only by a dramatic decrease in systemic PO_2. They do not begin to fire significantly until the PaO_2 falls to 50-60 mm Hg. Under these conditions there is an increased drive for ventilation, and alveolar ventilation usually increases. In most situations where the systemic arterial PO_2 is dramatically reduced the main drive for ventilation is the low PO_2 stimulation of the peripheral chemoreceptors. Sensitivity to hypoxia increases with CO_2 retention.

Therefore, although the peripheral chemoreceptors also monitor carbon dioxide, they play a *much more important role in monitoring oxygen levels* comparing with the central chemoreceptors. These receptors exert little control over ventilation until the PO_2 has dropped below 60 mm Hg.

Unlike the central sensor, peripheral chemoreceptors are sensitive to rising arterial blood H^+ and falling PO_2. They alone cause the stimulation of breathing by hypoxia; hypoxia in the brain has little effect on breathing unless severe, at which point breathing is depressed.

3.2. Lung and Chest Wall Receptors

Lung and chest wall receptors provide information on the status of breathing in terms of airway resistance and lung expansion. According to their nature these receptors are *baroreceptors*. Reflexes arising from the periphery provide feedback for fine-tuning, which adjusts frequency and tidal volume to minimize the work of breathing. Reflexes from the upper airways and lungs also act as defensive reflexes, protecting the lungs from injury and environmental insults.

There are five types of lung receptors — stretch, irritant, and juxtacapillary (juxtaalveolar) receptors. They are accompanied with the receptors located within the upper respiratory tract (receptors of the upper respiratory tract) and respiratory muscles (respiratory muscles receptors). Afferent fibers of all three lung receptors types lie predominantly in the vagus nerves, although some pass with the sympathetic nerves to the spinal cord. The role of the sympathetic afferents is uncertain.

Stretch receptors are located in the smooth muscle layers of the conducting airways. They respond to changes in pressure in the walls of the airways and are slowly adaptating receptors. Slowly adapting receptors fire in proportion to applied airway transmural pressure, and their usual role is to sense lung volume. When the lungs are inflated, these receptors inhibit inspiration and promote expiration. Because of these two effects, inflating the lungs with a sustained pressure at the mouth terminates an inspiration in progress and prolongs the time before a subsequent inspiration occurs. This sequence is known as the *Hering-Breuer reflex* or *lung inflation reflex.*

The Hering-Breuer reflex probably plays a more important role in infants than in adults. In adults, particularly in the awake state, this reflex may be overwhelmed by more prominent central control. Because increasing lung volume stimulates slowly adapting receptors, which then excite the inspiratory off-switch, it is easy to see how they could be responsible for a feedback signal that results in cyclic breathing. Therefore, stretch receptors are important in establishing breathing patterns and minimizing the work of breathing by adjusting respiratory rate and tidal volume to accommodate changes in lung compliance and airway resistance.

The *irritant receptors* are sensory terminals of myelinated afferent fibers located between the airway epithelial cells. By nature they are *rapidly adapting receptors.* The definition — irritant receptors — is used because these nerve endings respond to irritation of the airways by touch or by noxious substances, such as cigarette smoke, inhaled dust, and cold air. Rapidly adapting receptors are stimulated also by histamine, serotonin, and prostaglandins released locally in response to allergy and inflammation. They are also stimulated by lung inflation and deflation but their firing rate rapidly declines when a volume change is sustained. Because of this rapid adaptation bursts of activity occur that are in proportion to the change of volume and the rate at which that change occurs.

These receptors are probably nearly inactive in normal quiet breathing. Based on what stimulates them, their role would seem to be to sense the onset of pathological events. As a general rule, these receptors stimulation causes excitatory responses such as coughing, gasping and prolonged inspiration time. Irritant receptors stimulation leads to airway constriction and a pattern of rapid, superficial breathing. This pattern of breathing probably protects respiratory tissues from the damaging effects of toxic inhalants. It also is thought that the mechanical stimulation of these receptors may ensure more uniform lung expansion by initiating periodic sighing and yawning. It is possible that

these receptors are involved in the bronchoconstriction response that occurs in some persons with bronchial asthma.

The *juxtacapillary* or *J-receptors* are located in the alveolar wall, close to the pulmonary capillaries. It is thought that these receptors sense lung congestion. Like irritant receptors J-receptors play a protective role. They are stimulated by lung injury, large inflation, acute pulmonary vascular congestion and certain chemical agents.

These receptors may be responsible for the rapid, superficial breathing bronchoconstriction, increased airway secretion and cardiovascular activity depression (bradycardia, hypotension) that occurs with pulmonary edema, pulmonary embolism, and pneumonia. Apnea (cessation of breathing) and a marked fall in systemic vascular resistance occur when they are stimulated acutely and severely.

Receptors of the upper respiratory tract react on both the mechanical and chemical irritations. They bear a resemblance to irritant receptors with different localization. Their response reaction when stimulated might be in a form of sneezing, coughing as well as to bronchoconstriction.

Respiratory muscles receptors — actually proprioceptors by nature. These receptors may play a role in breathing, particularly when more than quiet breathing is called for or when breathing efforts are opposed by increased airway resistance or reduced lung compliance. Muscle spindles are present in considerable numbers in the intercostal muscles but are rare in the diaphragm. It has been proposed that muscle spindles may adjust breathing effort by sensing the discrepancy between tensions of the intrafusal and extrafusal fibers of the intercostal muscles. If a discrepancy exists, information from the spindle receptor alters the contraction of the extrafusal fiber, thereby minimizing the discrepancy. This mechanism known as gamma-loop provides increased motor excitation when movement is opposed.

3.3. Cough Reflex

Coughing is a neurally mediated reflex that protects the lungs from the accumulation of secretions and from entry of irritating and destructive substances. It is one of the primary defense mechanisms of the respiratory tract. The cough reflex is initiated by receptors located in the tracheobronchial wall; these receptors are extremely sensitive to irritating substances and to the presence of excess secretions. Afferent impulses from these receptors are transmitted through the vagus to the medullary center, which integrates the cough response.

Coughing itself requires the rapid inspiration of a large volume of air (usually about 2.5 l), followed by rapid closure of the glottis and forceful contraction of the abdominal and expiratory muscles. As these muscles contract, intrathoracic pressures are elevated to levels of 100 mm Hg or more. The rapid opening of the glottis at this point leads to an explosive expulsion of air.

Many conditions can interfere with the cough reflex and its protective function. The reflex is impaired in persons whose abdominal or respiratory muscles are weak. This problem can be caused by disease conditions that lead to muscle weakness or paralysis, by prolonged inactivity or as an outcome of surgery involving these muscles. The cough reflex also is impaired when there is depressed function of the medullary centers in the brain that integrate the cough reflex. Interruption of the central integration aspect of the cough reflex can arise as the result of disease of this part of the brain or the action of drugs that depress the cough center.

Although the cough reflex is a protective mechanism, frequent and prolonged coughing can be exhausting and painful and can exert undesirable effects on the cardiovascular and respiratory systems and on the elastic tissues of the lungs. This is particularly true in young children and elderly persons.

Unit IX Gastrointestinal physiology

Chapter 25. Digestive tract functions. Digestion in the oral cavity. Swallowing

1. SHORT OVERVIEW OF THE GASTROINTESTINAL TRACT FUNCTIONAL ORGANIZATION

Structurally, the gastrointestinal tract is a long, empty tube with its lumen inside the body and its wall acting as an interface between the internal and external environments.

In the digestive tract food and other materials move slowly along its length as they are systematically broken down into ions and molecules that can be absorbed into the body. In the large intestine, unabsorbed nutrients and wastes are collected for later elimination.

For simplicity and understanding, the digestive system can be divided into four parts. *The upper part* — the mouth, esophagus and stomach — acts as an intake source and receptacle through which food passes and in which initial digestive processes take place.

The middle portion consists of the small intestine — the duodenum, jejunum, and ileum. Most digestive and absorptive processes occur in the small intestine.

The lower segment — the cecum, colon, and rectum — serves as a storage channel for the efficient elimination of waste.

The fourth part consists of the accessory organs — the teeth, tongue, salivary glands, liver, gallbladder and pancreas.

These structures produce digestive secretions that help dismantle foods and regulate the use and storage of nutrients.

Fig. 25–1 gives you an impression about the functional organization of the gastrointestinal tract.

Solid food is chewed and mixed with *saliva* which lubricates it and contains immunocompetent substances and enzymes. The esophagus rapidly transports the food bolus to the stomach. The lower esophageal sphincter opens only briefly to allow the food to pass. The proximal stomach mainly serves as a food reservoir. Its tone determines the rate at which food passes to the distal stomach, where it is further processed (chyme formation) and its proteins are partly broken down. The distal stomach (including the pylorus) is also responsible for portioning chyme delivery to the small intestine.

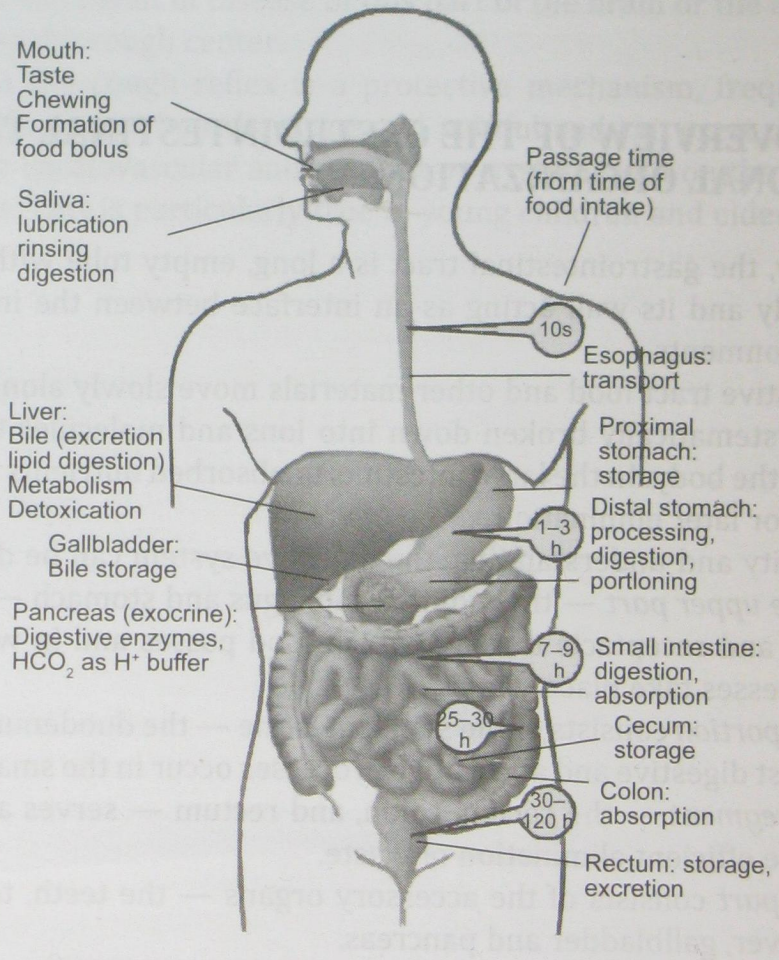


Fig. 25–1. Functions of the organs of the digestive tract. Time of food/chyme passage from the moment of food intake is indicated by yellow.

In the small intestine, enzymes from the pancreas and small intestinal mucosa break down the nutrients into absorbable components.

HCO_3^- in pancreatic juices neutralizes the acidic chyme. Bile salts in bile are essential for fat digestion. The products of digestion (monosaccharides, amino acids, dipeptides, monoglycerides and free fatty acids) as well as water and vitamins are absorbed in the small intestine.

Waste products to be excreted reach the feces via bile secreted by the liver. The large intestine is the last stop for water and ion absorption. It is colonized by bacteria and contains storage areas for feces (cecum, rectum).

We will emphasize these peculiarities of the oral cavity digestion. Three broad areas will be discussed: 1) motility, 2) secretion and 3) digestion and absorption.

2. MOUTH AND ESOPHAGUS MOTOR FUNCTION

In the mouth, food is mixed with saliva and propelled into the esophagus. Peristaltic waves in the esophagus move the food into the stomach.

The movement of foods and fluids in the gastrointestinal tract begins with *chewing*, in the case of solid foods, and *swallowing*. Chewing (*mastication*) begins the digestive process, it breaks up large food particles and mixes the food with the secretions of the salivary glands. This wetting and homogenizing action aids swallowing and subsequent digestion. Large food particles can be digested but they cause strong and often painful contractions of the esophageal musculature. Particles that are small tend to disperse in the absence of saliva and also make swallowing difficult because they do not form a bolus. The number of chews that is optimal depends on the food, but usually ranges from 20 to 25.

Resuming, one could classify the following functions of mastication: 1) chewing *breaks food into smaller pieces* which (i) makes it easier for the food to be swallowed, (ii) breaks off the undigestible cellulose coatings of fruits and vegetables and (iii) increases the surface area of the food particles, making it easier for them to be digested by the digestive enzymes; 2) chewing *mixes the food with salivary gland secretions* which (i) initiates the process of starch digestion by salivary alpha-amylase, (ii) initiates the process of lipid digestion by lingual lipase and (iii) lubricates and softens the bolus of food, making it easier to swallow; 3) chewing *brings food into contact with taster receptors* and *releases odors* that stimulate the olfactory receptors. The sensations generated by these receptors increase the pleasure of eating and initiate gastric secretions.

Although *chewing* usually is considered a voluntary act, it can be carried out involuntarily by a person who has lost the function of the cerebral cortex.

The *swallowing* (deglutition) reflex is a rigidly ordered sequence of events that results in the propulsion of food from the mouth to the stomach through the esophagus. Although swallowing is initiated as a voluntary activity, it becomes involuntary as food or fluid reaches the pharynx. Sensory impulses for the reflex begin at tactile receptors in the pharynx and esophagus and are integrated with the motor components of the response in an area of the reticular formation of the medulla and lower pons called the *swallowing center.*

The motor impulses for the oral and pharyngeal phases of swallowing are carried in the trigeminal (V), glossopharyngeal (IX), vagus (X) and hypoglossal (XII) cranial nerves and impulses for the esophageal phase are carried by the vagus nerve. These impulses are integrated in the nucleus of the tractus solitarius and the nucleus ambiguus. Diseases that disrupt these brain centers or their cranial nerves disrupt the coordination of swallowing and predispose an individual to food and fluid lodging in the trachea and bronchi, leading to risk of asphyxiation or aspiration pneumonia.

Swallowing consists of three phases: an oral (voluntary) phase, a pharyngeal phase and an esophageal phase. During the *oral phase*, the bolus is collected at the back of the mouth so the tongue can lift the food upward until it touches the posterior wall of the pharynx (Fig. 25–2, 1). At this point, the *pharyngeal phase* of swallowing is initiated. The soft palate is pulled upward, the palatopharyngeal folds are pulled together so that food does not enter the nasopharynx (Fig. 25–2, 2), the vocal cords are pulled together and the epiglottis is moved so that it covers the larynx. Respiration is inhibited, and the bolus is moved backward into the esophagus by constrictive movements of the pharynx. Although the striated muscles of the pharynx are involved in the second stage of swallowing, it is an involuntary stage.

The third phase of swallowing is the *esophageal stage*. As the upper esophageal sphincter opens (Fig. 25–2, 4) and food enters the esophagus (Fig. 25–2, 5) and stretches its walls, local and central nervous system reflexes that initiate peristalsis are triggered (Fig. 25–3).

The motor impulses for the *oral* and *pharyngeal* phases of swallowing are carried in the trigeminal (V), glossopharyngeal (IX), vagus (X) and hypoglossal (XII) cranial nerves. Impulses for the *esophageal* phase are carried by the vagus nerve.

There are two types of esophageal peristalsis: *primary* and *secondary*.

Primary peristalsis is controlled by the swallowing center in the brain stem and begins when food enters the esophagus.

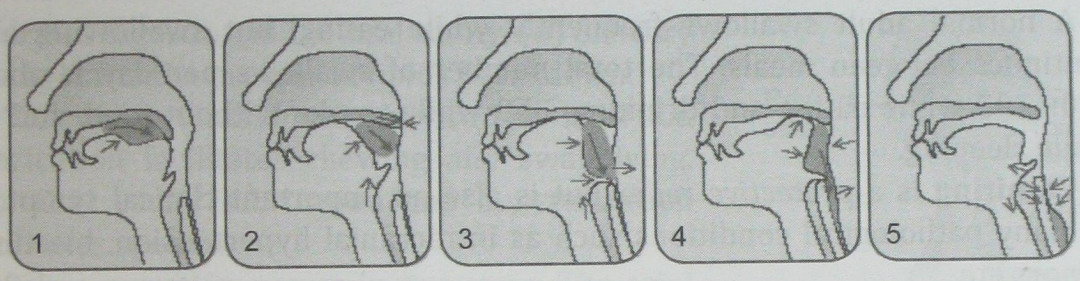


Fig. 25–2. The mechanics of swallowing.

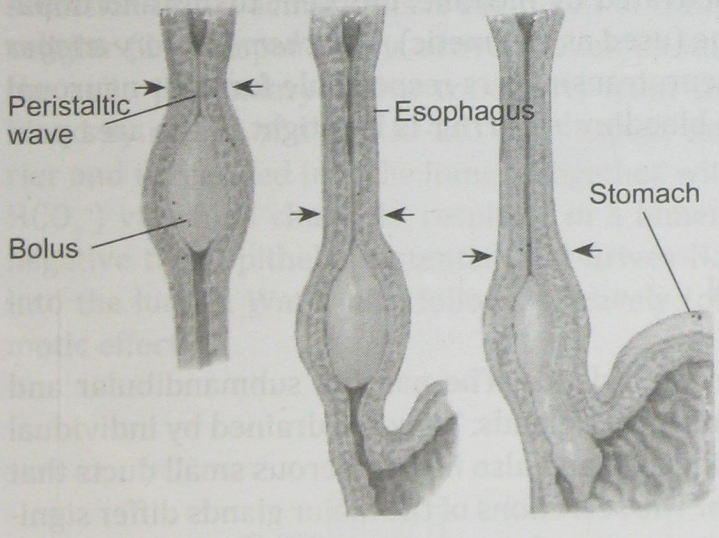


Fig. 25–3. The schematic representation of esophageal peristalsis.

Secondary peristalsis is partially mediated by smooth muscle fibers in the esophagus and occurs when primary peristalsis is inadequate to move food through the esophagus. Peristalsis begins at the site of distention and moves downward. Before the peristaltic wave reaches the stomach, the lower esophageal sphincter relaxes to allow the bolus of food to enter the stomach. The pressure in the lower esophageal sphincter normally is greater than that in the stomach, an important factor in preventing the reflux of gastric contents. Regurgitation would occur because the pressure in the abdominal cavity is greater than the pressure in the thoracic cavity as a result of respiratory movements. The lower esophageal sphincter must therefore remain closed until food is pushed through it by peristalsis into the stomach.

Swallowing is difficult if not impossible when the mouth is open, as anyone who has spent time in the dentist's chair feeling saliva collect in the throat is well aware.

A normal adult swallows frequently while eating, but swallowing also continues between meals. The total number of swallows per day is about 600: 200 while eating and drinking, 350 while awake without food, and 50 while sleeping.

Vomiting is a *protective reflex* but is also an important clinical symptom of many pathological conditions such as intracranial hypertension, bleeding, tumors etc.

The *vomiting center* is located in the medulla oblongata within the reticular formation. It is mainly controlled by chemosensors of the *area postrema* which is located on the bottom of the IV-th ventricle — this is called the *chemosensory trigger zone* which might be activated by nicotine, different toxins and dopamine agonists like apomorphine (used as an emetic). The *chemosensory trigger zone* cells have receptors for neurotransmitters responsible for their neuronal control. Importanet is that the bloodbrain barrier is less tight in the area postrema.

3. SALIVARY SECRETION

Saliva is secreted by the salivary glands. The parotid, submandibular and sublingual glands are the major salivary glands. They are drained by individual ducts into the mouth. The sublingual gland also has numerous small ducts that open into the floor of the mouth. The secretions of the major glands differ significantly. The parotid glands secrete saliva that is rich in water and electrolytes, whereas the submandibular and sublingual glands secrete saliva that is rich in mucin. There are also minor salivary glands located in the labial, palatine, buccal, lingual and sublingual mucosae.

Saliva has three functions. The first is *defensive*. (i) Saliva is rich in mucus which protects the oral mucosa and coats the food as it passes through the mouth, pharynx and esophagus. The sublingual and buccal glands produce only mucus-type secretions. (ii) Additionally, saliva regulates the temperature regime of the ingested foods. (iii) Saliva performs also *protective antimicrobial action*. The saliva cleans the mouth and contains the enzyme lysozyme which has an antibacterial action. (iv) Removing the food residual particles from the teeth saliva responsible for caries prevention. (v) Saliva pH=7.0 helps in acid substances neutralization.

The second saliva function is *digestive*. (i) Saliva contains ptyalin and alpha-amylase, which initiate the digestion of dietary starches, predominantly carbo-

hydrates digestion in the oral cavity. (ii) Salivary lipase starts the process of lipids breakdown.

The third function ifs *lubricative*. (i) Food lubrication facilitates chewing and swallowing. (ii) Oral cavity lubrication and moisturization facilitates speech.

Saliva secretion occurs in *two steps*: The acini (end pieces) produce *primary saliva* (Fig. 25–4) which has an electrolyte composition similar to that of plasma (Fig. 25–5). Primary saliva secretion in the acinar cells is the result of transcellular Cl^- transport: Cl^- is actively taken up into the cells (secondary active transport) from the blood by means of a Na^+- K^+-$2Cl^-$ cotransport carrier and is released into the lumen (together with HCO_3^-) via anion channels, resulting in a lumen-negative transepithelial potential that drives Na^+ into the lumen. Water also follows passively (osmotic effect).

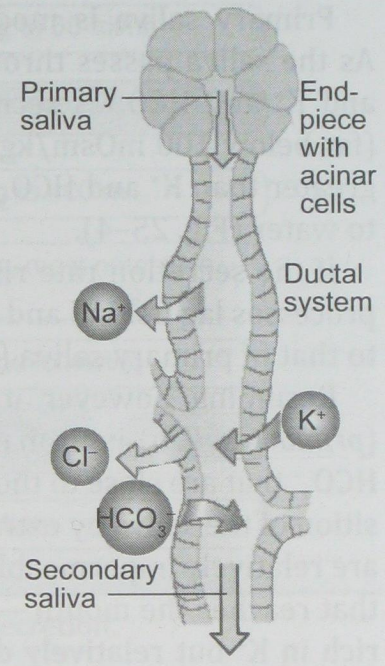


Fig. 25–4. The schematic representation of saliva secretion.

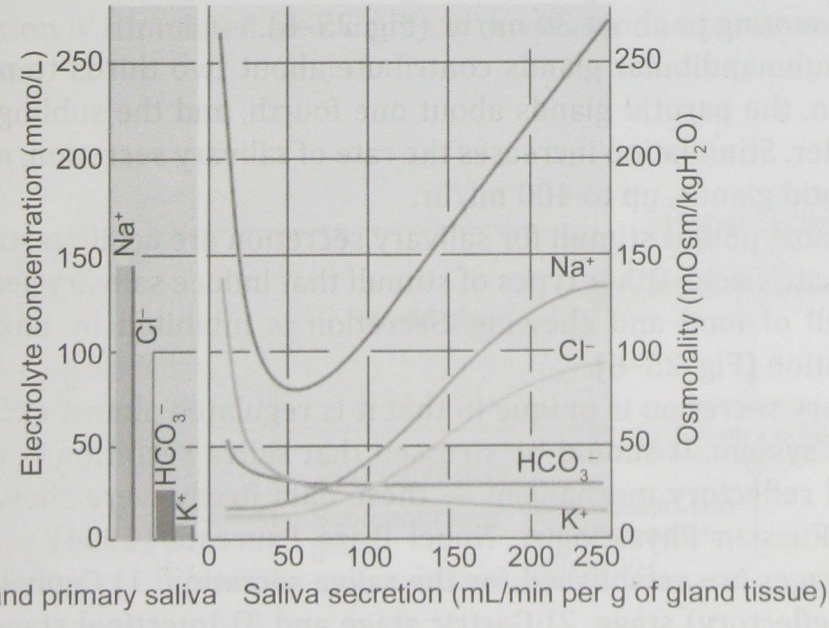


Fig. 25–5. Saliva composition and osmolality.

Primary saliva is modified in excretory ducts, yielding *secondary saliva.* As the saliva passes through the excretory ducts, Na^+ and Cl^- are reabsorbed and K^+ and HCO_3^- is secreted into the lumen. The saliva becomes *hypotonic* (far below 100 mOsm/kg H_2O; Fig. 25–4) because Na^+ and Cl^- reabsorp-tion is greater than K^+ and HCO_3^- secretion and the ducts are relatively impermeable to water (Fig. 25–4).

If the secretion rate rises to values much higher than 100 µl/min/g, these processes lag behind and the composition of secondary saliva becomes similar to that of primary saliva (Fig. 25–5).

Resuming, however, it should be stressed that saliva secreted in the acini (*primary saliva*) is probably isotonic, with concentrations of Na^+, K^+, Cl^-, and HCO_3^- that are close to those in plasma. The excretory ducts modify the composition of the saliva by extracting Na^+ and Cl^- and adding K^+ and HCO_3^-. The ducts are relatively impermeable to water. Therefore, at low salivary flows, the saliva that reaches the mouth — *secondary saliva* — is hypotonic, slightly acidic, and rich in K^+ but relatively depleted of Na^+ and Cl^-. When salivary flow is rapid, there is less time for ionic composition to change in the ducts. Consequently, although still hypotonic in humans, saliva is closer to isotonic, with higher concentrations of Na^+ and Cl^-.

The rate of saliva secretion varies from 0.1 to 4 ml/min (10–250 µl/min per gram gland tissue), depending on the degree of stimulation (Fig. 25–5). This adds up to about 0.5 to 1.5 l per day. In the resting state, salivary secretion is low, amounting to about 30 ml/hr (Fig. 25–6).

The submandibular glands contribute about two thirds to resting salivary secretion, the parotid glands about one fourth, and the sublingual glands the remainder. Stimulation increases the rate of salivary secretion, most notably in the parotid glands, up to 400 ml/hr.

The most potent stimuli for salivary secretion are acidic-tasting substances, such as citric acid. Other types of stimuli that induce salivary secretion include the smell of food and chewing. Secretion is inhibited by anxiety, fear, and dehydration (Fig. 25–6).

Salivary secretion is unique in that it is regulated almost exclusively by the nervous system. It should be stressed that saliva secretion is under the conditioned reflectory mechanism — these data firstly were shown by the well-known Russian Physiologist, Nobel Prize Laureate (1904) prof. I.P. Pavlov. Three stages are established for the saliva secretion: 1) Cephalic (oral, conditioned reflectory) stage, 2) Gastric stage and 3) Intestinal stage. The last two stages are of unconditioned nature.

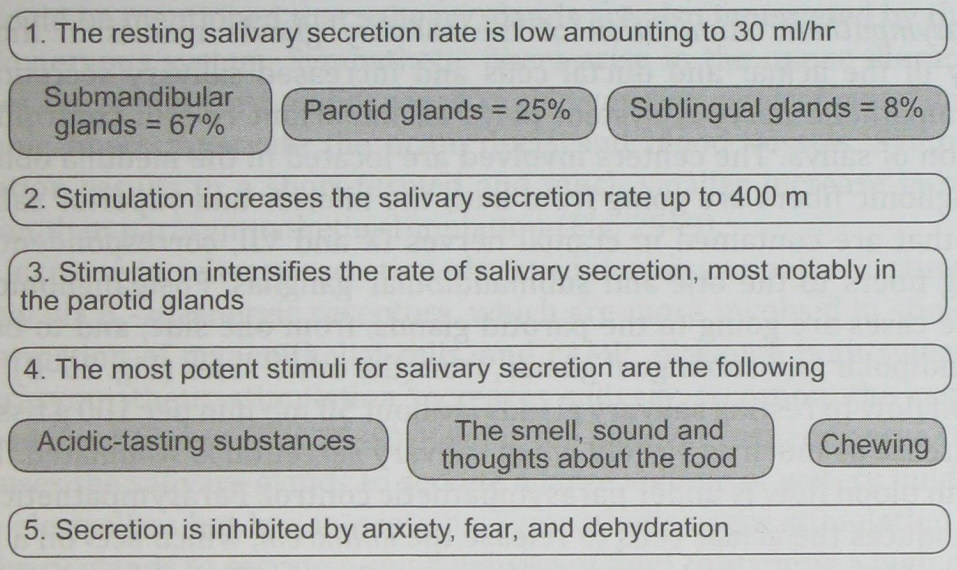


Fig. 25–6. The schematic overview concerning the saliva secretion.

The mostly important conditioned reflectory stage of saliva secretion results in saliva secretion stimulation in response to conditioned stimuli — food smell, food taste (it contact with the taste receptors), thoughts about the food, noise coincides with the food and the mastication itself (Fig. 25–7). Reflectory stimulation of saliva secretion occurs in the larger salivary glands.

Saliva secretion is stimulated also via the sympathetic and parasympathetic nervous systems.

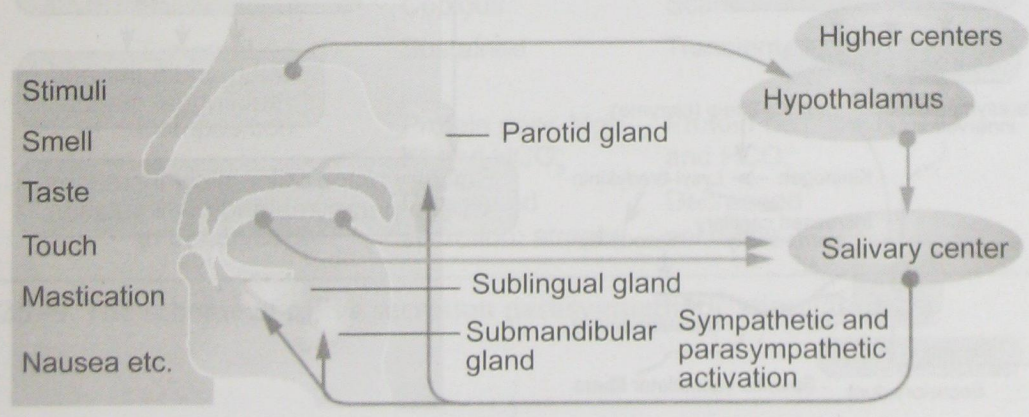


Fig. 25–7. The scheme of saliva secretion reflectory stimulation.

Parasympathetic stimulation of the salivary glands results in increased activity of the acinar and ductal cells and increased salivary secretion. The parasympathetic nervous system plays an important role in controlling the secretion of saliva. The centers involved are located in the medulla oblongata. Preganglionic fibers are going from both the inferior and superior salivatory nuclei that are contained in cranial nerves IX and VII, correspondently, and sending fibers to the otic and submandibular ganglias. Postganglionic fibers in these cases are going to the parotid glands, from one side, and to both the submandibular and sublingual glands, from the other sides (Fig. 25–8).

Blood flow to resting salivary glands is about 50 ml/min per 100 g tissue and can increase as much as 10-fold when salivary secretion is stimulated. This increase in blood flow is under parasympathetic control. Parasympathetic stimulation induces the acinar cells to release the *kallikrein*, which acts on a plasma globulin, *kininogen* which causes dilation of the blood vessels supplying the salivary glands (Fig. 25–8). Atropine, an anticholinergic agent, is a potent inhibitor of salivary secretion. Agents that inhibit acetylcholinesterase (e.g., pilocarpine) enhance salivary secretion. Some parasympathetic stimulation also increases blood flow to the salivary glands directly, apparently via the release of the neurotransmitter *vasoactive intestinal peptide (VIP).* VIP is known to increase the cAMP concentration of acinar cells.

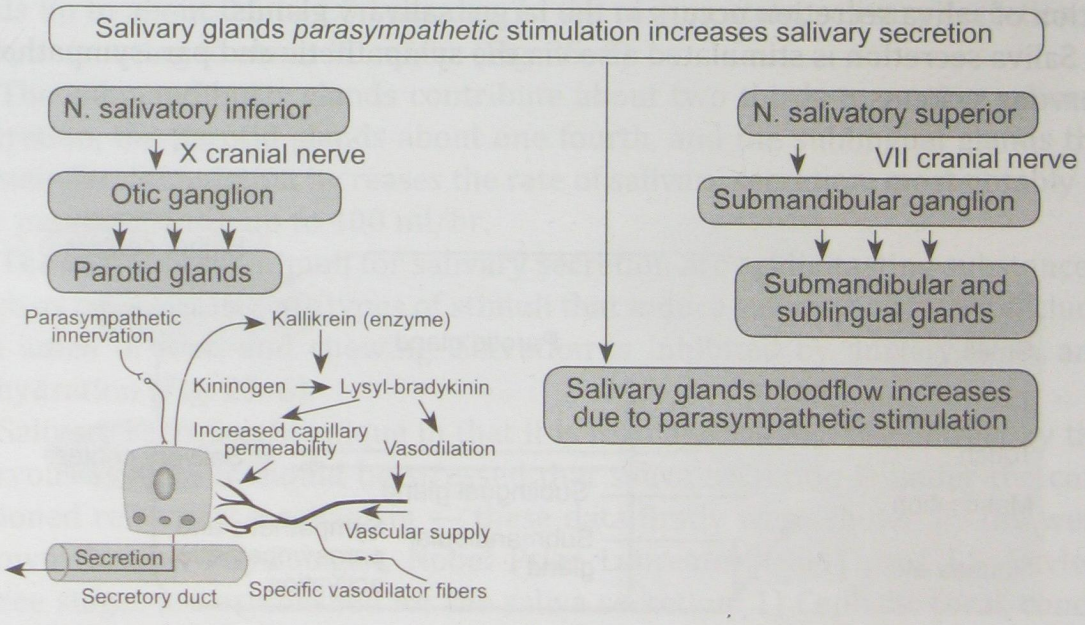


Fig. 25–8. The scheme of saliva secretion parasympathetic stimulation.

It should be mentioned that salivary glands are also innervated by the sympathetic nervous system. Sympathetic fibers arise in the upper thoracic segments of the spinal cord and synapse in the *superior cervical ganglion*. Postganglionic fibers innervate the acini, ducts and blood vessels. Sympathetic stimulation results in a short-termed and much smaller increase in salivary secretion than parasympathetic stimulation (Fig. 25–9).

The increase in salivary secretion observed during sympathetic stimulation is mainly via β_2-adrenergic receptors, which are more involved in stimulating the contraction of myoepithelial cells, and cAMP. Although both sympathetic and parasympathetic stimulation increases salivary secretion, the responses produced are different (Fig. 25–9). Briefly, the parasympathetic nerve stimulation causes the salivary glands to secrete a large volume of watery fluid that is high in electrolites but low in proteins. Sympathetic nerve stimulation causes the salivary glands to secrete a small volume of fluid containing a high concentration of mucus.

Salivary glands *sympathetic* stimulation increases salivary secretion

1. Sympathetic stimulation tends to result in a short-lived and much smaller increase in salivary secretion than parasympathetic stimulation.

2. The sympathetic increase in salivary secretion is mainly via B-adrenergic receptors.

3. Although both sympathetic and parasympathetic stimulation increases salivary secretion, the responses produced are different:

Responses	Parasympathetic	Sympathetic
Saliva output	Copious	Scant
Temporal response	Sustained	Transient
Composition	Protein poor, high K^+ and HCO_3^-	Protein-rich, low K^+ and HCO_2^-
Response to denervation	Decreased secretion, atrophy	Decreased secretion

Fig. 25–9. The scheme of saliva secretion parasympathetic stimulation.

4. DIGESTION AND ABSORPTION IN THE ORAL CAVITY

As it was mentioned earlier, salivary secretions begin the process of carbohydrates and lipids digestion. Salivary α-amylase can digest most of the ingested *carbohydrates* in to disaccharides before thay reach the small intestine. α-amylase is ultimately inactivated by the low pH in the stomach. It is difficult to speak about any carbohydrates absorption in the oral cavity in case of the normal chewing — food have to enter the esophagus after approximately 10 sec (Fig. 25–1). Some di- and monosaccharides absorption in the the blood vessels richly supplied the oral cavity is possible in case of prolonged chewing.

Salivary lipase begins to break down ingested *lipids* while they are in the oral cavity, stomach snd upper portions of the small intestine. The same condition — low pH in the stomach — halted lipase activity in the stomach.

Proteins digestion doesn't happen in the oral cavity, it starts only in the stomach where are the appropriate conditions for these biochemical reactions — both gastric juice and gastric pepsin. Both proteins and lipids are not absorpted in the oral cavity.

Chapter 26. Digestion in the stomach

1. ANATOMIC AND PHYSIOLOGICAL PROPERTIES OF STOMACH

The stomach is a pocket-like structure that lies in the upper part of the abdomen and serves as a food storage reservoir during the early stages of digestion. Although the residual volume of the stomach is only approximately 50 ml' it can increase to almost 1000 ml before the intraluminal pressure begins to rise. Stomach *size* is dependent on the degree of its filling but this distension is mainly limited to the proximal stomach (Fig. 26–1).

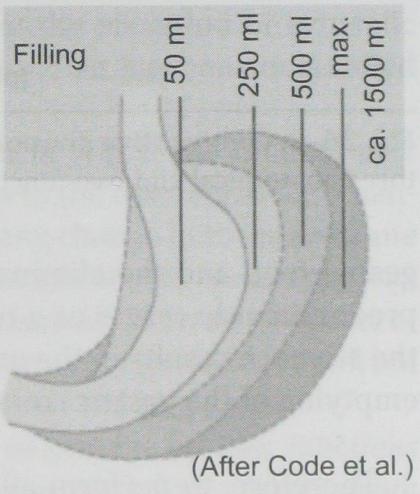


Fig. 26–1. Stomach filling.

The functional regions of the stomach do not correspond to the anatomic regions. Functionally, the stomach is divided into a proximal *reservoir* and distal *antral pump* on the basis of distinct differences in motility between the two regions (Fig. 26–2). The reservoir consists of the fundus and approximately one third of the corpus; the antral pump includes the caudal two thirds of the corpus, the antrum, and the pylorus. The reservoir is specialized for receiving and storing a meal.

2. GASTRIC MOTILITY

Differences in motility between the reservoir and antral pump reflect adapta-tions for different functions.

The muscles of the proximal stomach are adapted for maintaining continuous contractile tone (tonic contraction) and do not contract phasically. By contrast, the muscles of the antral pump contract phasically. The spread of phasic contractions in the region of the antral pump propels the gastric contents toward

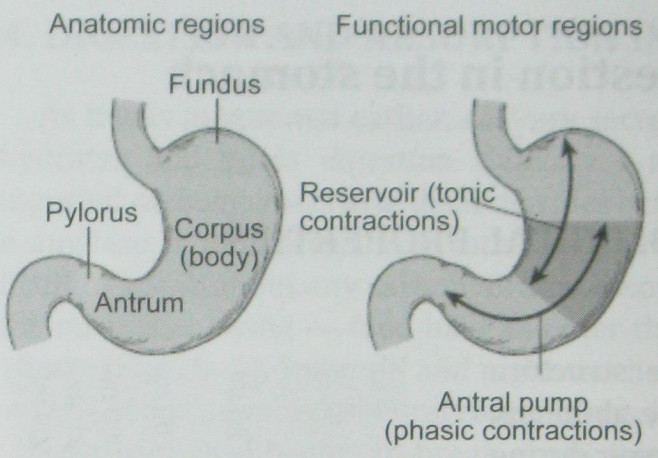


Fig. 26–2. Comparative divisions of stomach into three anatomical and two functional regions.

the small intestine. Strong propulsive waves of this nature do not occur in the proximal stomach.

2.1. Short overview of gastric motility

Motility in the digestive tract accounts for the propulsion, mixing and reservoir functions necessary for the orderly processing of ingested food and the elimination of waste products. The body of the stomach predominantly serves as a reservoir for ingested solids and liquids. Motility of the stomach results in the mixing and crushing of solid foods and regulates the emptying of the gastric contents or chyme into the duodenum.

Therefore, to perform all the named above functions of the stomach *there are four types of the stomach motility*:

1. The peristalsis (peristaltic movements).
2. The retropulsions.
3. The receptive relaxation.
4. Migrating motor complex.

As the result of all these types of gastric motility stomach have to prepare the chyme for digestion in the small intestine. *Chyme* is the semi-fluid material produced by the gastric digestion of food. Chyme results partly from the conversion of large solid particles into smaller particles via the combined peristaltic movements of the stomach and contraction of the pyloric sphincter. The peristaltic, cutting and retropulsive movements will be discussed later.

2.2. Electric discharges in the gastric myocytes

Smooth muscles of the stomach and intestine behave like *unitary type* smooth muscle. These muscles contract spontaneously in the absence of neu-

ral or endocrine influence and contract in response to stretch. There are no structured neuromuscular junctions and neurotransmitters travel over extended diffusion distances to influence relatively large numbers of muscle fibers.

Gastrointestinal smooth muscle differs from skeletal muscle in having two mechanisms that initiate the processes leading to contractile shortening and development of tension. In both skeletal muscle and gastrointestinal smooth muscle depolarization of the membrane electrical potential leads to the opening of voltage-gated calcium channels, followed by the elevation of cytosolic Ca^{2+} which, in turn, activates the contractile proteins. This mechanism is called *electromechanical coupling*.

Smooth muscles have an additional mechanism in which the binding of a ligand to its receptor on the muscle membrane leads to the opening of Ca^{2+} channels and the elevation of cytosolic calcium without any change in the membrane electrical potential. This mechanism is called *pharmacomechanical coupling*. The ligands may be chemical substances released as signals from nerves (neurocrine), from nonneural cells in close proximity to the muscle (paracrine) or from endocrine cells as hormones delivered to the muscle by the blood.

Smooth muscle fibers are connected to their neighbors by *gap junctions* which are permeable to ions and, thereby, transmit electrical current from muscle fiber to muscle fiber. Ionic connectivity, without cytoplasmic continuity from fiber to fiber, accounts for the *electrical syncytial properties* of smooth muscle, which confers electrical behavior analogous to that of cardiac muscle (see Chapter #17). Electrical activity and associated contractions spread from a point of initiation (e.g., the pacemaker region) in three dimensions throughout the mass of the muscle. A failure of nervous control can lead to disordered motility that includes spasm and associated abdominal pain.

Gastrointestinal smooth muscles electrical activity represents mainly the *slow-wave activity. Electrical slow waves* are universal and responsible for triggering action potentials in some regions, whereas in other regions (e.g., the gastric antrum and large intestinal circular muscle) they represent the only form of electrical activity (Fig. 26–3). They are always present in the small intestine where they decrease in frequency along a gradient from the duodenum to the ileum.

In the gastric antrum, the terms *slow wave* and *action potential* are used interchangeably for the same electrical event.

Slow waves in gastrointestinal muscles occur in four phases determined by specific ionic mechanisms — changes in calcium and potassium conductances (Fig. 26–3).

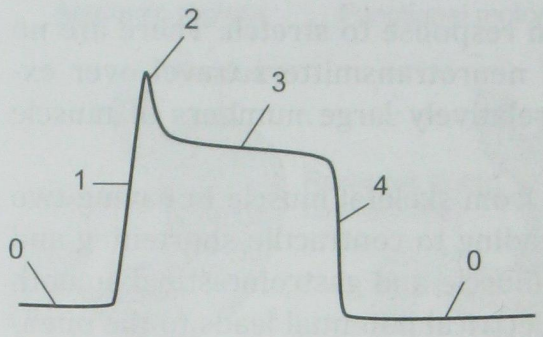


Fig. 26–3. The scheme of the stages of slow-wave potentials in the gastrointestinal myocytes.

Phase 0 — resting membrane potential. The mechanism of this stage is the outward K^+ current.

Phase 1 — the rising phase (upstroke depolarization). Both voltage-gated L-type calcium channels and voltage-gated potassium channels are activated.

Phase 2 — the beginning of the repolarization. The opening of potassium channels as the calcium channels are closing. Calcium inward current is prevailed upon outward potassium current.

Phase 3 — the plateau phase. It constitutes the balance between inward calcium current and outward potassium current.

Phase 4 — repolarization. Voltage-gated calcium channels are inactivated and calcium-gated potassium channels are activated.

It should be stressed that L-type calcium channels in gastrointestinal smooth muscle are essentially the same as those found in cardiac and vascular smooth muscle. Therefore, disordered gastrointestinal motility may be an adverse effect of treating of cardiovascular disease with drugs that block L-type calcium channels.

The frequencies of the slow-wave electrical activity equals to 3–5 per min, each with a duration of 2 to 20 seconds (Fig. 26–4).

As the peristaltic wave approaches the antrum, it speeds up and the entire terminal 5 to 10 cm of the antrum contracts, occluding the pyloric opening. Contraction of the antrum reverses the movement of the chyme, returning the larger particles to the body of the stomach for further churning and kneading. Because the pylorus is contracted during antral contraction, the gastric contents are emptied into the duodenum between contractions.

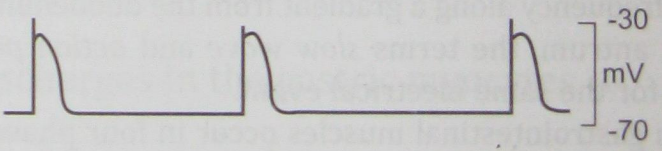


Fig. 26–4. The frequency of the electrical slow-wave in the stomach.

2.3. Peristaltic motor movement

Peristalsis is the organized propulsion of material over variable distances within the intestinal lumen. *Peristalsis* is the controlled movement of ingested foods, liquids, gastrointestinal secretions and sloughed cells from the mucosa through the digestive tract. It moves the food from the stomach into the small intestine and along the small intestine, with appropriate timing for efficient digestion and absorption. Propulsive forces move undigested material into the large intestine and eliminate waste through defecation.

Being a leading type of contraction, peristaltic waves with a relatively constant amplitude are associated with the rising phase of the action potential. A trailing contraction, of variable amplitude, is associated with the plateau phase (Fig. 26–5).

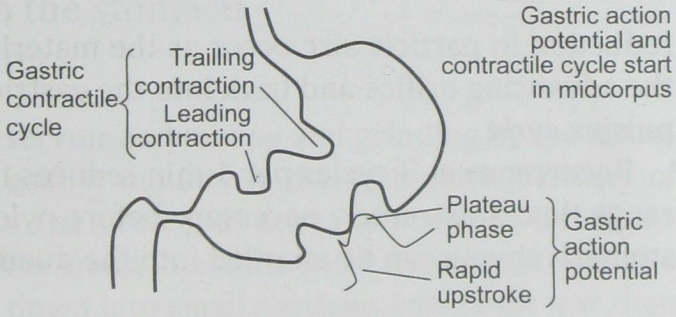


Fig. 26–5. The scheme of the two subtypes of peristaltic waves in the stomach.

The leading contractions produced by the rising phase of the gastric action potential have insignificant amplitude as they propagate to the pylorus. As the rising phase reaches the terminal antrum and spreads into the pylorus, contraction of the pyloric muscle closes the pyloric sphincter located between the stomach and duodenum. The trailing contraction follows the leading contraction by a few seconds.

As the peristaltic wave approaches the closed pylorus, the gastric contents are forced into an antral compartment of ever-decreasing volume and progressively increasing pressure. The second cycle of the peristaltic wave starts.

2.4. Retropulsion

Retropulsion is the same peristaltic wave (peristalsis) that is moving in the stomach in the retrograde direction. The origin of this wave is due to peristaltic wave meeting with the closed pyloric sphincter. This results in jet-like retropulsion through the internal surface of the stomach (Fig. 26–6). Cutting and

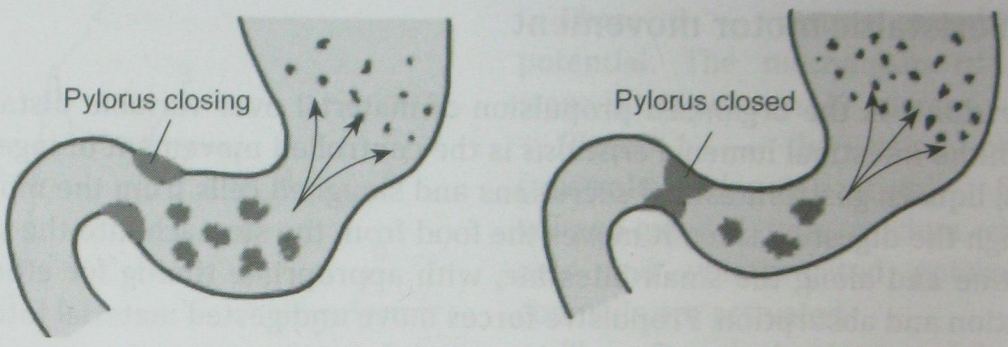


Fig. 26–6. The schematic representation of gastric retropulsion.

reduction in particle size occur as the material is forcibly retropelled through the advancing orifice and back into the gastric reservoir to await the next propulsive cycle.

Recurrence at 3 cycles per 1 min reduces food particles size to the 1–7-mm range that is extremely necessary before pyloric sphincter can be opened and stomach chyme can be emptied into the duodenum.

2.5. Receptive relaxation

Total inhibition of the gastric motility defines as *receptive relaxation*. Neurally mediated, it decreases in tonic contracture of the musculature are responsible for relaxation in the gastric reservoir (i.e., increased volume; Fig. 26–7). Physiologically, receptive relaxation makes 'quiet' ideal conditions for the ingested food particles with the aim of the better HCl and other humoral factors biochemical interaction with the food within the stomach.

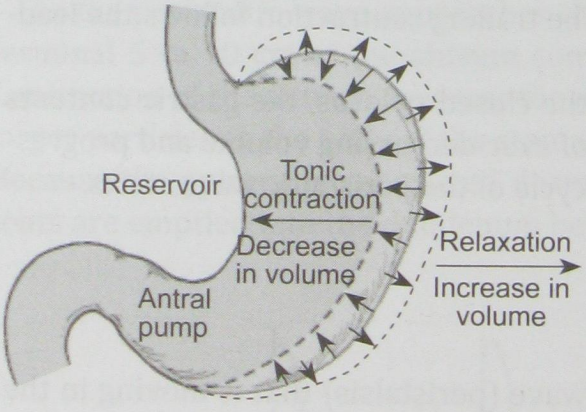


Fig. 26–7. The scheme of gastric muscles tonic contraction and relaxation.

Receptive relaxation is a vago-vagal reflex triggered by distension of the gastric reservoir. Sectioning the vagus nerve will prevent or significantly diminish receptive relaxation since the process is controlled by the enteric nervous system.

2.6. Migrating motor complexr

The pattern of motility during the interdigestive condition is called the *migrating motor complex*. It represents a single pulsatile wave that originates in the stomach and moves toward the large bowel during the interdigestive period with the frequency of 1 wave per 60–90 min.

Physiological importance of the migrating motor complex is mainly adaptive. Migrating motor complex appears to be a mechanism for large indigestible particles clearing from the stomach during the interdigestive state.

2.7. Chyme evacuation from the stomach. Pyloric sphincter opening

In addition to storage in the reservoir and mixing and grinding by the antral pump, an important function of gastric motility is the systematic delivery of the gastric chyme to the duodenum at a rate that does not overload the digestive and absorptive functions of the small intestine. Solid food remains in the stomach until it has been broken down into small particles (diameter less than 1 mm) and suspended in chyme. The chyme then passes to the duodenum. The time required for 50 % of the ingested volume to leave the stomach varies, e.g., 10–20 min for water and 1–4 hrs for solids (carbohydrates < proteins < fats).

The rate of gastric emptying is regulated by neural control mechanisms to compensate for variations in the volume, composition, and physical state of the gastric contents.

The following factors determined the rate of the gastric emptying and pyloric sphincter opening:

1. *Chyme osmolality.* Hypotonic and hypertonic liquids empty more slowly than isotonic liquids.
2. *Chyme osmolarity.* The bigger are sizes of the ingested food particles the slower are the rate of the pyloric sphincter opening.
3. *Chyme acidity*. Gastric contents (chyme) acidity in the pyloric part of the stomach have to slightly higher The rate of gastric emptying decreases as the of the increases.
4. Caloric content of the gastric chyme. Meals with a high caloric content empty from the stomach at a slower rate than meals with a low caloric content.

5. Intragastric or intraduodenal pressure. Sometimes increased intragastric pressure due to the stomach overloading results in earlier the pyloric sphincter opening. In the contrast, much chyme accumulation within the duodenum delays stomach emptying.

The mechanisms of control of gastric emptying keep the rate of delivery of calories to the small intestine within a narrow range, regardless of whether the calories are presented as carbohydrate, protein, fat, or a mixture. Of all of these, fat is emptied the most slowly, or stated conversely, fat is the most potent inhibitor of gastric emptying. Part of the inhibition of gastric emptying by fats may involve the release of the hormone cholecystokinin, which itself is a potent inhibitor of gastric emptying.

The emptying of the stomach is regulated by hormonal and neural mechanisms (Fig. 26–8).

Abbreviations: ACh — acetylcholine, BER — basic electrical rhythm, CCK — cholecystokinin, GIP — gastric inhibitory polypeptide, GRP — gastrin-releasing peptide, VIP — vasoactive intestinal peptide.

The hormones cholecystokinin and gastric inhibitory peptide, which are thought to control gastric emptying, are released in response to the pH and the osmolar and fatty acid composition of the chyme.

Local and central circuitries are involved in the neural control of gastric emptying. Afferent receptor fibers synapse with the neurons in the intramural

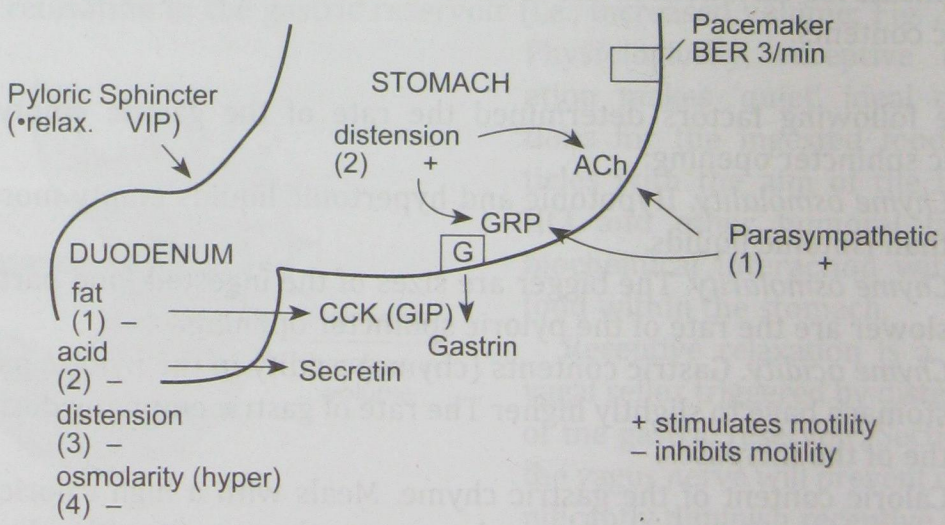


Fig. 26–8. Factors that regulated gastric motility.

plexus or trigger intrinsic reflexes by means of vagal or sympathetic pathways that participate in extrinsic reflexes.

Short overview concerning the primary and additional humoral factors influencing on the stomach motility is given on Fig. 26–8.

3. GASTRIC SECRETORY FUNCTION

Each day, approximately 7000 ml of fluid is secreted into the gastrointestinal tract. Among this amount stomach secretes appr. 2000 ml of fluid. Approximately 50 to 200 ml of this fluid leaves the body in the stool; the remainder is reabsorbed in the small and large intestines. These secretions are mainly water and have sodium and potassium concentrations similar to those of extracellular fluid. Because water and electrolytes for digestive tract secretions are derived from the extracellular fluid compartment, excessive secretion or impaired absorption can lead to extracellular fluid deficit.

The secretory activity of the gut is influenced by local, mainly, mechanical, humoral and neural influences. Neural control of gastrointestinal secretory activity is mediated through the autonomic nervous system. Secretory activity, like motility, is increased with parasympathetic stimulation and inhibited with sympathetic activity. Many of the local influences, including pH, osmolality and chyme consistently act as stimuli for neural and humoral mechanisms.

3.1. Gastric Secretions

The stomach's mucosal lining, the glandular gastric mucosa, contains three main types of glands: cardiac, pyloric and oxyntic. These glands contain mucous cells that secrete mucus and HCO_3^- ions, which protect the stomach from the acid in the stomach lumen.

The oxyntic glands contain *parietal (oxyntic) cells, chief cells, mucous neck cells* and *some endocrine cells* (Fig. 26–9). Surface mucous cells occupy the gastric pit (foveola); in the gland, most mucous cells are located in the neck region. The base of the oxyntic gland contains mostly chief cells, along with some parietal and endocrine cells. Mucous neck cells secrete mucus, parietal cells principally secrete *hydrochloric acid* (HCl) and *intrinsic factor* and chief cells secrete *pepsinogen*. Gastric intrinsic factor which is produced by the parietal cells, is necessary for the vitamin B_{12} absorption. The pepsinogen that is secreted by

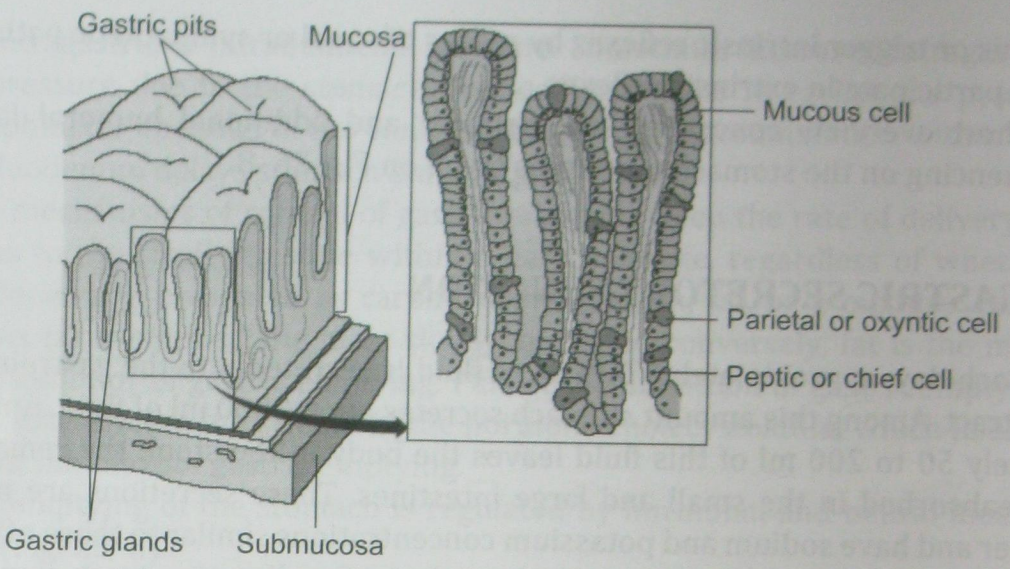


Fig. 26–9. Gastric pit from body of the stomach. The increased imagination gives a simplified diagram of the oxyntic gland with their main types of secreting cells.

the chief cells is rapidly converted to pepsin when exposed to the low pH of the gastric juice.

Resuming, the oxyntic gland area of the stomach is composed of glands and pits (Fig. 26–9). The surface area and gastric pits are lined with mucus-producing epithelial cells. The bases of the gastric pits contain the parietal (or oxyntic) cells, which secrete hydrochloric acid and intrinsic factor, and the chief (peptic) cells, which secrete large quantities of pepsinogen. There are approximately 1 billion parietal cells in the stomach; together they produce and secrete approximately 20 mEq of hydrochloric acid in several hundred milliliters of gastric juice each hour. Gastric intrinsic factor, which is produced by the parietal cells, is necessary for the absorption of vitamin B_{12}. The pepsinogen that is secreted by the chief cells is rapidly converted to pepsin when exposed to the low pH of the gastric juices.

The important constituents of human gastric juice are HCl, electrolytes, pepsinogen and intrinsic factor. Compared to extracellular fluid gastric secretions are rich in H^+, K^+ and Cl^- but low in Na^+. Gastric juice pH is low, about 0.7 to 3.8. Swallowed food buffers it to a pH of 1.8-4.0 which is optimal for most pepsins.

One of the important characteristics of the gastric mucosa is resistance to the highly acid secretions that it produces. This raises a question: 'How does the gastric mucosa protect itself from acidity?' As mentioned earlier, the sur-

face mucous cells secrete a fluid containing mucus and HCO_3^- ions. The mucus forms a *mucus gel layer* covering the surface of the gastric mucosa. Bicarbonate trapped in the mucus gel layer neutralizes acid, preventing damage to the mucosal cell surface.

HCl functions: 1) it participates in the proteins breakdown, 2) it provides an optimal pH for the pepsin action, 3) antibacterial efficacy — it hinders the growth of pathogenic bacteria.

HCl secretion. Normally, the secretion of hydrochloric acid by the parietal cells of the stomach is accompanied by secretion of bicarbonate ions (HCO_3^-). For every hydrogen ion (H^+) that is secreted, a HCO_3^- is produced, and as long as HCO_3^- production is equal to H^+ secretion, mucosal injury does not occur.

On stimulation (by a cholinergic agonist, gastrin or histamine) the lumenal surface develops a negative potential as compared to the nutrient surface. Active transport of Cl^- and H^+ is carried out by separate pumps in the apical membrane (Fig. 26–10). Generation of H^+ within the cell occurs with the simultaneous generation of equal quantities of OH^- which must be neutralized by H_2CO_3 formed by the hydration of CO_2, as catalysed by carbonic anhydrase.

The mechanism which generates $H^+ + OH^-$ is believed lo be a proton pump in the apical plasma membrane of the parietal cell. This pump is a *H^+/K^+-dependent*

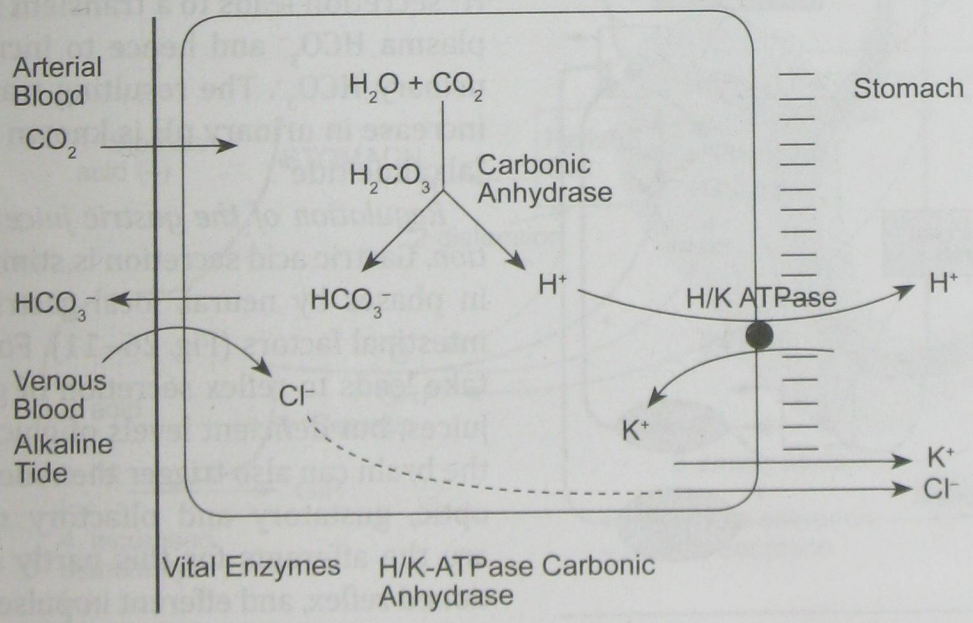


Fig. 26–10. The simplified scheme of HCl secretion by the gastric parietal cells.

ATPase which has been identified in and purified from the apical membrane. The hydrolysis of a molecule of ATP to ADP + inorganic phosphate at the cytosolic face of the membrane pump is accompanied by the active transport of H^+ out of the cell in exchange for a K^+ ion transported into the cell. Thus, the H^+/K^+-ATPase recycles K^+ ions back into the cell in exchange for H^+ ions. The basolateral cell membrane has an electroneutral Cl^-/HCO_3^- exchanger that balances the entry of Cl^- into the cell with an equal amount of HCO_3^- entering the bloodstream. The Cl^- inside the cell then leaks into the lumen through Cl^--channels, down an electrochemical gradient. Consequently, HCl is secreted into the lumen.

The HCO_3^- diffuses out of the cell into the extracellular fluid. Osmotic gradients resulting from the active transport of H^+ and Cl^- result in net movement of H_2O into the lumen.

For each H^+ transported into the gastric secretion, one OH^- remains in the cytosol. The rate of production of OH^- ions is such that the intracellular pH would rapidly rise and the cell suffer serious damage. Consequently, the OH^- ions must be neutralized by H_2CO_3.

Overall, one HCO_3^- is balanced by the equal amount of HCl produced, and HCO_3^- diffuses into the extracellular fluid and plasma compartments. On eating a meal the stimulation of gastric H^+ secretion leads to a transient rise in plasma HCO_3^- and hence to increased urinary HCO_3^-. The resulting transient increase in urinary pH is known as the "alkaline tide".

Regulation of the gastric juice secretion. Gastric acid secretion is stimulated in phases by neural, local gastric and intestinal factors (Fig. 26–11). Food intake leads to reflex secretion of gastric juices, but deficient levels of glucose in the brain can also trigger the reflex. The optic, gustatory and olfactory nerves are the afferents for this partly conditioned reflex, and efferent impulses flow via the vagus nerve. The stimulation of HCl secretion resulting from the inges-

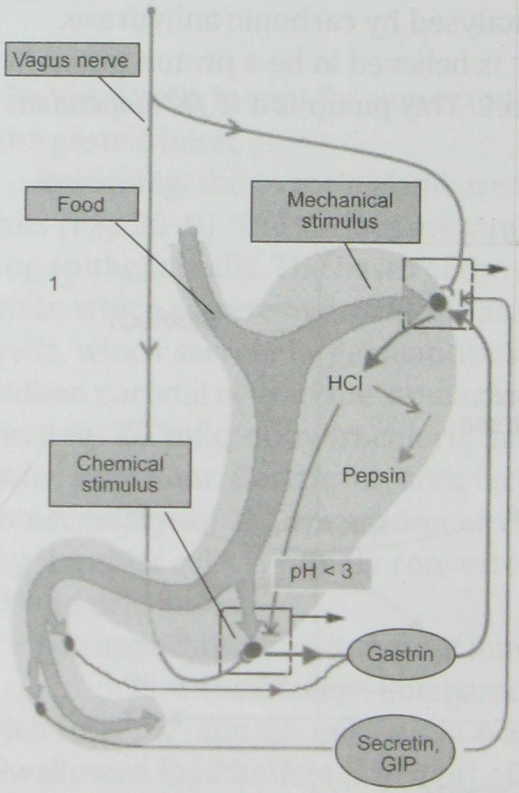


Fig. 26–11. Regulation of gastric juice secretion.

tion of food can be divided into three phases: the cephalic phase, the gastric phase and the intestinal phase.

The cephalic phase involves mainly the CNS activity as it was established by Russian physiologist Pavlov in the 1890's. He found that before food enters the stomach, neural mechanisms predominate.

Sight, smell, taste (plus *food swallowing),* thoughts *about the* food, *the sounds that usually followed food intake* are the *main stimuli* for HCl secretion during this phase. In the cases listed above impulses are transmitted via the vagus nerve (Fig. 26–11) to the parietal and G-cells in the stomach (Fig. 26–12).

The nerve endings release *ACh* which directly stimulates acid secretion from parietal cells (neural regulation). The nerves also release gastrin-releasing peptide (GRP), which stimulates G-cells to release *gastrin*, indirectly stimulating parietal cell acid secretion (endocrine regulation).

Important that parietal cells possess special *histamine* receptors — H_2- receptors, whose stimulation results in increased HCl secretion (paracrine regulation). Special endocrine cells of the stomach known as enterochromaffin-like cells are believed to be the source of this histamine. The importance of histamine as an effector of gastric acid secretion has been indirectly demonstrated by the effectiveness of cimetidine, an H_2-blocker, in reducing gastric juice secretion. H_2-blockers are commonly used for the treatment of peptic ulcer disease or gastroesophageal reflux disease.

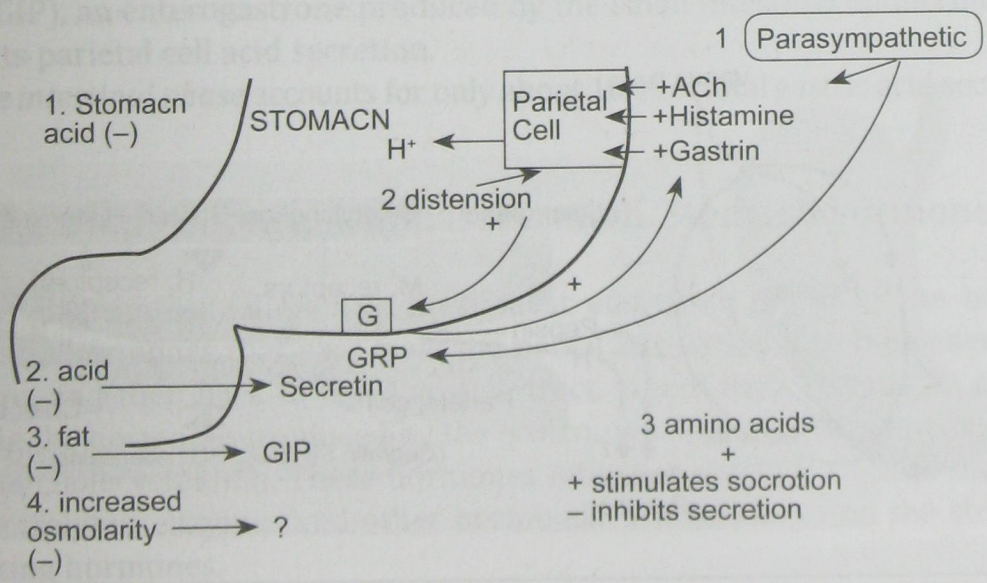


Fig. 26–12. The scheme of the overall regulation of the parietal cells secretion.

As a result, there are 3 humoral substances that stimulate HCl secretion during the cephalic stage — i.e. acetylcholine, gastrin and histamine (Fig. 26–12).

Inhibition of HCl secretion during this phase accounts for the mechanism of the conditioned inhibition.

The cephalic phase probably accounts for about 40 % of total acid secretion.

The gastric phase is mainly a result of gastric distension and chemical agents such as digested proteins. Distension of the stomach stimulates mechanoreceptors (Fig. 26–11) which stimulate the parietal cells directly through short local (enteric) reflexes and by long *vago-vagal reflexes.* Vago-vagal reflexes are mediated by afferent and efferent impulses travelling in the vagus nerves. Hence, neural regulation in the gastric phase is via the vagus. Vagal efferents have both direct and potentiating effects. Stimulation is provided by *acetylcholine*, the postganglionic transmitter of vagal parasympathetic fibers (via muscarinic M_1 receptors).

Digested proteins in the stomach are also potent stimulators of gastric acid secretion, an effect mediated through gastrin release (Fig. 26–13).

Several other chemicals, such as *alcohol* and *caffeine*, stimulate gastric acid secretion through mechanisms that are not well understood.

Inhibition of HCl secretion during gastric phase occurs due to enterogastral peptide *somatostatin* that secretes by the endocrine cells (D cells) in the antrum. Somatostatin inhibits the release of gastrin and, thus, gastric acid secretion. Another mechanism for inhibiting gastric acid secretion is acidification of the duodenal

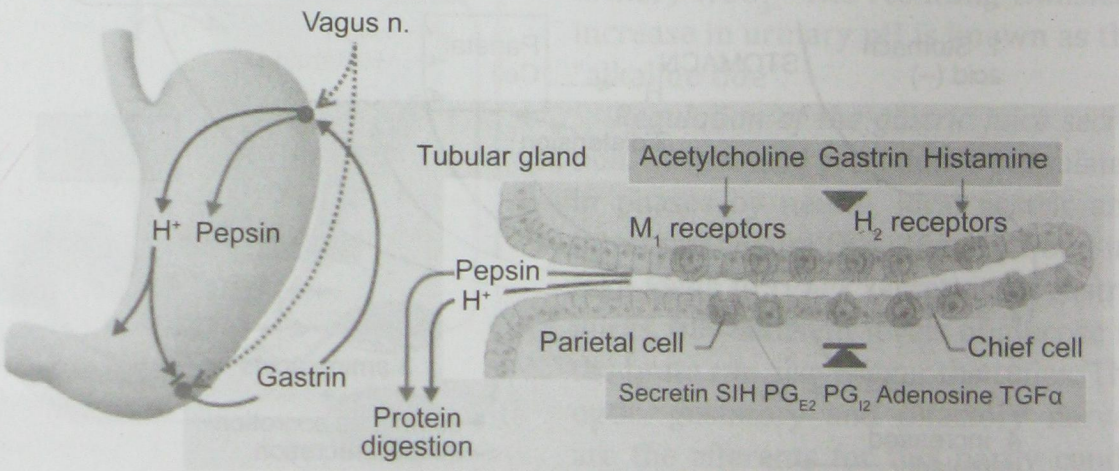


Fig. 26–13. Schematic illustration of neural, endocrine, paracrine and autocrine mechanisms of gastric juice secretion regulation.

lumen. *Secretin* also known to inhibit the release of gastrin. The inhibition of gastric juice secretion by a high concentration of H^+ ions (pH<3.0, see Fig. 26–11) in the gastric lumen is also an important regulatory mechanism (negative feedback; Fig. 26–13, left side).

Prostaglandins, chemical messengers derived from cell membrane lipids, play an important role in HCl secretion inhibition. The prostaglandins (especially E_2 and I_2) probably exert their effect through improved blood flow, increased bicarbonate ion secretion, and enhanced mucus production. Transforming growth factor-alpha (TGF-a) and adenosine (both — paracrine and autocrine regulation) are also known to suppress gastric secretion (Fig. 26–13, right side).

The gastric phase accounts for about 50 % of total gastric acid secretion.

During *the intestinal phase* protein *digestion products* in the duodenum stimulate gastric juice secretion through the action of the circulating amino acids on the parietal cells.

Distension of the small intestine, probably via the release of the hormone *enterooxyntin* from intestinal endocrine cells, stimulates acid secretion.

Food in the duodenum releases one or more hormones which stimulate gastric secretion. This has been demonstrated in man and in experimental animals. The particular hormones have not been identified, although *cholecystokinin* may be one.

Acid, fatty acids or hyperosmolar solutions in the duodenum stimulate the release of enterogastrones, which inhibit HCl secretion. Gastric inhibitory peptide (GIP), an enterogastrone produced by the small intestinal endocrine cells, inhibits parietal cell acid secretion.

The intestinal phase accounts for only about 10 % of total gastric acid secretion.

3.2. Gastrointestinal hormonal secretion. Stomach hormones

The gastrointestinal tract is the largest endocrine organ in the body. It produces hormones that pass from the portal circulation into the general circulation and then back to the digestive tract, where they exert their actions. Among the hormones produced by the gastrointestinal tract are gastrin, secretin and cholecystokinin. These hormones influence motility and the secretion of electrolytes, enzymes and other hormones. We will focus on the stomach-secreting hormones.

The primary function of *gastrin* is the stimulation of gastric acid secretion. Gastrin also has a trophic, or growth-producing, effect on the mucosa of the

small intestine, colon, and oxyntic (acid-secreting) gland area of the stomach. Removal of the tissue that produces gastrin results in atrophy of these structures. This atrophy can be reversed by the administration of exogenous gastrin.

3.3. Protection of the gastric and duodenal mucosa

Because the acid–pepsin mixture of gastric secretion denatures and digests protein, the protein-containing wall of the stomach and duodenum has to be protected from the harmful action of gastric juice. The following mechanisms are involved in this:

1. A gel-like *mucus film*, 0.1–0.5mm thick, protects the surface of the gastric epithelium. The mucus is secreted by epithelial cells (and depolymerized by the pepsins so that it can then be dissolved; Fig. 26–14, a).
2. The *epithelium secretes* HCO_3^- ions that are enriched not only in the liquid layer directly over the epithelium, but also diffuse into the mucus film, where they buffer H^+ ions that have penetrated from the gastric lumen. *Prostaglandins* are important stimulants of this HCO_3^- secretion (Fig. 26–14, b).
3. In addition, the *epithelium itself* (apical cell membrane, tight junctions) has *barrier properties* that largely prevent the penetration of H^+ ions or can very effectively remove those H^+ ions that have already penetrated (Na^+/H^+ exchange carrier only basolaterally). These properties are regulated, among others, by the *epidermal growth factor* contained in saliva and bound to receptors of the apical epithelial membrane. Glutathione-dependent antioxidative mechanisms are also part of this *cytoprotection* (Fig. 26–14, c).
4. Finally, good *mucosal blood flow* serves as the last "line of defense" that, among other actions, quickly removes H^+ ions and provides a supply of HCO_3^- and substrates of energy metabolism (Fig. 26–14, d).

4. DIGESTION AND ABSORPTION IN THE STOMACH

Digestion and absorption occur mainly in the small intestine. The stomach is a poor absorptive structure and only a few lipid-soluble substances, including alcohol, are absorbed from the stomach.

Digestion is the process of dismantling foods into their constituent parts. Digestion requires hydrolysis, enzyme cleavage and fat emulsification. Most pro-

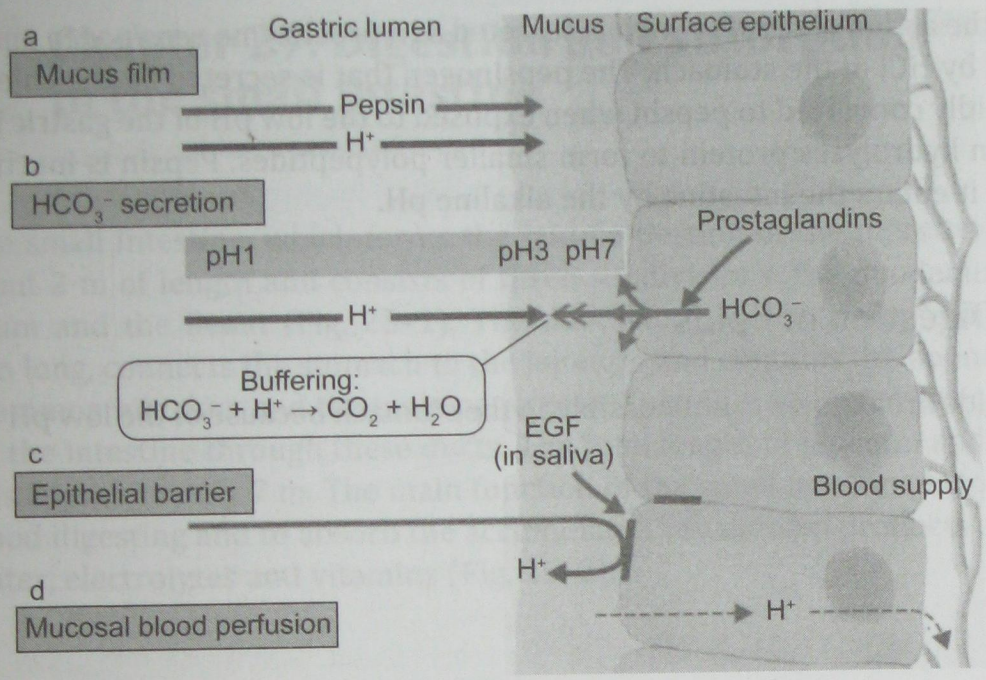


Fig. 26–14. Mucosa protection

teins, fats and carbohydrates must be broken down into smaller particles before they can be absorbed. Although some digestion of carbohydrates and proteins begins in the stomach, digestion also takes place mainly in the small intestine.

4.1. Digestion of carbohydrates

In the stomach, especially in its initial, antral part one could observe the residual activity of the alpha-amylase of the saliva that makes possible a little digestive activity of the stomach concerning carbohydrates digestion. Low pH value of the rest stomach compartments makes impossible any digestive processes with carbohydrates.

4.2. Digestion of proteins

Proteins form the fundamental structure of cells and are the most abundant of all organic compounds in the body. Protein digestion starts in the stomach

with the action of *pepsin* which is secreted as a proenzyme *pepsinogen* and activated by HCl in the stomach. The pepsinogen that is secreted by the chief cells is rapidly converted to pepsin when exposed to the low pH of the gastric juices. Pepsin hydrolyzes protein to form smaller polypeptides. Pepsin is inactivated when it enters the intestine by the alkaline pH.

4.3. Digestion of lipids

Lipids breakdown is not possible in the stomach because of the low pH value.

Chapter 27. Digestion and absorption in the small intestine

The small intestine, which forms the middle portion of the digestive tract, is about 2 m of length and consists of three subdivisions: the duodenum, the jejunum and the ileum (Fig. 25–1). The duodenum, which is approximately 22 sm long, connects the stomach to the jejunum and contains the opening for the common bile duct and the main pancreatic duct. Bile and pancreatic juices enter the intestine through these ducts. The total length of jejunum and ileum equals approximately 7 m. The main function of the small intestine is to finish the food digesting and to absorb the accumulated breakdown products as well as water, electrolytes and vitamins (Fig. 25–2).

1. SMALL INTESTINE MOTILITY

The small intestine is the major site for the digestion and absorption of food. Therefore, its movements are mixing and propulsive. The functions of small intestine motility are providing (a) chyme further degradation and successful absorption and (b) chyme passage toward the large intestine for its evacuation from the gastrointestinal tract. To perform all the named above functions of the stomach, *there are three types of the small intestine movement*:

1. The peristalsis (peristaltic movements).
2. The segmentation.
3. Migrating motor complex.

Intestinal motility is autonomously regulated by the enteric nervous system but is influenced also by hormones and external innervation. The movement of contents through the gastrointestinal tract is controlled by neurons located in the submucosal and myenteric plexuses of the gut. The axons from the cell bodies in the myenteric plexus innervate the circular and longitudinal smooth muscle layers of the gut. These neurons receive impulses from local receptors located in the mucosal and muscle layers of the gut and extrinsic input from the parasympathetic and sympathetic nervous systems. As a general rule, the parasympathetic nervous system tends to increase the motility of the bowel, whereas sympathetic stimulation tends to slow its activity.

1.1. Peristalsis

Peristalsis is the organized propulsion of material over variable distances within the intestinal lumen. It moves the food from the stomach into the small intestine and along the small intestine, with appropriate timing for efficient digestion and absorption. Propulsive forces move undigested material into the large intestine and eliminate waste through defecation.

The muscle layers of the intestine behave in a stereotypical pattern during peristaltic propulsion (Fig. 27–1). This pattern is determined by the integrated circuits of the enteric nervous system. During peristalsis, the longitudinal muscle layer in the segment ahead of the advancing intraluminal contents contracts while the circular muscle layer simultaneously relaxes. The intestinal tube behaves like a cylinder with constant surface area. The shortening of the longitudinal axis of the cylinders is accompanied by a widening of the cross-sectional diameter.

Propulsive movements occur with synchronized activity in a section 10 to 20 cm long. The simultaneous shortening of the longitudinal muscle and relaxation of the circular muscle results in expansion of the lumen, which prepares a *receiving segment* for the forward-moving intraluminal contents during peristalsis.

The second component of stereotyped peristaltic behavior is contraction of the circular muscle in the segment behind the advancing intraluminal contents. The longitudinal muscle layer in this segment relaxes simultaneously with contraction of the circular muscle, resulting in the conversion of this region to *a propulsive segment* that propels the luminal contents ahead into the receiving segment (Fig. 27–1).

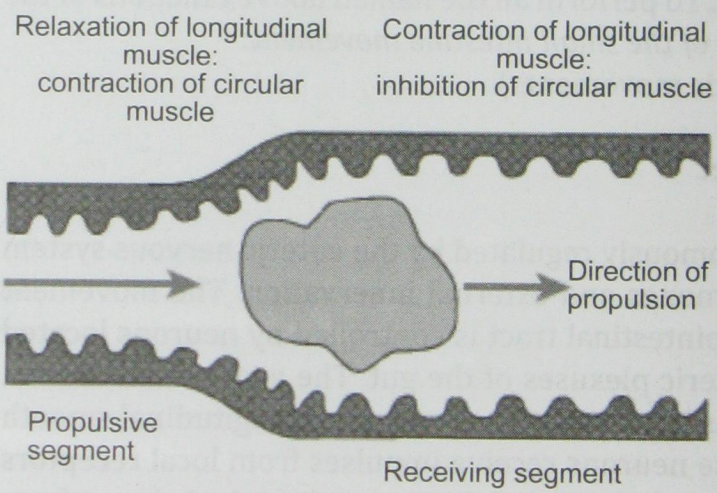


Fig. 27–1. Schematic draw of the single peristaltic wave. Peristaltic propulsion involves formation of a propulsive and a receiving segment.

Regular peristaltic move-ments start in the duodenum near the entry sites of the common duct and the main hepatic duct where the

pacemaker cells (*interstitial Cajal cells*) localized. A series of local pacemakers maintains the frequency of intestinal contraction. The membrane potential of these cells oscillates between 10 mV and 20 mV every 3-15 min, producing *slow waves* (Fig. 27–2, fragment 1). Their amplitude can rise (less negative potential) or fall in response to neural, endocrine or paracrine stimuli. A series of action potentials (spike bursts) are fired once the membrane potential rises above a certain threshold (appr. –40 mV) (Fig. 27–2, fragment 2).

The peristaltic movements (approximately 12 per minute in the jejunum) become less frequent as they move further from the pylorus, becoming approximately 9 per minute in the ileum.

Muscle spasms occur if the trough of the wave also rises above the threshold potential (Fig. 27–2, fragment 3).

1.2. Segmentation

The peristaltic contractions produce *segmentation* waves and propulsive movements of the small intestine. With segmentation waves, slow contractions of circular muscle occlude the lumen and drive the contents forward and backward. Most of the contractions that produce segmentation waves are local events involving only 1 to 4 cm at a time. They function mainly to mix the chyme with the digestive enzymes from the pancreas and to ensure adequate exposure of all parts of the chyme to the mucosal surface of the intestine, where absorption takes place.

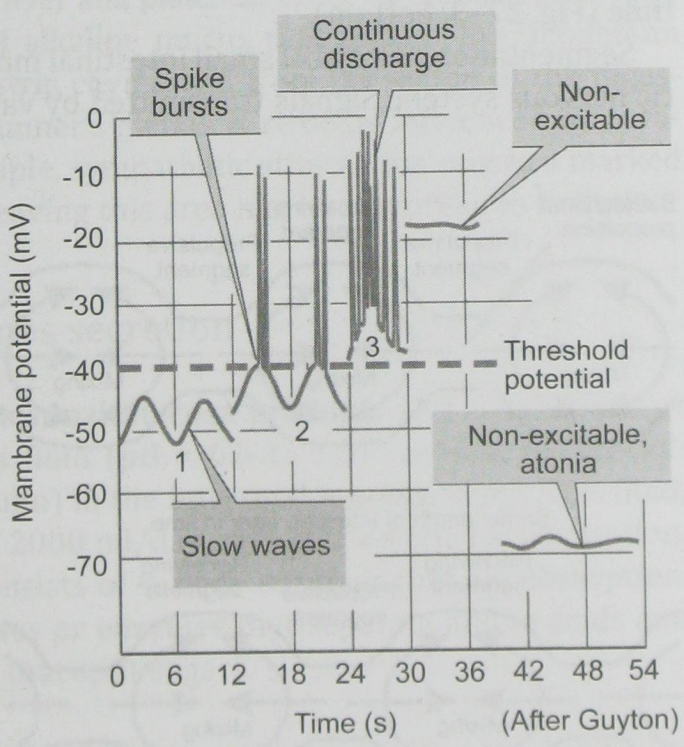


Fig. 27–2. The scheme of the electrical activity (slow wave and spikes) in the myenteric cells of the small intestine.

Segmentation appears to be the most common motor event in the small intestine. The frequency of segmenting activity increases after a meal, presumably stimulated by receptors in the stomach and intestine. Circular muscle contractions in short propulsive segments are separated on either end by relaxed receiving segments. Each propulsive segment jets the chyme in both directions into the relaxed receiving segments where stirring and mixing occur (Fig. 27–3). This happens continuously at closely spaced sites along the entire length of the small intestine.

The segmentation pattern of motility is characteristic of the digestive state. Propulsive segments separated by receiving segments occur randomly at many sites along the small intestine. Mixing of the luminal contents occurs in the receiving segments. Receiving segments convert to propulsive segments, while propulsive segments become receiving segments (Fig. 27–3).

The intervals of time between mixing contractions are the same as for electrical slow waves. A higher frequency of electrical slow waves and associated contractions in more proximal regions and the peristaltic nature of the mixing movements result in a net rare propulsion of the luminal contents over time (Fig. 27–3, bottom).

Segmentation pattern of small intestinal motility is programmed by the enteric nervous system. Signals transmitted by vagal efferent nerves to the enteric nervous system interrupt the migrating motor complex and initiate mixing motility during ingestion of a meal. When the vagus nerves are blocked during the digestive state, migrating motor complexes reappear in the intestine but not in the stomach.

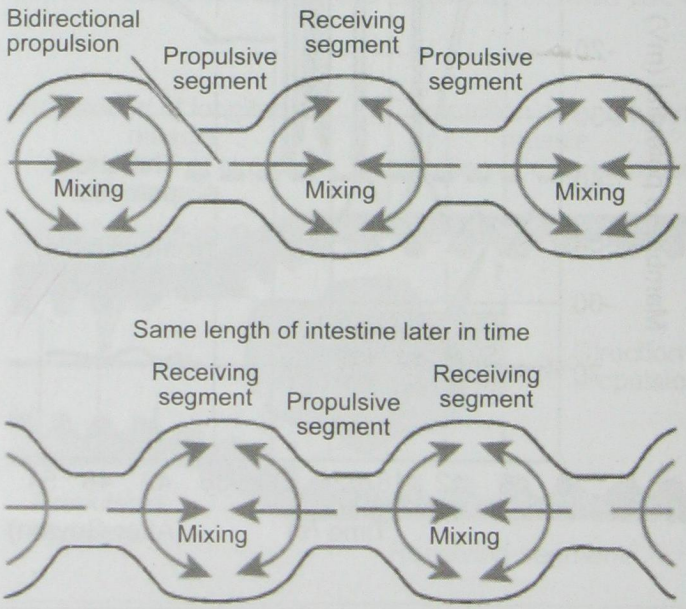


Fig. 27–3. Segmentation in the small intestine. Mixing movements.

1.3. Migrating motor complex

The small intestine is in the *digestive state* when nutrients are present and the digestive processes are

ongoing. It converts to the *interdigestive state* 2 to 3 hrs after a meal when the digestion and absorption of nutrients are complete. The pattern of motility in the interdigestive state is called the *migrating motor complex*.

So, it represents a single pulsatile wave that originates in the stomach and moves toward the large bowel during the interdigestive period with the frequency of 1 wave per 60–90 min.

2. INTESTINAL SECRETIONS

2.1. Mucus secretion

The small intestine secretes digestive juices and receives secretions from the liver and pancreas. An extensive range of mucus-producing glands, called *Brunner's glands*, are concentrated at the site where the contents from the stomach and secretions from the liver and pancreas enter the duodenum. These glands secrete large amounts of alkaline mucus that protect the duodenum from the acid content in the gastric chyme and from the action of the digestive enzymes. The activity of Brunner's glands is strongly influenced by autonomic nervous system. For example, sympathetic stimulation causes a marked decrease in mucus production, leaving this area more susceptible to irritation.

2.2. Serous fluid and enzymes secretion

In addition to mucus, the intestinal mucosa produces two other types of secretions. The first is a serous fluid (pH = 6.5 to 7.5) secreted by specialized cells (*i.e.*, crypts of Lieberkuhn) in the intestinal mucosal layer. This fluid, which is produced at the rate of 2000 ml/day, acts as a vehicle for absorption. The second type of secretion consists of surface enzymes that aid absorption. These enzymes are the peptidases or enzymes that separate amino acids and the disaccharidases or enzymes that split sugars.

2.3. Gastrointestinal hormonal secretion. Intestinal hormones

Secretin is secreted by S cells in the mucosa of the duodenum and jejunum in an inactive form called *prosecretin*. When an acid chyme with a pH of less than

4.5 to 5.0 enters the intestine, secretin is activated and absorbed into the blood. Secretin causes the pancreas to secrete large quantities of fluid with a high HCO_3^- concentration and low Cl^- concentration. When administered to experimental subjects secretin also decreases gastric motility and delays gastric emptying, inhibits the action of gastrin on parietal cells and decreases release of antral gastrin in response to stimuli. These actions would all tend to slow the rate of transfer of gastric H^+ to the duodenal lumen, thereby facilitating neutralization in the intestine.

The primary function of *cholecystokinin* (CCK) is stimulation of pancreatic enzyme secretion. It is secreted into the blood by cells of the upper small intestine in response to food protein breakdown products, fatty acids and triglycerides in the lumen. It potentiates the action of secretin, increasing the pancreatic bicarbonate response to low circulating levels of secretin. In addition to its effects on the pancreas, cholecystokinin also stimulates biliary secretion of fluid and bicarbonate and it regulates gallbladder contraction and gastric emptying.

Gastric inhibitory peptide (GIP) which is related structurally to secretin and to glucagon released from the intestinal mucosa in response to increased concentration of glucose and fats, inhibits gastric acid and pepsinogen secretion, gastric motility resulting in delayed gastric emptying time. GIP stimulates also insulin release from pancreatic islet cells.

GIP is a major component of the so-called "enterogastrone" mechanism described many years ago. In this mechanism humoral factors are released by the proximal small intestine in response to food lipids. These factors then inhibit gastric secretion, reduce gastric motility and delay gastric emptying. Although GIP is one hormone involved in this mechanism, other hormones such as secretin or cholecystokinin may also participates

Motilin, which stimulates intestinal motility and contributes to the control of the interdigestive actions of the intestinal neurons, is released from the upper small intestine.

3. DIGESTION AND ABSORPTION IN THE SMALL INTESTINE

The stomach is a poor absorptive structure and only a few lipid-soluble substances, including alcohol, are absorbed from the stomach. That's why it's reasonable to repeat that small intestine is the place where digestion and absorption occur.

Digestion is the process of dismantling foods into their constituent parts. Digestion requires hydrolysis, enzyme cleavage and fat emulsification. Concerning absorption, one should remember that the intestinal mucosa is impermeable to most large molecules. Most proteins, fats, and carbohydrates must be broken down into smaller particles before they can be absorbed. Although some digestion of carbohydrates and proteins begins in the stomach, digestion takes place mainly in the small intestine.

The breakdown of fats to free fatty acids and monoglycerides takes place entirely in the small intestine. The liver, with its production of bile, and the pancreas, which supplies a number of digestive enzymes, play important roles in digestion.

Absorption is the process of moving nutrients and other materials from the external environment of the gastrointestinal tract into the internal environment. Absorption is accomplished by active transport and diffusion. The distinguishing characteristic of the small intestine is its large surface area, which in the adult is estimated to be approximately 250 m^2.

Fats rely largely on the lymphatics for absorption. Each villus is covered with cells called *enterocytes* that contribute to the absorptive and digestive functions of the small bowel and goblet cells that provide mucus. The crypts of Lieberkuhn are glandular structures that open into the spaces between the villi. The enterocytes have a life span of approximately 4 to 5 days; their replacement cells differentiate from progenitor cells located in the area of the crypts. The maturing enterocytes migrate up the villus and eventually are extruded from the tip.

The enterocytes secrete enzymes called *brush border enzymes* that aid in the digestion of carbohydrates and proteins because they adhere to the border of the villus structures. In this way they have access to the carbohydrates and protein molecules as they come in contact with the absorptive surface of the intestine.

This mechanism of secretion places the enzymes where they are needed and eliminates the need to produce enough enzymes to mix with the entire contents that fill the lumen of the small bowel. The digested molecules diffuse through the membrane or are actively transported across the mucosal surface to enter the blood or, in the case of fatty acids, the lacteal. These molecules are then transported through the portal vein or lymphatics into the systemic circulation.

Therefore, resuming, digestion and absorption in the small intestine are facilitated by a series of specializations:

1) different segments of the intestine contain specialized transport mechanisms;
2) mucosal epithelial cells differentiate with respect to specific enzymes and transport mechanisms as the cells ascend from the crypts of Lieberkuhn to the villus tips;
3) each mucosal cell contains specialized organelles:

- the apical microvillus membrane clearly increases the absorptive surface greatly (approximately 25–fold) and contains specific hydrolases and transport proteins;
- the basolateral membranes are similarly differentiated;
- the tight junctions provide the barriers between the external and internal environments of the body;
- the cytosolic organelles perform not only the usual functions of biosynthesis, compartmentalization and molecular sorting found in all cell types, but specialized reactions involved in transcellular transport of nutrients.

3.1. Carbohydrates digestion and absorption

Carbohydrates provide half to two-thirds of the energy requirement. Before they can be absorbed from the small intestine carbohydrates must be broken down into monosaccharides or single sugars.

Digestion of carbohydrates begins in the mouth with the action of salivary α-amylase (Fig. 27–4, left fragment). This digestive process continues in the proximal stomach but is interrupted in the distal stomach as the food is mixed with acidic gastric juices. A pancreatic α-amylase, with a pH optimum of 8 is mixed into the chyme in the duodenum.

The carbohydrates can be only absorbed in the form of monosaccharides. Amylases break down carbohydrates into several disaccharides including maltose, isomaltose, and α-dextrins. The brush border enzymes convert the disaccharides into monosaccharides that can be absorbed. Sucrose yields glucose and fructose, lactose is converted to glucose and galactose, and maltose is converted to two glucose molecules. Therefore, *fructose, glucose and galactose are three end-products of carbohydrates metabolism* in the gastrointestinal tract that can be absorbed in the small intestine (Fig. 27–4, right fragment).

When the disaccharides are not broken down to monosaccharides, they cannot be absorbed but remain as osmotically active particles in the contents of the digestive system, causing diarrhea. Persons who are deficient in lactase, the

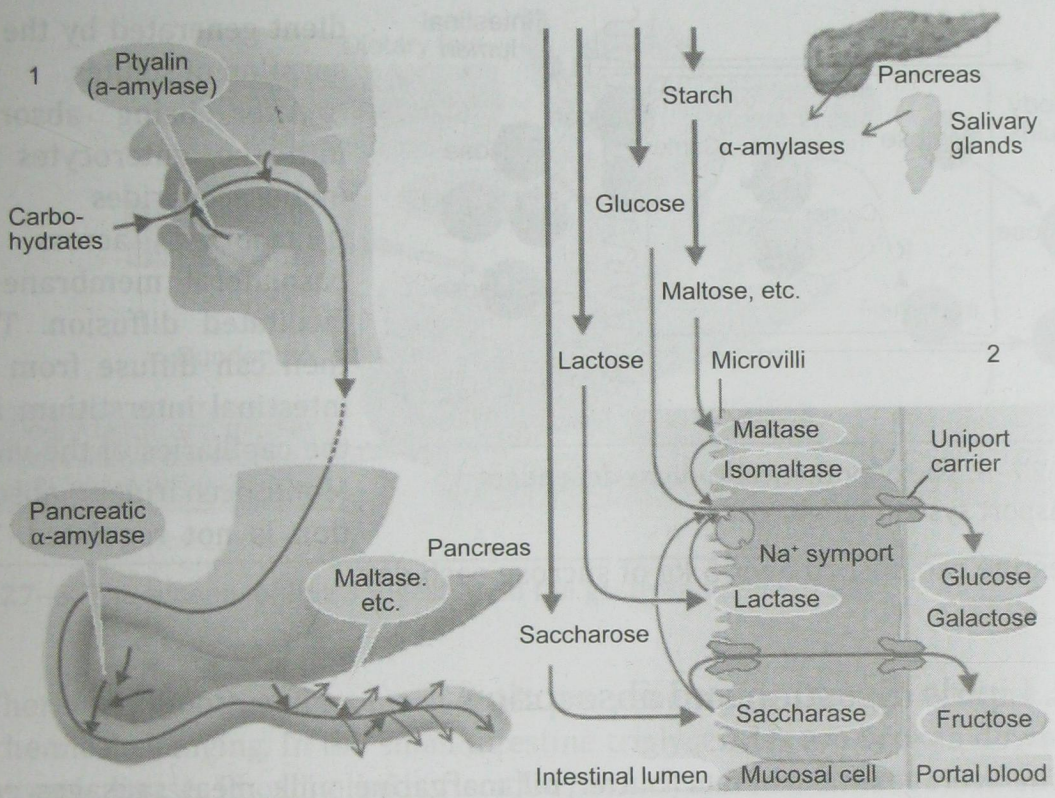


Fig. 27–4. The schematic representation of carbohydrate digestion and monosaccharides absorption.

enzyme that breaks down lactose, experience diarrhea when they drink milk or eat dairy products.

Fructose is transported across the intestinal mucosa *by facilitated diffusion*, which does not require energy expenditure. In this case, fructose moves along a concentration gradient (Fig. 27–4, right fragment).

Glucose and *galactose* are transported by way of a *sodium-dependent transport system* that uses adenosine triphosphate and the Na^+/K^+-ATPase pump as an energy source (Fig. 27–4, right fragment).

Both sodium and glucose must attach to the transport carrier before either can be transported into the cell (Fig. 27–5). The concentration of glucose builds up in the intestinal cell until a diffusion gradient develops causing glucose to move into the body fluids. Sodium is transported out of the cell by the energy dependent Na^+/K^+-ATPase pump. This creates the gradient needed to operate the transport system. Water is passively absorbed along a concentration gra-

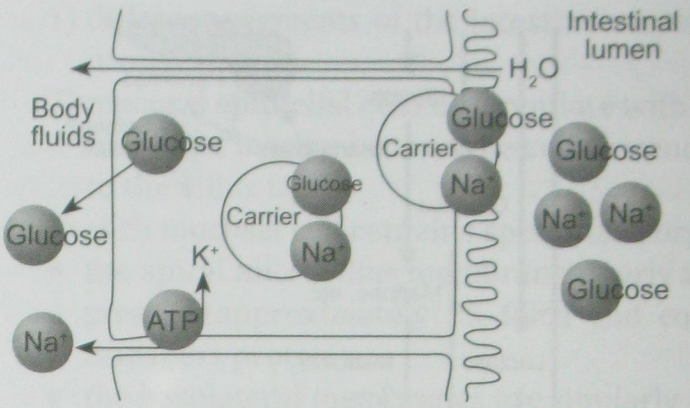


Fig. 27–5. The hypothetical sodium-dependent transport system for glucose.

dient generated by the absorption of solutes.

After being absorbed into the enterocytes the monosaccharides are transported across the basolateral membrane by facilitated diffusion. They then can diffuse from the intestinal interstitium into the capillaries of the villus. Monosaccharides absorption is not regulated. The intestine can absorb over 5 kg of sucrose each day.

3.2. Lipids digestion and absorption

The average intake of fats (butter, oil, margarine, milk, meat, sausages, eggs, nuts etc.) is roughly 60–100 g/day but there is a wide range of individual variation (10–250 g/day). Most fats in the diet (90 %) are neutral fats or triglycerides. The rest are phospholipids, cholesterol esters and lipid-soluble vitamins (A, D, E and K). Over 95 % of the lipids are normally absorbed in the small intestine.

Lipids are a concentrated form of energy. They are also essential for normal body functions, as they form part of cellular membranes and are precursors of bile acids, steroid hormones, prostaglandins, and leukotrienes.

Lipids are poorly soluble in water, so special mechanisms are required for their digestion in the watery environment of the gastrointestinal tract and for their subsequent absorption and transport in plasma. Although small quantities of undegraded triacylglycerol can be absorbed, dietary fats must be hydrolyzed by enzymes before they can be efficiently absorbed.

Lipases, the fat digesting enzymes, originate from the lingual glands, gastric fundus (chief and mucous neck cells) and *pancreas*. About 10–30 % of dietary fat intake is hydrolyzed in the stomach, while the remaining 70–90 % is broken down in the duodenum and upper jejunum. Lingual and gastric lipases have an acid pH optimum, whereas pancreatic lipase has a pH optimum of 7–8. Lipases become active at the fat/oil and water interface (Fig. 27–6).

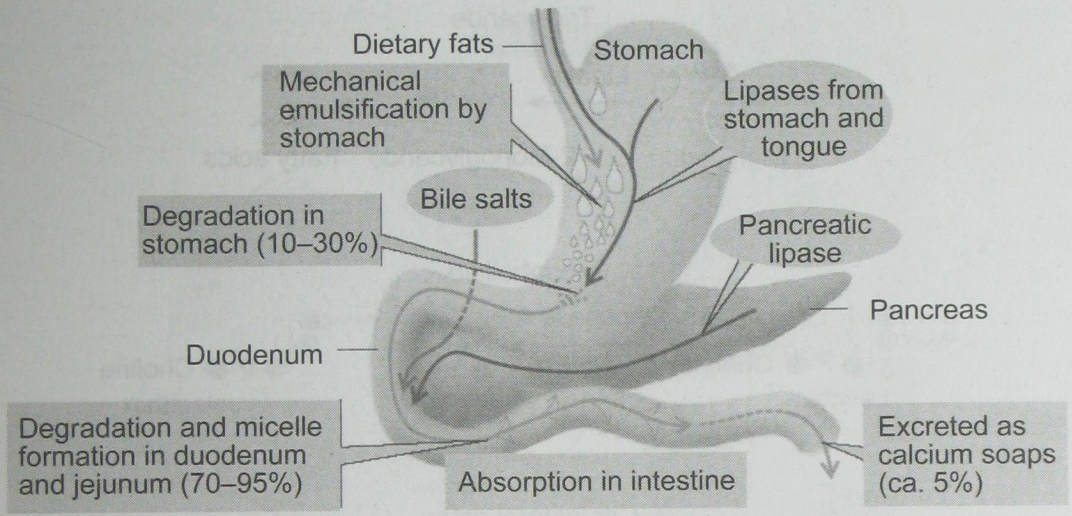


Fig. 27–6. The scheme of lipids digestion in the gastrointestinal tract.

Therefore, mouth and stomach lipids are penetrating through without any biochemical changing. In the small intestine triglycerides are broken down by pancreatic lipase. Phospholipids (lecithin) are cleaved by another pancreatic enzyme — phospholipase A_2. Dietary cholesterol is presented as sterol (cholesterol) ester. Its hydrolysis is catalyzed by the pancreatic enzyme cholesterol esterase (Fig. 27–7).

The major site of lipids absorption is the upper jejunum. Medium-chain triglycerides, with 6 to 10 carbon atoms in their structures, are absorbed better than longer-chain fatty acids because they are more completely broken down by pancreatic lipase and they form micelles more easily. The absorption of vitamins A, D, E, and K, which are lipid-soluble vitamins, requires bile salts.

Therefore, for the further easy explanation we would like to introduce the following *stages of the lipids absorption* in the small intestine:

1. **The lumenal phase involves** (a) lipids emulsification in the stomach and upper intestine for the optimal enzymatic activity of the pancreatic gland because emulsified lipid droplets (1–2 μm; Fig. 27–8, top of the right fragment) provide a much larger surface (relative to the mass of fat) for lipases. Mechanical energy for emulsification is supplied by the antral peristaltic waves of the stomach. Emulsion droplets are important as the physical state of the triglycerides required as substrate for pancreatic lipase. The action of the lipase in the lumen converts one mole of the triglyceride to 2 moles of fatty acid plus 1 mole of the 2-monoglycer-

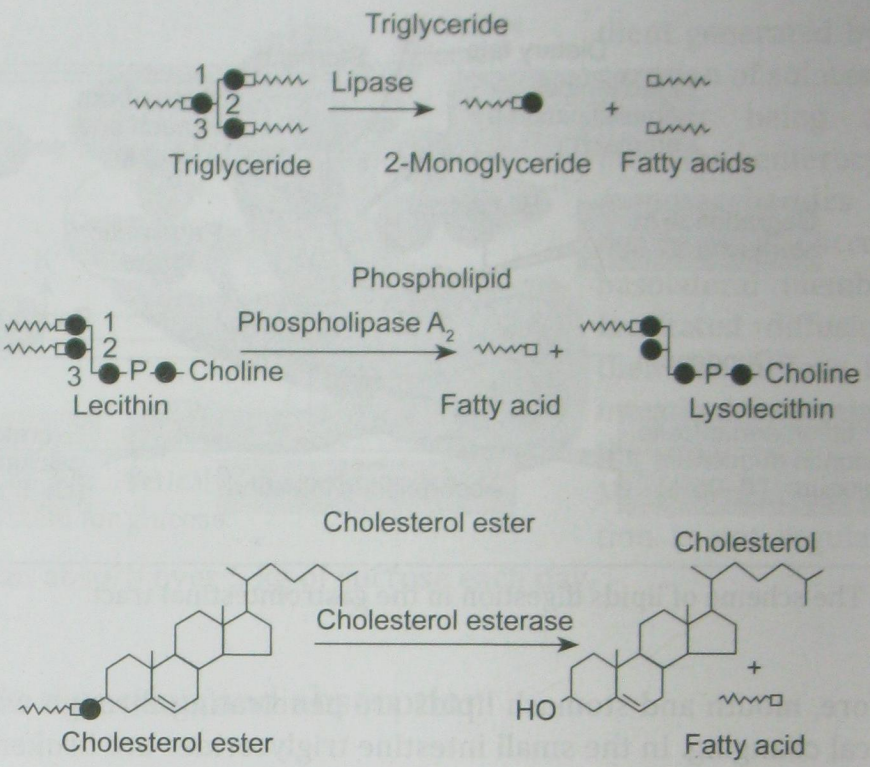


Fig. 27–7. The digestion of the dietary lipids by the pancreatic enzymes in the small intestine.

ide (Fig. 27–8). (b) The mix of aqueous and hydrophobic zones then forms around the enzyme (Fig. 27–8, left fragment). (c) The third process within this stage involves micelle formation. Emulsified products of lipid metabolism (i.e., the end-products of lipid degradation), bile salts, cholesterol and fat-soluble vitamins A, D, E, and K are the main compounds for the micelles formation (Fig. 27–8, bottom of the right fragment). The micelles are only about 20–50 nm in diameter, and their surface-to volume ratio is roughly 50 times larger than that of the lipid droplets in emulsions. They facilitate close contact between the products of fat digestion and the wall of the small intestine and are therefore essential for lipid absorption. The bile salts are on the outside of the micelles. The 2-monoglicerides and lysophosphatides have their hydrophobic chains facing the interior of the micelle and their polar ends facing the surrounding water phase. The cholesterol and lipid-soluble vitamins are located within the lipid-soluble interior of the micelle. Thus, mixed micelles in the unstirred water layer of

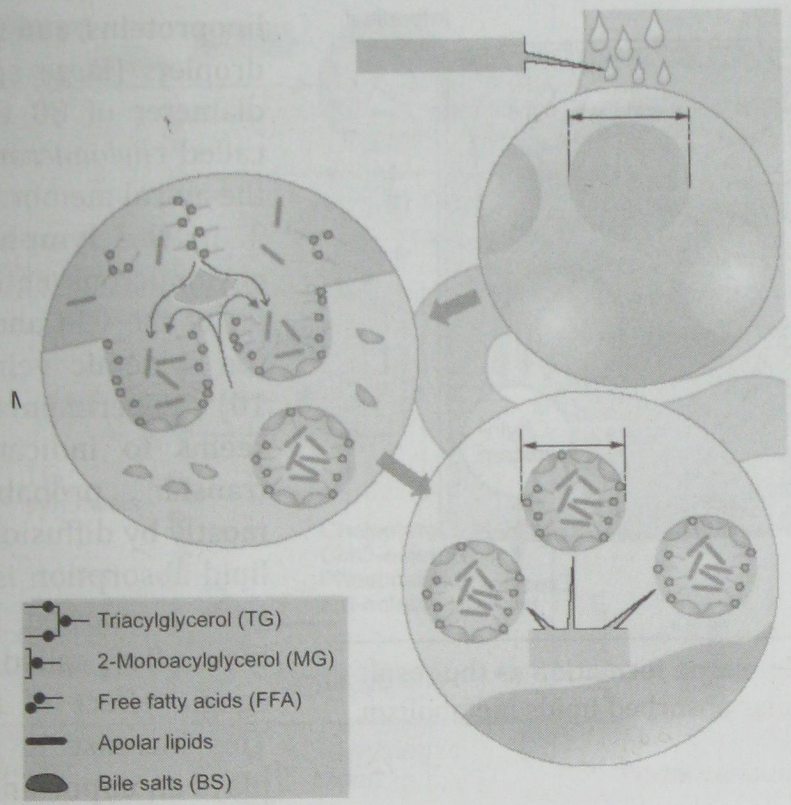


Fig. 27–8. The scheme of lipids degradation and micelles formation.

the microvillus membrane surface give up their contents for absorption across the apical membrane. Lipids inside the micelles cross the basal enterocytes' membrane by *simple diffusion*. The bile salts freed of their associated lipids are absorbed in the terminal ileum by a Na^+-dependent active transport process and re-used through the enterohepatic bile circulation (see Fig. 27–10).

2. **The intracellular phase** occurs inside the enterocytes. Lipids, cholesterol and the lipid-soluble vitamins earlier removed rapidly from the micelles once the micelles make contact with the microvilli. Once inside the enterocytes, the digested lipids enter are reconstituted with the help of the endoplasmic reticulum enzymes. Monoglycerides and fatty acids are rapidly reconstituted to form triglycerides (Fig. 27–9). Cholesterol can be transported out of the enterocytes as free cholesterol or as esterified cholesterol. Reesterified triglycerides are packaged with other lipids (phospholipid, cholesterol, cholesterol esters, vitamins A, D, E and K) by means of

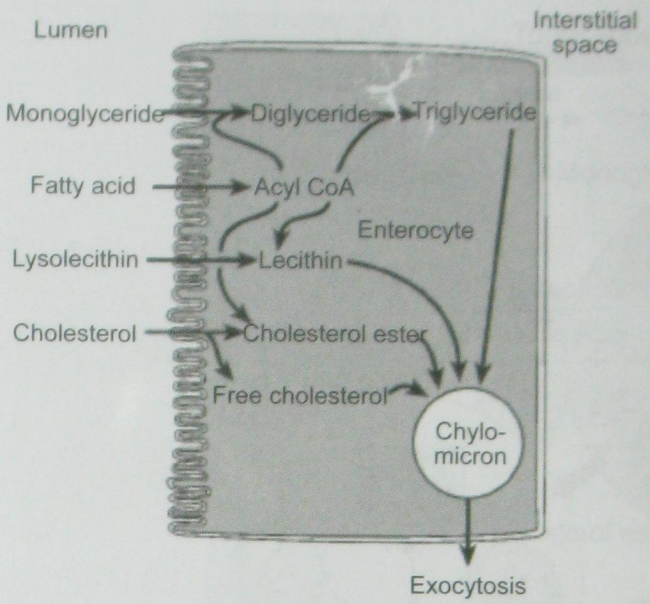


Fig. 27–9. Chylomicron formation as the result of the intracellular absorbed lipids metabolism.

lipoproteins, and the resulting droplets (large spheres with diameter of 80 to 500 nm), called *chylomicrons*, leave via the apical membrane.

3. In **the lymphatic phase** chylomicrons enter lymphatics of the villi and drain into the systemic veins (Fig. 27–10). Experimental evidence seems to indicate that the transfer probably occurs mostly by diffusion. Intestinal lipid absorption is associated with a marked increase in lymph flow called the lymphagogic effect of fat feeding. This increase in lymph flow plays an important role in the transfer of lipoproteins from the intercellular spaces to the central lacteal. After ingesting a fat meal the intestinal lymph and subsequently the peripheral plasma become milky ("lactescent"). Almost all digested lipids are totally reabsorbed by the time the chyme reaches the midjejunum, with the most of the absorption occurring in the duodenum.

Resuming, the following *processes constituting the background of the lipids absorption* in the small intestine:

1. *Lipids emulsification* within the intestinal lumen;
2. *Micelles formation* closely to the basal membrane of the enterocytes;
3. *Chylomicrons formation* inside the enterocytes after micelles rupture and lipids re-esterification;
4. Lipids passage into the lymph and finally — into blood (Fig. 27–10).

Fat that is not absorbed in the intestine is excreted in the stool. *Steatorrhea* is the term used to describe fatty stools. It usually indicates that there is 20 g or more of fat in a 24-hour stool sample. The pathophysiology of steatorrhea may involve the following mechanisms: (a) poor emulsification (stomach); (b) bile salt deficiencies; (c) pancreatic lipase deficiency; (d) decreased microvil-

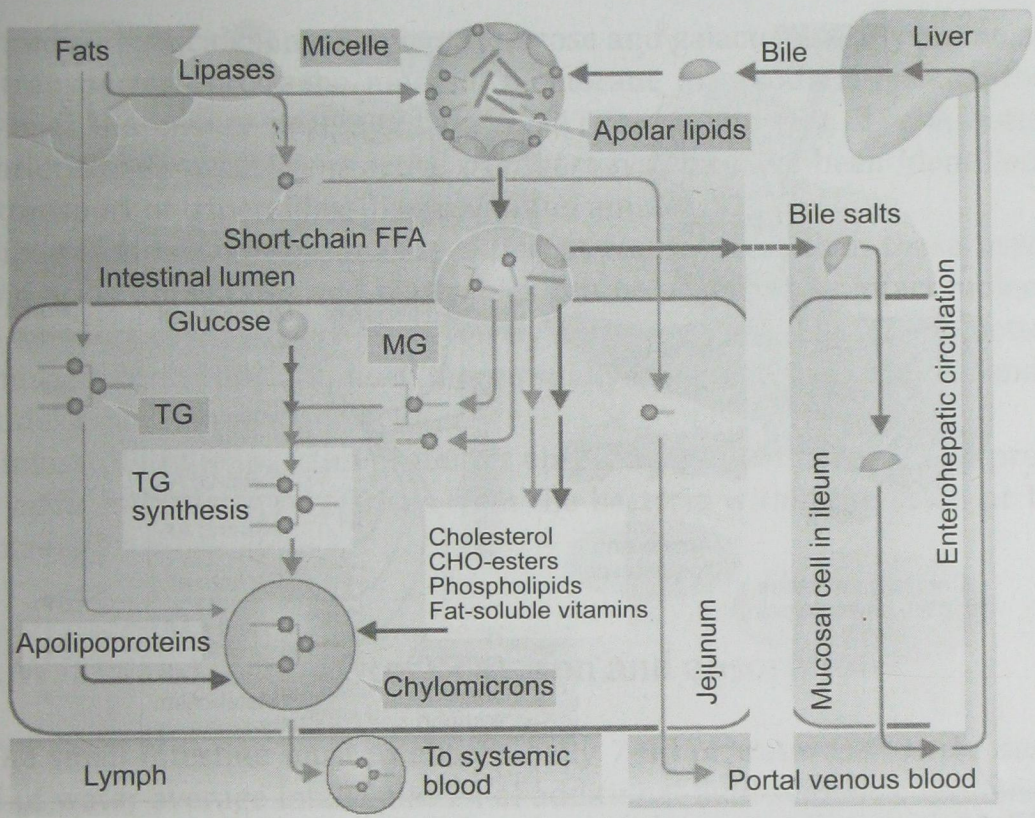


Fig. 27–10. The scheme of the lipids absorption in the small intestine.

lus surface area; (e) deficient formation of chylomicrons; (f) blockage of lymph flow.

3.3. Proteins digestion and absorption

Proteins from the diet must be broken into amino acids to be absorbed. Protein digestion begins in the stomach with the action of pepsin (Fig. 27–11, left fragment). Pepsinogen, the enzyme precursor of pepsin, is secreted by the chief cells in response to a meal and acid pH (0.5-2.0). Acid in the stomach is required for the conversion of pepsinogen to pepsin. Pepsin is inactivated when it enters the intestine by the alkaline pH.

Proteins are broken down further by pancreatic enzymes, such as trypsin, chymotrypsin, carboxypeptidase and elastase. They are secreted as inactive proenzymes and subsequently activated in the gut lumen. This is presumed to

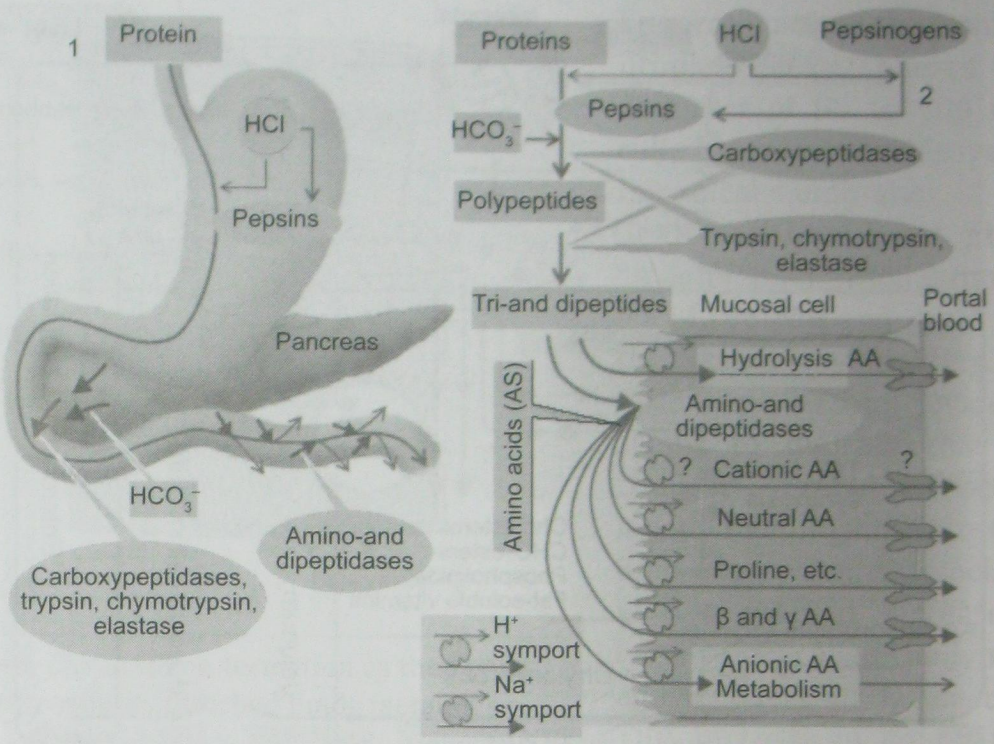


Fig. 27–11. Protein digestion and absorption of amino acids and oligopeptides.

protect the pancreas from self-digestion. Trypsinogen, which lacks enzymatic activity, is activated by an enzyme located on the brush border cells of the duodenal enterocytes. Activated trypsin activates additional trypsinogen molecules and other pancreatic precursor proteolytic enzymes (Fig. 27–11, right fragment).

The endopeptidases *trypsin, chymotrypsin* and *elastase* hydrolyze the protein molecules into short-chain peptides. *Carboxypeptidase A and B* (from the pancreas) as well as *dipeptidases* and *aminopeptidase* (brush border enzymes) act on proteins at the end of the peptide chain, breaking them down into tripeptides, dipeptides, and (mostly) individual amino acids. Therefore, *tripeptides, dipeptides and aminoacids are the end-product of the protein metabolism.* These cleavage products are absorbed in the duodenum and jejunum.

Absorption across the microvillus membrane of single amino acids, di- and tripeptides occurs. Tetrapeptides and larger peptides cannot be absorbed as such and are broken down by hydrolases in the microvillus membrane.

The amino acids are liberated intramurally or on the surface of the villi by brush border enzymes that degrade proteins into peptides that are one, two

or three amino acids long. Similar to glucose and galactose, many amino acids are transported across the mucosal membrane in a sodium-linked process that uses the energy dependent Na^+/K^+-ATPase pump (Fig. 27–10, bottom). A variety of *Na^+-dependent active transport systems* have been identified for the transport of tripeptides, dipeptides and aminoacids.

Dipeptides and tripeptides use different transporters than those used by amino acids. Dipeptides and tripeptides can be absorbed as intact molecules by a symport carrier which is driven by an H^+ gradient (Fig. 27–10, bottom). Amino acids generally are much more rapidly absorbed as dipeptides and tripeptides than as free amino acids.

Almost all of the ingested protein is absorbed by the intestine. Any protein that appears in the stool derives from the bacteria within the colon or from cellular debris.

3.4. Water and electrolytes digestion and absorption

The small intestine must to absorb daily 7-8 l of water that is the sum of 1.5 l of water average intake and 7 l of additional fluid secretion contained in salivary, gastric, biliary and pancreatic secretions. Approximately 0.1 l/day of water is eliminated in the feces.

Water has to be absorbed in the gastrointestinal tract mainly in the jejunum and ileum, with smaller quantities being absorbed by the colon (Fig. 27–11). Water is driven through the intestinal epithelium *passively, by osmosis.*

When solutes (Na^+, Cl^- etc.) are absorbed in the intestine, water follows (Fig. 27–12).

Water absorption is mainly driven by the *absorption of Na^+, Cl^- and organic compounds* (Fig. 27–12). The luminal concentration of Na^+ and Cl^- steadily decreases from the duodenum to the colon. That of Na^+, for example, is approximately 145 mmol/l in the duodenum, 125 mmol/l in the ileum and only 40 mmol/l in the colon.

Na^+ is absorbed by various mechanisms, and the *Na^+-K^+-ATPase* on the basolateral cell membrane is the primary driving mechanism for all of them (Fig. 27–13).

Firstly, Na^+ and Cl^- are transported from the lumen into the enterocytes. Then they are transported across the basolateral membrane into the intestinal interstitium.

Na^+ passively influxes into cells of the duodenum and jejunum in three ways — through a Na^+-glucose, Na^+-amino acid and Na^+-(di or tri-) peptide

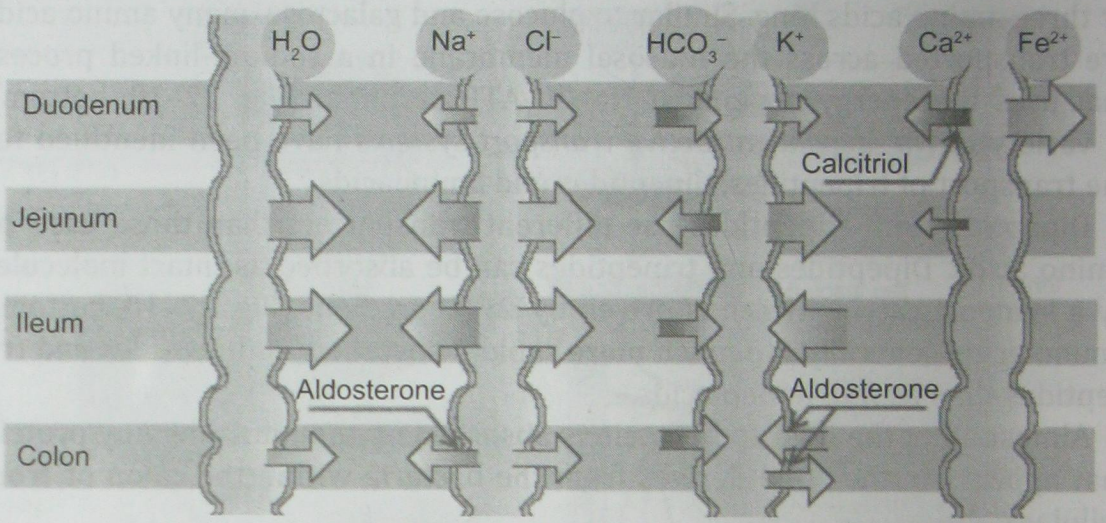


Fig. 27–12. Water and electrolyte absorption in the gastrointestinal tract.

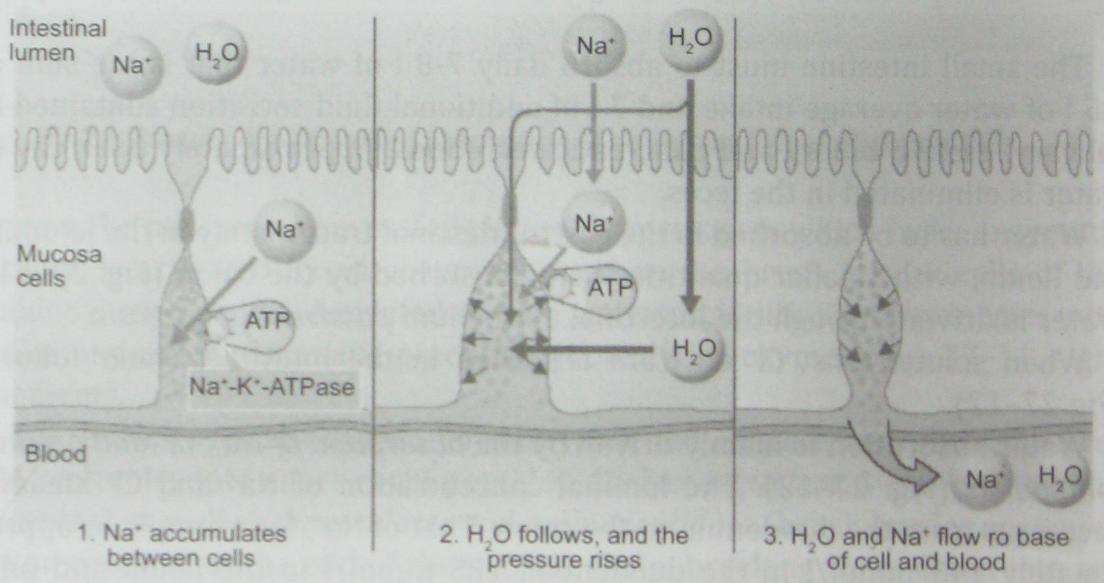


Fig. 27–13. Schematical representation of Na+ and water absorption in the gastrointestinal tract.

co- transport systems (Fig. 27–14, fragment 1). About 30 % of Na^+ transported into the cell by a neutral Na^+-Cl^- co-transport system. The rest Na^+ ions enter the cell passively down an electrochemical gradient that drives also Cl^- out of the lumen forms (Fig. 27–14, fragment 2).

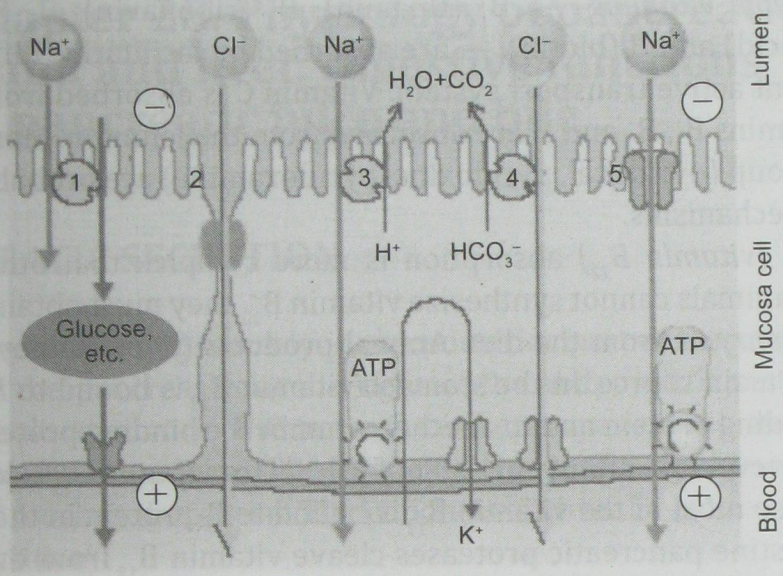


Fig. 27–14. Schematical representation of Na+ and water absorption in the gastrointestinal.

Na^+ ions in the lumen of the ileum are exchanged for H^+ ions (Fig. 27–14, fragment 3) while Cl^- is exchanged for HCO_3^- at the same time (Fig. 27–14, fragment 4).

Resuming, most Na^+, Cl^- and, by subsequent osmosis, H_2O is absorbed by this electroneutral transport mechanism.

Na^+ in the colon is mainly absorbed through luminal *Na^+ channels* (Fig. 27–14, fragment 5). This type of Na^+ transport is electrogenic and aldosterone-dependent. The related mechanism either leads to K^+ secretion or drives Cl^- out of the lumen (Fig. 27–12 and Fig. 27–14, fragment 2).

Cholera toxin inhibits the GTPase of the Gs proteins, thereby maintaining a maximal cAMP concentration and therefore a marked increase in Cl^- secretion. In response to it, large quantities of water and Na+ are secreted into the lumen, which can lead to severe diarrhea (up to 1 l/hour!).

3.5. Vitamins and minerals digestion and absorption

Fat-soluble vitamins — A (retinol), D_3 (cholecalciferol), E (tocopherol) and K — are incorporated into micelles for absorption (see 3.2 of this chapter) and are absorbed along with other lipids in the proximal intestine.

Water-soluble vitamins — B_1 (thiamin), B_2 (riboflavin), B_6 (pyridoxine), C (ascorbic acid) and H (biotin) — are absorbed by facilitative diffusion or via Na^+-dependent active transport system. Vitamin C is absorbed from the ileum, whereas vitamins B_1, B_2 and H are absorbed from the jejunum. Members of the vitamin B_6 group (pyridoxal, pyridoxine, pyridoxamine) are probably absorbed by passive mechanisms.

Cobalamin (vitamin B_{12}) absorption is more complex that other vitamins. Since higher animals cannot synthesize vitamin B_{12} they must obtain this *cobalt-containing coenzyme* from the diet. Animal products (liver, kidneys, fish, eggs, milk) are the main source. In the stomach vitamin B_{12} is bound to *R-protein* — a specific binding protein and to another vitamin B_{12}-binding protein — *intrinsic factor*. However, the affinity of the intrinsic factor vitamin B_{12} is less than that of R-protein, so most of the vitamin B_{12} is bound to R-protein in the stomach.

In the intestine pancreatic proteases cleave vitamin B_{12} from the R-protein, allowing it to bind to trypsin-resistant intrinsic factor. The intrinsic factor-vitamin B_{12} complex binds to a receptors on ileal enterocytes and is absorbed by *receptor-mediated endocytosis*.

Calcium (Ca^{2+}) absorption within the small intestine is regulated to maintain Ca^{2+} balance. The stool contains one-third of the dietary Ca^{2+} intake. Ca^{2+} is absorbed in the upper part of the small intestine (Fig. 27–14) with the aid of intracellular *calcium*-binding protein. It should be mentioned that calcitriol increases calcium-binding protein synthesis, thus increasing Ca^{2+} absorption. Vitamin D or substances that form water-insoluble compounds with Ca^{2+} (oxalate, fatty acids) deficiencies decrease Ca^{2+} absorption from the gastrointestinal tract.

Iron (Fe) is absorbed by a different mechanism (iron metabolism see in Chapter # 13). Iron absorption is necessary to maintain the normal iron balance. Iron is absorbed primarily within the *duodenum* and *jejunum* either as heme (derived from meat) or as a free ion. The ferrous ion (Fe^{2+}) is absorbed more efficiently than the ferric ion (Fe^{3+}).

An *iron deficiency* inhibits hemoglobin synthesis leading to hypochromic microcytic anemia. An *iron overload* most commonly damages the liver, pancreas and myocardium (hemochromatosis).

Chapter 28. Physiology of pancreatic gland and liver. Digestive functions of pancreatic juice and bile

1. PANCREATIC SECRETION

The pancreas lies transversely in the posterior part of the upper abdomen. The head of the pancreas is at the right of the abdomen; it rests against the curve of the duodenum in the area of the hepatopancreatic ampulla and its entrance into the duodenum. The body of the pancreas lies beneath the stomach. The tail touches the spleen. The pancreas is virtually hidden because of its posterior position.

The human pancreas is located in close apposition to the duodenum. It performs both endocrine and exocrine functions, but here we discuss only its exocrine function. (The endocrine functions of pancreatic gland are discussed in Chapter # 11.)

The exocrine pancreas is composed of numerous small, sac-like dilatations called acini composed of a single layer of pyramidal *acinar cells*. These cells are actively involved in the production of enzymes. Zymogen granules are observed in the apical region of acinar cells, which secrete digestive enzymes into a system of microscopic ducts. These ducts empty into the main pancreatic duct, which extends from left to right through the substance of the pancreas. Main pancreatic duct and the bile duct unite to form the hepatopancreatic ampulla, which empties into the duodenum. The sphincter of the pancreatic duct controls the flow of pancreatic secretion into duodenum.

Pancreatic juice functions. One of the major functions of pancreatic secretion is *to neutralize the acids* in the chyme when it enters the duodenum from the stomach. This mechanism is important because pancreatic enzymes operate optimally near neutral pH. Another important function is the *production of enzymes* involved in the digestion of dietary carbohydrate, fat and protein.

Pancreatic juice composition. The pancreas secretes about 1200–1500 ml per day of HCO_3^- rich fluid. The osmolality of pancreatic fluid is equal to that of plasma at all secretion rates. The Na^+ and K^+ concentrations of pancreatic juice are the same as those in plasma, but unlike plasma, pancreatic juice is enriched with HCO_3^- and has a relatively low Cl^- concentration. The HCO_3^- concentration

increases with increases in secretion rate and reaches a maximal concentration yielding a solution with a pH = 8.2.

HCO_3^- is secreted from the luminal membrane of the ductules via an anion exchanger that simultaneously reabsorbs Cl^- from the lumen (Fig. 28–1, fragment 1). Cl^- returns to the lumen via a Cl^- channel which is more frequently opened by secretin to ensure that the amount of HCO_3^- secreted is not limited by the availability of Cl^- (Fig. 28–1, fragment 2). The HCO_3^- involved is the product of the CO_2+OH^- reaction catalyzed by carbonic anhydrase. For each HCO_3^- molecule secreted, one H^+ ion leaves the cell on the blood side through an Na^+/ H^+ exchanger (Fig. 28–1, fragment 3).

Pancreatic juice contains the following elements:

1. Water and electrolytes which are secreted by ductular and duct cells.
2. Digestive enzymes which are secreted by acinar cells. Pancreatic enzymes could be classified as following:

- Pancreatic amylases — main representative is alpha-pancreatic amylase which is secreted in the active form and has to hydrolyze complex carbohydrates within the chyme in the duodenum.
- Pancreatic lipases — main representatives are pancreatic lipase and phospholipase A_2 which are also secreted in the active form and have to cleave the triglycerides, phospholipids, cholesterol within the chyme in the duodenum.

  Pancreatic proteases are classified as endopeptidases or exopeptidases — main representatives are trypsinogen, chymotrypsinogen and procarboxypeptidase which are secreted in the inactive form. Their activation sequences are the following as it is shown on the Fig. 28–2.

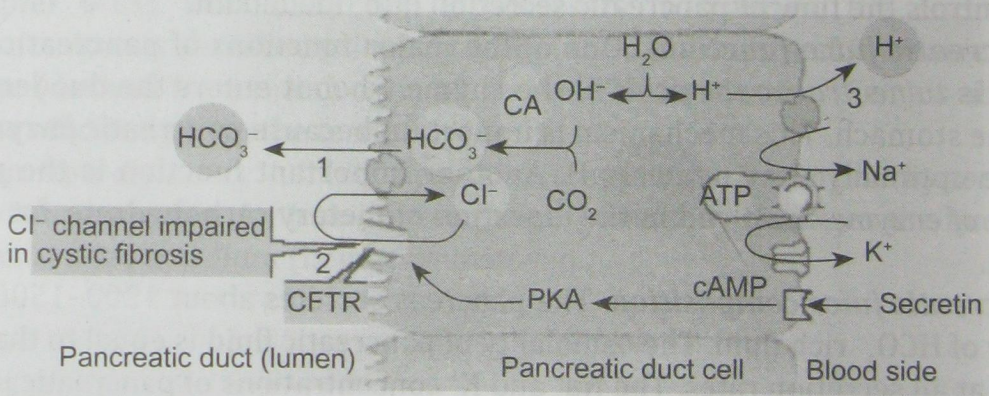


Fig. 28–1. Secretion in pancreatic duct cells.

When the pancreatic juice enters the duodenum, trypsinogen is converted to trypsin by *enterokinase*, an enzyme found on the luminal surface of enterocytes. The active *trypsin* then converts the other proenzymes to active enzymes.

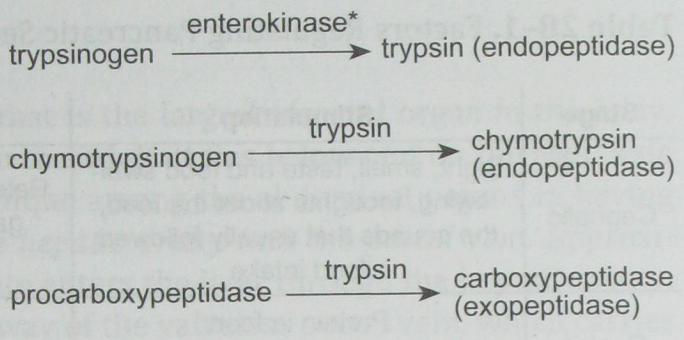


Fig. 28–2. The biochemical sequence of pancreatic endopeptidases activation.

Endopeptidases hydrolyze certain internal peptide bonds of proteins or polypeptides to release the smaller peptides.

3. *Trypsin* splits off basic amino acids from the carboxyl terminal of a protein.
4. *Chymotrypsin* attacks peptide bonds with an aromatic carboxyl terminal.
5. *Elastase* attacks peptide bonds with a neutral aliphatic carboxyl terminal.

The *exopeptidases* in pancreatic juice are *carboxypeptidase A* and *carboxypeptidase B*. As the result, the final products of protein digestion are amino acids and small peptides.

- Trypsin inhibitor synthesizes together with endopeptidases proforms and serves for the pancreatic parenchyma protection from the self-ingestion.

Stages of pancreatic secretion. The stimulation of pancreatic secretion can be divided into three phases: the cephalic phase, the gastric phase and the intestinal phase. The regulation of pancreatic secretion by various hormonal and neural factors is summarized below (Table 28–1).

Seeing, smelling, tasting, chewing, swallowing or thinking about food results in the secretion of a pancreatic juice rich in enzymes. In *cephalic phase* stimulation of pancreatic secretion is mainly mediated by direct efferent impulses sent by vagal centers in the brain to the pancreas and, to a minor extent, by the indirect effect of parasympathetic stimulation of gastrin release.

The *gastric phase* is initiated when food enters the stomach and distends it. Pancreatic secretion is then stimulated by vago-vagal reflex. Gastrin is also involved in this phase pancreatic secretion when aminoacids inside the stomach stimulated its release from G- cells of the stomach mucosa.

Table 28–1. Factors Regulating Pancreatic Secretion After a Meal

Stage	Stimulation	Mediators	Response
Cephalic	Sight, smell, taste and food swallowing, thoughts about the food, the sounds that usually followed food intake	Release of ACh and gastrin by vagal stimulation	Increased pancreatic secretion with greater effect on enzyme output
Gastric	Protein in food	Gastrin	
	Gastric distension	Vago-vagal reflex	
Intestinal	Amino acids and peptides	Cholecystokinin and vago-vagal reflex	
	Long-chain fatty acids		
	Acid in chyme (pH < 4.5)	Secretin	Increased HCO_3^- secretion

During the most important — *the intestinal phase* — the entry of acidic chyme from the stomach into the small intestine stimulates the release of secretin (Fig. 28–3). When the pH of the lumen in the duodenum decreases, the secretin concentration in plasma increases. This response is followed by an increase in HCO_3^- output by the pancreas. Cholecys-tokinin (CCK) potentiates secretin-induced pancreatic juice secretion in this situation.

When chyme with big amount of free amino acids and long-chain fatty acids enters duodenum pancreatic juice content is characterized by bigger pancreatic endopeptidases content. This type of pancreatic juice secretion is stimulated by cholecystokinin. Secretin potentiates secretin-induced pancreatic juice secretion in this situation. It should be mentioned also that any type of chyme penetrating through the pyloric sphincter distends the walls of the duodenum and thus stimulates pancreatic juice releasing via local vago-vagal reflex (Table 28–1). Thus, acetylcholine is known as the potent compound that could potentiate pancreatic juice secretion.

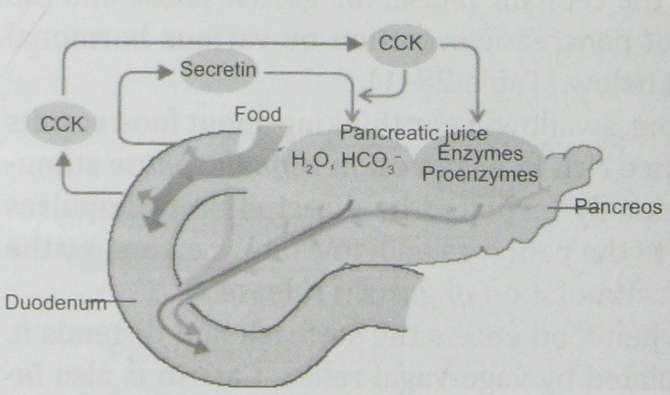


Fig. 28–3. Control of pancreatic juice secretion.

2. BILE SECRETION

Bile is the secret of the liver that is the largest visceral organ in the body, weighing approximately 1.3 kg in the adult. It lies below and on the right side of the diaphragm. The liver is unique among the abdominal organs in having a dual blood supply through the *hepatic artery* and the *portal vein*. Approximately 300 ml of blood per minute enters the liver through the hepatic artery; another 1050 ml/min enters by way of the valveless portal vein, which carries blood from the stomach, the small and the large intestines, the pancreas, and the spleen (Fig. 28–4). Therefore, blood from the gastrointestinal tract, spleen, and pancreas travels to the liver by way of the portal vein before moving into the vena cava for return to the heart. This will important in our further discussion about *enterohepatic bile circulation*.

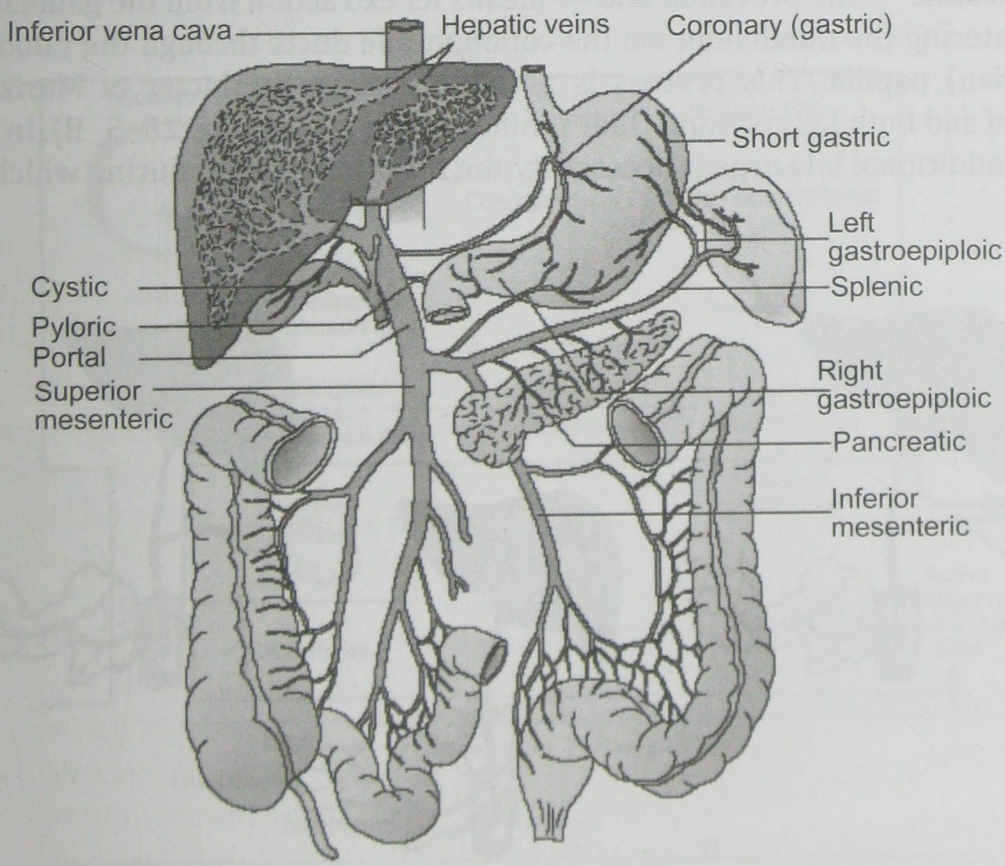


Fig. 28–4. The portal circulation.

The liver is one of the most versatile and active organs in the body. It produces bile; metabolizes hormones and drugs; synthesizes proteins, glucose, and clotting factors; stores vitamins and minerals; changes ammonia produced by deamination of amino acids to urea; and converts fatty acids to ketones. In its capacity for metabolizing drugs and hormones, the liver serves as an excretory organ. We will discuss here only bile-secreting function of the liver.

Bile secretes inside the hepatocytes. It goes after into the gallbladder where it accumulates and stores until the following food intake when it will be necessary for the lipid products emulsification. Bile entering into the initial part of the small intestine — the duodenum — is the reflectory process that has 3 stages. During the *first stage* bile moves from the hepatocytes into the gallbladder through the hepatic ducts under the pressure of 400 H_20 sm (Fig. 28–5, A). Both sphincters of Lutkens (located at the hepatic exit and the hepatic ducts set up) and Merizzi (located just before the gallbladder) are opened and sphincter of Oddi is closed at this moment. The *second stage* of bile admission into the intestine — the prevailed one — means its extraction from the gallbladder and entering the duodenum via the common bile ducts through the duodenal (Faterian) papilla. This process is possible because sphincter of Merizzi is opened and both Lutkens and Oddi sphincters are closed (Fig. 28–5, B). In case of the additional bile urgent necessisity the *third stage* starts during which the

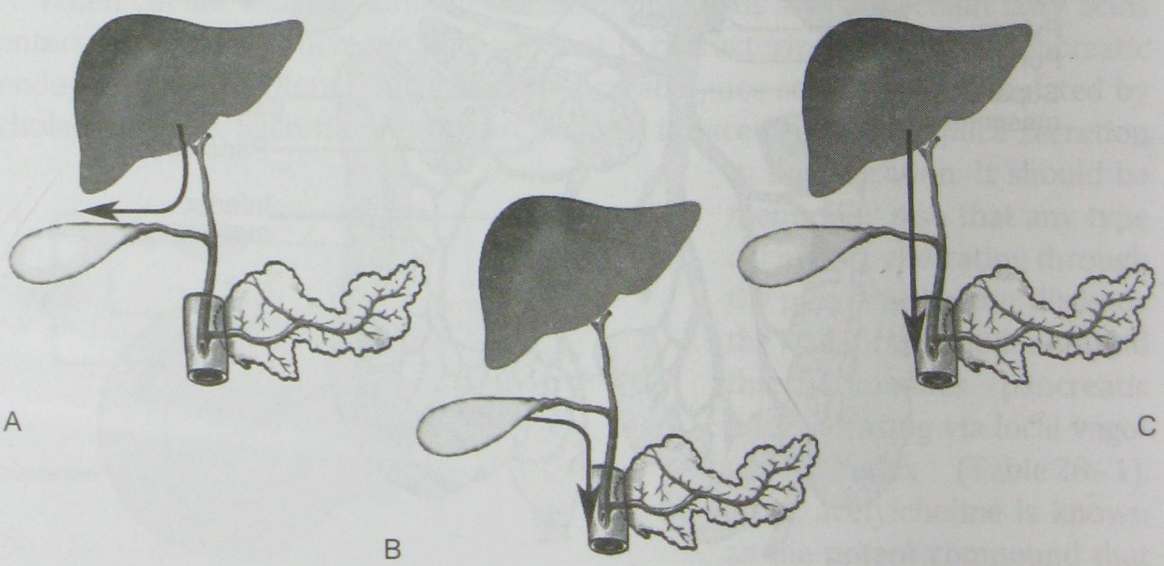


Fig. 28–5. The scheme of the bile secretion.

pure hepatic bile directly enters duodenum without gallbladder transit. The only Lutkens sphincter is opened, both Merizzi and Oddi sphincters are closed at this moment (Fig. 28–5, C).

Bile functions. The secretion of bile, approximately 600 to 1200 ml daily, is one of the many functions of the liver. Bile functions in the digestion and absorption of fats and fat-soluble vitamins from the intestine, and it serves as a vehicle for excretion of bilirubin, excess cholesterol, and metabolic end-products that cannot be eliminated in the urine.

Bile content. Bile contains water, electrolytes, bile salts, bile pigments (e.g., bilirubin), cholesterol and certain products of organic metabolism. Fig. 28–6 summarizes the major components of the bile with the following description.

Bile acids. Bile acids are classified as primary and secondary. Primary bile acids known as cholic acid and chenodeoxycholic acid are synthetized by the liver from cholesterol. Bile acids are secreted as conjugates of glycine or tau-

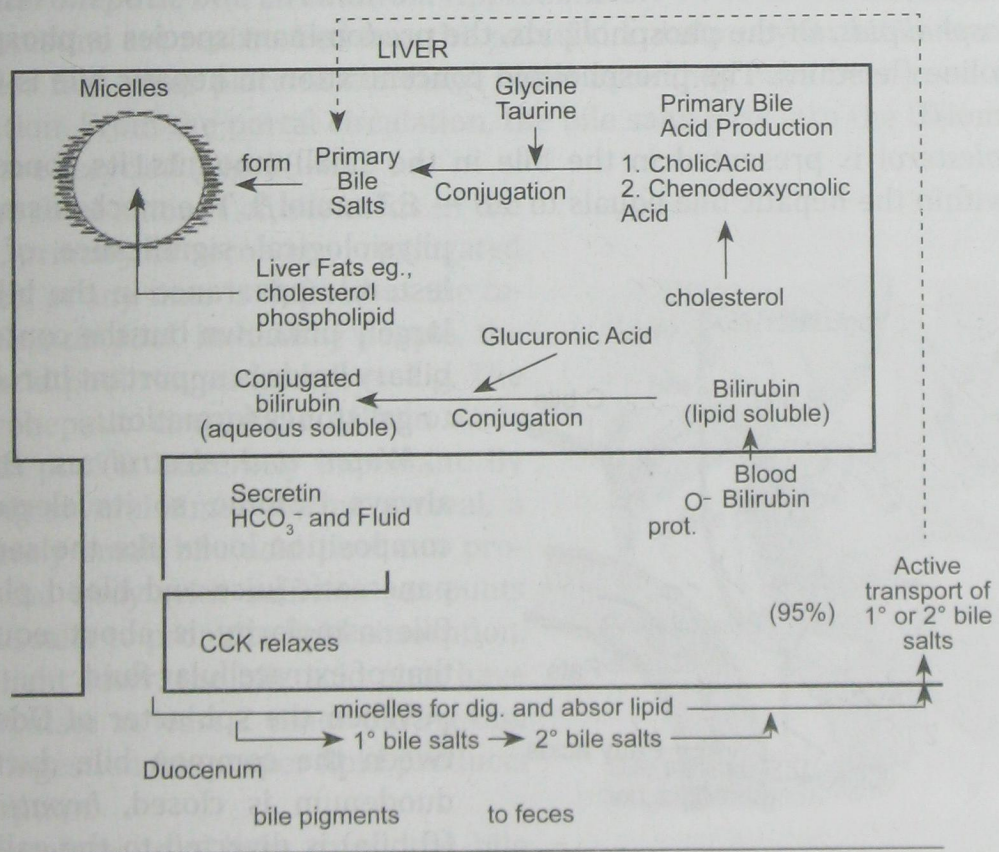


Fig. 28–6. The schematic description of the bile content.

rine. When bile enters the duodenum bacteria present in the lumen act on the primary bile acids and convert them to secondary bile acids by dehydroxylation. Cholic acid is converted to deoxycholic acid and chenodeoxycholic acid to lithocholic acid.

At a neutral pH, the bile acids are mostly ionized and are referred to as *bile salts* which are much more polar than bile acids and thus have greater difficulty penetrating cell membranes. Consequently, bile salts are absorbed much more poorly by the small intestine than bile acids. This property of bile salts is important because they play an integral role in the intestinal absorption of lipid. Therefore, it is important that bile salts are absorbed by the small intestine only after all of the lipid has been absorbed.

Bile pigments — mainly, bilirubin and biliverdin. Bilirubin is secreted by hepatocytes via an active process. Shortly, *bilirubin* is an end-product of hemoglobin degradation in the monocyte-macrophage system in the spleen, bone marrow and liver. Full physiological mechanism of bilirubin formation will be given below.

Phospholipids. Of the phospholipids, the predominant species is phosphatidylcholine (lecithin). The phospholipid concentration in hepatic bile is 0.3 to 11 mmol/l.

Cholesterol is presented in the bile in the small amounts, its concentration within the hepatic bile equals to 1.6 — 8.3 mmol/l. The mechanisms and physiological significance of cholesterol appearance in the bile are largely unknown but the content of biliary lipids is important in relation to gallstones formation.

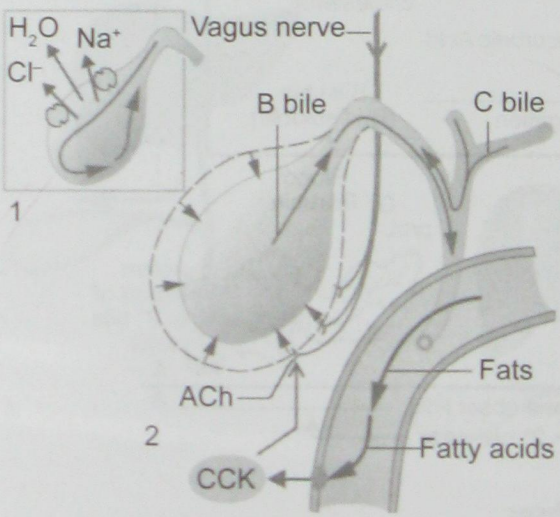


Fig. 28–7. The scheme of gallbladder bile accumulation and extraction.

Water and electrolytes. Bile is always isotonic, so its electrolyte composition looks like the same as pancreatic juice and blood plasma. Bile osmolarity is about equal to that of extracellular fluid.

When the sphincter of Oddi between the common bile duct and duodenum is closed, *hepatic bile* (C bile) is diverted to the gallbladder, where it is highly (1 : 10) concentrated and stored (Fig. 28–7).

That's why *gallbladder bile* (B bile) has a very different composition from hepatic bile. Water absorption is the major mechanism involved in concentrating hepatic bile by the gallbladder. It should be mentioned that water, Na^+ and Cl^- reabsorption by the gallbladder epithelium is passive (Fig. 28–7, fragment 1). As a result of isotonic fluid absorption from the gallbladder bile, the concentration of the various unabsorbed components of hepatic bile increases dramatically — till 10-20-fold. Having gallbladder volume appr. 50 ml it it possible that duodenal lumen could enter as much as till 0.5-1.0 l of the bile. This bile is used for fat digestion, in this case the gallbladder contracts and its contents are mixed in portions with the duodenal chyme (Fig. 28–7, fragment 2).

Gallbladder contraction is *triggered* by CCK and the neuronal plexus of the gallbladder wall, which is innervated by preganglionic parasympathetic fibers of the vagus nerve (Fig. 28–7, fragment 2). The sympathetic nervous system *inhibits* gallbladder contractions via a_2- adrenoreceptors located on cholinergic fiber terminals.

Enterohepatic bile circulation. Approximately 94 % of bile salts that enter the intestine are reabsorbed into the portal circulation by an active transport process that takes place in the distal ileum once they have been used for fat digestion. From the portal circulation, the bile salts pass into the liver, where they are recycled (Fig. 28–8).

The total amount of bile acids in the body, primary or secondary, conjugated or free, at any time is defined as the *total bile acid pool.* In healthy people, the bile acid pool ranges from 2 to 4 g. The enterohepatic circulation of bile acids in this pool is extremely important. By cycling several times during a meal, a relatively small bile acid pool can provide the body with sufficient amounts of bile salts to promote lipid absorption. In a light eater, the bile acid pool have to circulate at least 2 times during one food ingestion for proper lipid products absorption.

Therefore, in a light eater, the bile acid pool may circulate 6 to 8 times a day; in a heavy eater, it may circulate 14

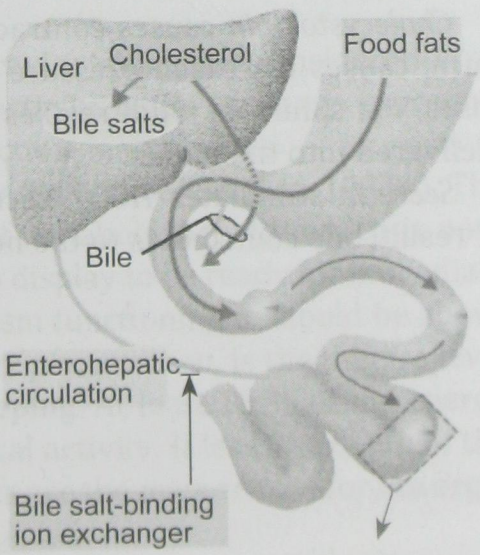


Fig. 28–8. The scheme of enterohepatic circulation.

to 16 times a day before being expelled in the feces. Such system for bile recirculation is called the *enterohepatic circulation* (Fig. 28–8).

Absorbed bile salts are transported in the portal blood bound to albumin or high-density lipoproteins. The uptake of bile salts by hepatocytes is extremely efficient. In just one pass through the liver, more than 80 % of the bile salts in the portal blood is removed. Once taken up by hepatocytes, bile salts are secreted into bile. The uptake of bile salts is a primary determinant of bile salt secretion by the liver.

Bile secretion regulation. The biliary system is supplied by parasympathetic and sympathetic nerves. Parasympathetic (vagal) stimulation increases bile formation through gallbladder contraction and sphincter of Oddi relaxation. By contrast, sympathetic nervous system stimulation results in reduced bile secretion and relaxation of the gallbladder.

The major determinant of bile acid synthesis and secretion by hepatocytes is the bile acid concentration in hepatic portal blood, which exerts a *negative-feedback* effect on the synthesis of bile acids from cholesterol.

Bile could be regulated directly and indirectly.

Gastrin stimulates bile secretion directly by affecting the liver and indirectly by stimulating increased bile acid production that results in increased secretin release.

Steroid hormones (for instance, estrogen and some androgens) are known to inhibited bile secretion.

Cholecystokinin causes contraction of the gallbladder (Fig. 28–7) which, in turn, causes increased pressure in the bile ducts. As the bile duct pressure rises, the sphincter Oddi relaxes (another cholecystokinin effect), and bile is delivered into the lumen.

Secretin stimulates HCO_3^- secretion by the cells lining the bile ducts. As a result, bile contributes to the neutralization of acid in the duodenum.

Unit X ENERGY EXCHANGE AND TEMPERATURE PHYSIOLOGY

Chapter 29. Energy exchange. Basal exchange. Nutrition

1. ENERGY TRANSFER IN AN ORGANISM

Energy transfer is one of the signs of the living organism, it correlates with its levels of functional activities and might accompany both normal and/or abnormal manifestation of the internal processes. Here we are not keeping about the exact biochemical details concerning the processes of nutrients biochemical transformation inside the human organism. The basic physiological processes will be observed.

There are 3 types of the cells' metabolic activity taking into consideration all explained above. The first one, *the level of activity*, is the exchange processes intensity throughout the specific cells' function maintenance — i.e., excitation, secretion, contraction and others (it should be not less than 100 %). The second, *the level of readiness*, is that level of metabolic activity which is non-active at one concrete moment the cell has to display to be ready for immediate work at any moment (for the normal organism functioning it should be above 50 %). The third, *the level of the cell structure preservation*, is the precise level of energy required for the cell structure keeping. At least 15 % of the energy is required for the last one level of energetical activity. If less, the death of the cell occurs. Important that energetical deficits are the main reason for the organism death.

For all these levels of activity organism needs nutrients intake which enter organism during the food ingestion. Nutrients (mainly, carbohydrates, lipids and proteins) provide the energy and materials necessary for performing the

activities of daily living, for maintaining healthy skin, muscles and other tissues, for replacing and healing tissues and for the effective functioning of all body systems including the immune and respiratory systems. Therefore, the energy appears in the organism with the food after nutrients successful absorption from the small intestine. Its losses directed to the organism physical activity, energy for nutrients absorption (and deposits) and energy losses at a basal conditions (the so-called basal metabolic rate or base balance; Fig. 29–1).

Resuming, one should know nutrients should at least cover all the energy lost by the organism daily. Thus, we are speaking about the *nutrition* and it is the *adequate diet* that must cover all the body's energy requirements and provide a minimum of carbohydrates, proteins (including all essential amino acids) and fats (including essential fatty acids). Minerals (including trace elements), vitamins and sufficient quantities of water are also essential [For more information concerning nutrition aspects of the chapter see part 29.7]. To ensure a normal passage time, especially through the colon, the diet must also provide a sufficient amount of *roughage* (indigestible plant fibers — cellulose, lignin and others). Of the great importance is the process of the end-substances of the energy exchanges processes elimination from the organism — one should mention that *excretory organs* are responsible for this.

Being inside the body, nutrients are used for energy or as the building blocks for tissue growth and repair. When excess nutrients are available, they frequently are stored for future use. Some chemical reactions after food intake in an organism occur spontaneously, whereas others require energy input. For

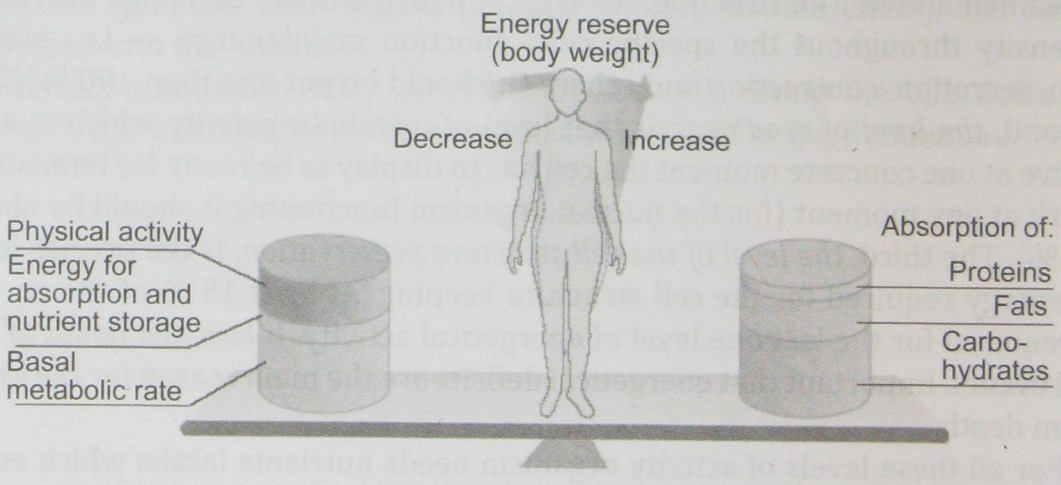


Fig. 29–1. The energy appearances and losses in the human organism.

example, some compounds are unstable and on keeping at room temperature they breakdown into simpler products. This breaking down involve rupture of some bonds in the compounds. As the rupture occurs, some energy, called "*free energy*", is liberated.

Some other reactions do not occur spontaneously, and for them to occur, energy from outside (mainly in the form of heat), which is again called "free energy", must be supplied.

Free energy, is that form of energy which can do work. For example, the energy input that is required for various chemical reactions, is free energy, as stated already. (Or, to lift a weight against gravity will also require an input of free energy, and so on).

Chemical reactions which release free energy (which can be utilized), are called *exergonic*, whereas they who require free energy input are called *endergonic* reactions. From the other side, chemical reactions which liberate heat are called *exothermic* reactions. These ones which need energy are called *endothermic*.

So, the mathematical sum of all the energetical processes occurred inside the organism (basal balance, the activity energy cost and the diet-induced thermogenesis) defines as *energetical balance* or *total metabolic rate*.

Energetical balance increases during the physical activity — up to 1.2-times in persons sitting quietly, up to 3.2-times in normal walking and till 8-times for forestry workers. Top athletes (e.g., in a marathon) can perform as much as 1600W (= J/s) for 2 hrs but their daily energy loss is much lower.

Energy drop also increases at various degrees of injury (1.6-fold for sepsis, 2.1-fold for burns). Each 1 °C of fever increases energetical balance by 13 %.

2. UNITS OF ENERGY

Heat is only a form of energy and free energy may or may not appear as heat. In fact, in our body, only a part of free energy appears as heat. However, standard practice in physics and biochemistry is to express the free energy in calories.

A *calorie* is the amount of heat or energy required to raise the temperature of 1 g of water by 1 °C. A *kilocalorie* (kcal) is the amount of energy needed to raise the temperature of 1 kg of water by 1 °C. Because a calorie is so small, kilocalories often are used in nutritional and physiologic studies.

The current day practice also uses joules [J; Joule is the unit of work] for the energy losses expression. Therefore, 1Kcal equals to 4.184 KJ.

3. BASAL BALANCE

Some cells and organs demonstrate a certain level of the exchange processes that differs from the of level of readiness. Exchange processes in heart, respiratory muscles, liver, kidneys, CNS etc. have to exceed the level of the cell structure preservation because these vitally important organs have to work even during the organism' absolute physical relaxation. The other organs (skeletal muscles, digestive tract etc.) might be in the level of readiness (ready to immediate work) during prolonged time without any harm for the organism. Correspondently, certain energetical exchange occurs during the so-called basal conditions which defines as the *basal balance* (synonyms are *basal metabolic rate, basal exchange*).

Basal conditions are the following: a) in the state of *muscular rest* — morning, after arousal (but not sleeping); b) complete mental and physical relaxation; c) *while fasting,* i. e. 12 to 14 hrs after the last food intake; d) at a *"comfortable" temperature* of 18 to 20 °C, i. e. at the temperature that neither causes the feeling of cold or shivering nor produces overheating of the body, and e) *free from all illness.*

Most of the basal metabolic energy is utilized for the work of skeletal muscles (40-50 %), the rest one goes to the gastrointestinal tract and liver (20–30 %), respiratory muscles (10–15 %), kidneys (4–6 %) and for CNS (2–5 %) activity.

Basal balance differs in different peoples because of their sex, height, weight, body compartments ration (i.e., bone, muscles and fats prevalence) as well as their rate of physical activity. Base balance values are expressed as Kcal (or KJ) per 24 hrs or additionally per 1 kg of the organism' weight. Rubner's idea was to recalculate the indexes of the energy production according to the total organism surface (square meters) but not pertaining to the weight units.

Hence, in adults base balance for healthy males equals approximately to 1300–1700 Kcal (7200 KJ) /24 hrs [about 1 Kcal (42 KJ)/kg/hr]. The females data are comparable but insignificantly (10 %) less. The same data recalculation to the total body surface (assuming this index equals to 1.8 m^2) gives adult healthy males base balance index is 40 Kcal (appr. 168 KJ)/m^2/hr and in healthy females — 37 Kcal (155 KJ)/m^2/hr.

Energy requirements are greater during growth periods. Infants require approximately 115 Kcal/kg at birth, 105 Kcal/kg at 1 year and 80 Kcal/kg of body weight between 1 to 10 years of age. During adolescence, boys require 45 Kcal/ kg of body weight and girls require 38 Kcal/kg of body weight. During pregnancy, a woman needs an extra 300 Kcal/day above her usual requirement and during the first 3 months of breast feeding she requires an additional 500 Kcal.

Basal balance has direct dependence with metabolism that is the organized process characterizes by nutrients biochemical degradation and energy release and conversion into cellular energy. Important to know how many units of energy releases after the nutrients oxidation: both carbohydrates and proteins give 4 Kcal/g, fats — two times more — 9 Kcal/g. The process of metabolism is unique in that it enables the continual release of energy, and it couples this energy with physiologic functioning. Because most of our energy sources come from the nutrients in the food that is eaten, the ability to store energy and control its release is important. Normally, energy utilization is balanced with energy expenditure (Fig. 29–1).

When a person is overfed and intake of food consistently exceeds energy expenditure, the excess energy is stored as fat and the person becomes overweight. Conversely, when food intake is less than energy expenditure, fat stores and other body tissues are broken down, and the person loses weight.

Energy losses are increased as the result of the skeletal muscles work. That's why energy exchange has direct dependence upon the level of the organism physical activity. There are 5 groups of the adult people's physical activity:

- The 1st group — persons working mentally.
- The 2nd group — persons with easy physical activity -operators of machines (turners, shop-assistants, medical nurses, milling machine operators, textile workers, public transport drivers, etc.).
- The 3rd group — persons with moderate physical activity (fitters and toolmakers, surgeons, stokers, agricultural workers).
- The 4th group — persons with hard physical activity (dockers and porters, navvies, wood-cutters, etc.).
- The 5th group — persons with extra-hard physical activity (miners, metallurgists, etc.)
- The energy values for different groups taking into consideration sex and age are given below (table 29–1).

Energy needs in persons with mental works (students, teachers — the 1st group) who prepared for diploma, exam or scientific thesis (dissertation) defense increased significantly and equals the same in persons of the 5th group.

Factors influencing basal balance values.

1. *Age.* In childhood it is high, after 60 years the basal balance tends to fall.
2. *Sex.* Throughout life, basal balance is somewhat lower in the females.

Table 29-1. The approximate expenditures of the adult persons according to their sex, age and physical work intensity

The groups of the work intensity	Age	Energy needs			
		Males		Females	
		Kcal	KJ	Kcal	KJ
1	18-29	2800	11750	2409	10042
	30-39	2700	11297	2300	9623
	40-59	2500	10669	2200	9205
2	18-29	3000	12552	1550	10669
	30-39	2900	12133	2450	10950
	40-59	2750	11506	2350	9832
3	18-29	3200	13388	2700	11296
	30-39	3100	12970	2600	10878
	40-59	2950	12342	2500	10460
4	18-29	3700	15480	3150	13179
	30-39	3600	15062	3050	12761
	40-59	3450	14434	2900	12133
5	18-29	4300	17991	—	
	30-39	4100	16154		
	40-59	3900	16317		

3. *Food*. If the food contains protein excess, the basal balance value increases. This is due to the high specific-dynamic action of the food protein.
4. *Time*. Starting from 8 a.m. basal balance index has a tendency to rise and vice versa after 8 p.m.
5. *Climate*. Colder the climate the basal balance is higher and vice versa.
6. *Pregnancy*. In later months of pregnancy, the basal balance rises because at this stage, the maternal basal balance includes that of the fetus which is considerable. In early stage of pregnancy also, there is some basal balance rise. This is perhaps due to human chorionic gonadotropin which might have a TSH-like action.

7. *Endocrine factors.* The basal balance is related to thyroid activity. In thyrotoxicosis it increases whereas in hypothiroidism it falls. Prior to the advent of modern methods like radio-immunoassay, thyroid disorders used to be diagnosed by basal balance estimation. A value of BMR over 40 % than normal (i.e., more than +40 %) or less than — 20 % (i.e., less than — 20 %) used to be usually regarded as diagnostic of thyrotoxicosis and mixedema, respectively.

4. THE SPECIFIC-DYNAMIC ACTION OF THE FOOD

The exchange processes intensification occurs till 10–12 hrs after food intake that is well known as the principle of the *specific-dynamic action of the food.* The energy goes on digestion itself (intestinal motor, secretory functions, absorption etc.) and on the biochemical processes intensification.

Maximal food specific-dynamic action — till +30 % pertaining the initial level — occurs after proteins prevalence inside the food. In case of the carbohydrates and lipids prevalence in the food this index doesn't exceed 15 %.

5. NITROGEN BALANCE

A physical state, in which nitrogen intake and output are equal, is known as *nitrogen equilibrium* or *nitrogen balance.*

$$N_{food} = N_{urine} + N_{persp} + N_{fec}$$

When the amount of nitrogen in food is so low that it corresponds to the minimum protein loss, the *nitrogen balance is negative.*

$$N_{food} < N_{urine} + N_{persp} + N_{fec}$$

It often occurs with fever, illness, infection, trauma, or burns, when more nitrogen is excreted than is consumed, and therefore, represents a state of tissue breakdown. *Thyroxine and adrenaline* can be responsible for it.

In certain physiological conditions (during growth, pregnancy or healing after surgery or injury), when the organism's requirements of protein are

completely satisfied, but they are consumed in excess, the *nitrogen balance becomes positive.*

$$N_{food} > N_{urine} + N_{persp} + N_{fec}$$

Somatotropic hormone, testosterone and progesterone can be responsible for positive nitrogen balance.

Retention of nitrogen occurs in growing organisms, owing to the increasing body mass, and during pregnancy owing to the growth of the fetus; during convalescence after a severe illness. It this physical state, nitrogen intake is more than output. Thus, positive nitrogen balance takes place in children, pregnant women, and others states when body mass increases.

Strenuous training in sports leading to an increase in the bulk of muscle tissue is also attended with nitrogen retention. It is known that *6,25 g of protein contains 1 gram of nitrogen.*

6. ISODYNAMIC LAW (RUBNER)

The amount of energy required by the organism can be supplied by the oxidation of proteins, fats or carbohydrates. Taking only their energetic value into account, Rubner formulated an *isodynamic law* states that *different nutrients can replace one another in accordance with their heat of combustion values.*

Rubner took into account only the energy requirement of the organism, but the substances contained in food, actually perform other functions, serving as building material et. al.. The body must receive adequate amounts of proteins, fats, carbohydrates, salts and vitamins.

7. ENERGY RICH COMPOUNDS

All energies inside the organism must be provided from the food. If no food is available the body has to utilize its own reserve food (glycogen of the liver and muscle, and triglycerides of the adipose tissue) as well as break down its own muscle for nitrogen supply — this condition will continue as long as possible and then death will occur.

Catabolism of food stuff, ultimately speaking, requires oxygen supply. Temporarily, even in absence of oxygen, partial catabolism and generation of little

amount of ATP (anaerobic oxidation) can occur but for long-term purposes, O_2 must be supplied from the atmosphere. Ultimately therefore, in a given subject, the greater the O_2 utilization, the greater is the intensity of metabolism. Determination of oxygen consumption per unit time therefore will determine the metabolic rate as the two parameters are interrelated.

This energy comes from ATP breakdown and other energy rich compounds. The ATP and related energy rich compounds are resynthesized by extracting energy which was potential in the food stuffs like carbohydrate, protein and fat, by catabolic processes.

It is the most fundamental feature of biochemistry that much of the free energy, which is liberated by catabolism of foodstuff in our body, is not wasted but trapped or stored in the form of ATP (adenosine triphosphate) and some other compounds (see below). These compounds are called energy rich compounds. ATP is the most studied and most important of all such energy rich compounds.

When, in our body, due to a chemical reaction, free energy is released, the free energy is, as already stated, trapped by ATP or some other energy rich compound. Therefore, ATP and other energy rich compounds are storage forms of energy. When input of energy, for some other biochemical reaction in our body will be required, the ATP will break down (hydrolyzed) into ADP (adenosine diphosphate) and Pi (inorganic phosphate), and energy will be released and available for doing the work of the biochemical reaction.

The following compounds are important energy rich compounds: a) ATP; b) CP (creatine phosphate, also called phosphocreatine); c) acyl-CoA (coenzyme A); d) UTP (uridine triphosphate); e) GTP (guanosine triphosphate) and f) CTP (cytosine triphosphate). ADP is also high energy compound; upon hydrolysis it yields AMP + Pi.

8. METHODS OF ENERGY EXCHANGE DETERMINATION

Different methods are known. These methods can be classified as direct and indirect.

Direct calorimetry. The aim of this method is to determine directly the amount of the heat produced by an individual throughout 24 hrs.

The subject stays inside a specially constructed by Athwater-Benedict chamber for some time. A tube encircles the room and water circulates through the pipe. The heat evolved by the subject escapes into the water of the pipe, so it is possible

to calculate the total heat evolved by the subject during the period of observation using (1, 2) the temperatures of the water inside the tubes, (3) the volume of the water circulated during the period of observation (Fig. 29–2).

The window (4) is used for the food intake and excrements elimination. The pump (5) is needed for the internal air circulation through the boxes filled by H_2SO_4 (6, 8) and special powder (7) for CO_2 extraction. Oxygen enters the chamber from balloon (10) supplied by manometer (11). The air pressure inside the chamber controls by the device with the rubber membrane (9).

Thus, knowing the period of observation and total body surface of the subject (in m^2), the total energy expenditure (or the metabolic rate) can be found out in the form of the heat liberated per hr/m^2 of body surface. The chamber is made up of such material that does not cause loss or gain of heat to or from the exterior. The heat given out via the subject's excreta is also added.

Experimental animals might be used for the direct calorimetry also (Fig. 29–3).

The testing animal is placed in a small chamber immersed in a known volume of ice. The amount of heat produced is equivalent to the amount of heat

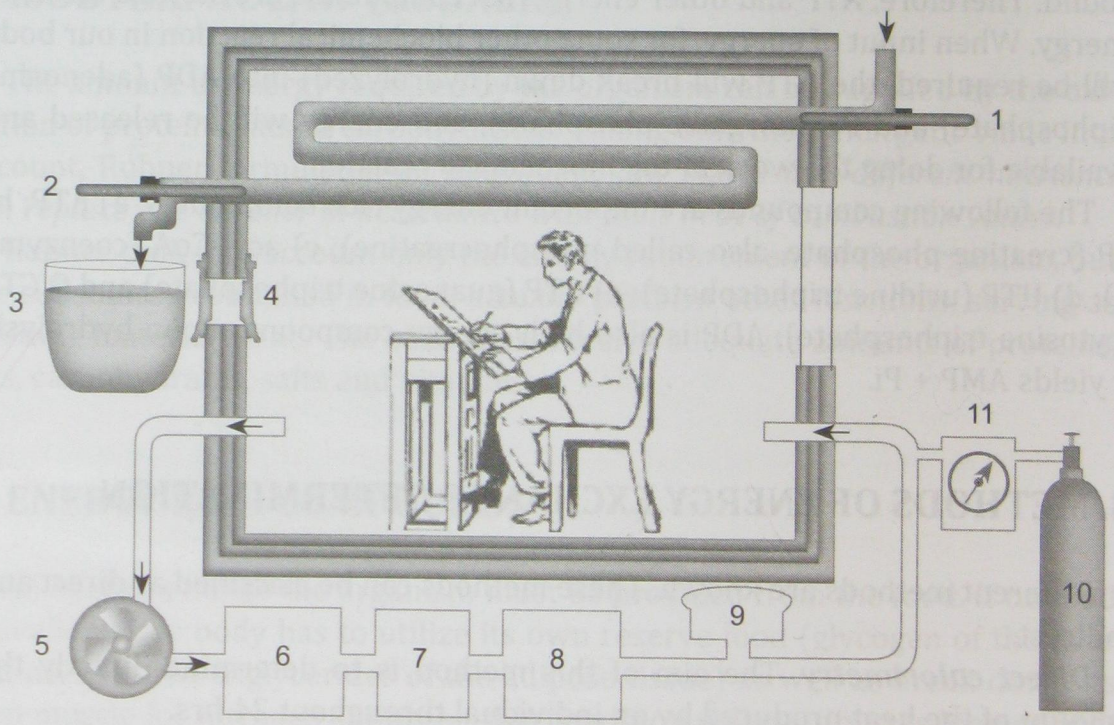


Fig. 29–2. The scheme of the Athwater-Benedict chamber.

absorbed by the surrounding water or ice. This is respectively calculated as the rise in water temperature or the amount of ice that melts.

This method is absolutely accurate but very expensive and cumbersome and unsuitable for routine use.

Indirect Calorimetry. In indirect calorimetry, the heat evolved is not directly measured. The special respiratory chambers might be used (the closed methods of the indirect calorimetry, Fig. 29–4) as well as non-chambers methods (the opened methods of the indirect calorimetry, Fig. 29–5) might be involved with this reason.

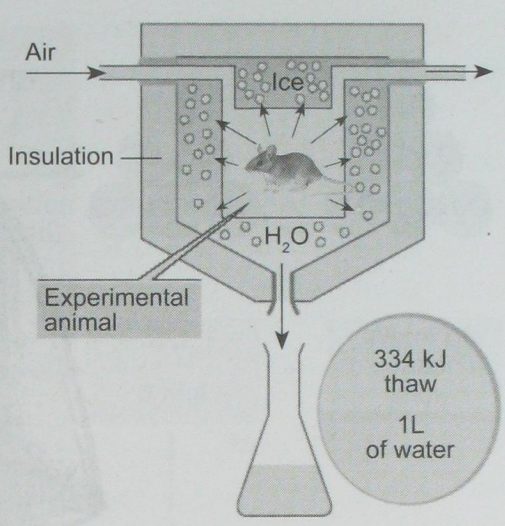


Fig. 29–3. The chamber of Lavoisier for the direct calorimetry

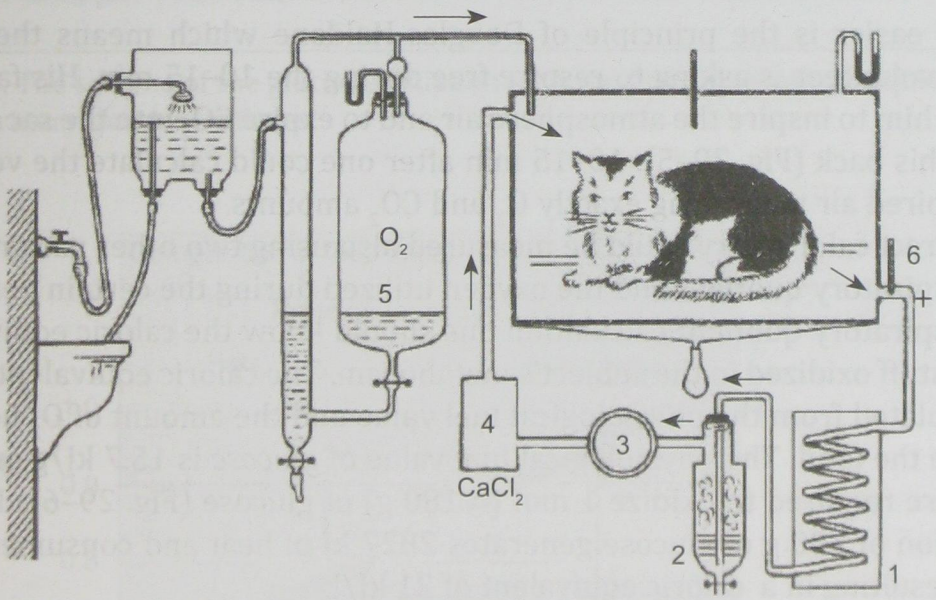


Fig. 29–4. The respiratory chamber arranged by Shaternikov.

The animal placed inside the closed chamber with the dosed O_2 access from the balloon (5) and CO_2 elimination through the pump (3). The heated air (measured with the help of the thermometers — 6) out the chamber through the refrigerator (1) consequently enters the device filled with base solution (2) and the same filled with $CaCl_2$ (4) for water air elimination.



Fig. 29–5. Pulmonary ventilation determination with the help of Douglas-Haldane sac.

The easier is the principle of Douglas-Haldane which means the following. A volunteer is asking to respire free during the 10–15 min. His face mask allows him to inspire the atmosphere air and to expire CO_2 into the sac of Douglas on his back (Fig. 29–5). 10–15 min after one could calculate the volume of the expired air measuring exactly O_2 and CO_2 amounts.

Indirect calorimetry could be measured also using two other parameters — the respiratory quotient and the oxygen utilized during the certain period. For the respiratory quotient calculation one should know the caloric equivalent of a foodstuff oxidized in the subject's metabolism. The caloric equivalent, in turn, is calculated from the physiological fuel value and the amount of O_2 needed to oxidize the food. The physiological fuel value of *glucose* is 15.7 kJ/g and 6 mol of O_2 are required to oxidize 1 mol (= 180 g) of glucose (Fig. 29–6). Thus, the oxidation of 180 g of glucose generates 2827 kJ of heat and consumes 134.4 l of O_2 resulting in a caloric equivalent of 21 kJ/l.

The mean caloric equivalent of the basic nutrients at 37 °C is 18.8 kJ/l of O_2 (carbohydrates), 17.6 kJ/l of O_2 (lipids) and 16.8 kJ/l of O_2 (proteins). The oxidized nutrients must be known in order to calculate the metabolic rate from the caloric equivalent.

The respiratory quotient (RQ) is a rough measure of the nutrients oxidized. For pure carbohydrates oxidized, RQ = 1.0. The RQ of lipids (fats) equals to 0.7 (Fig. 29–7).

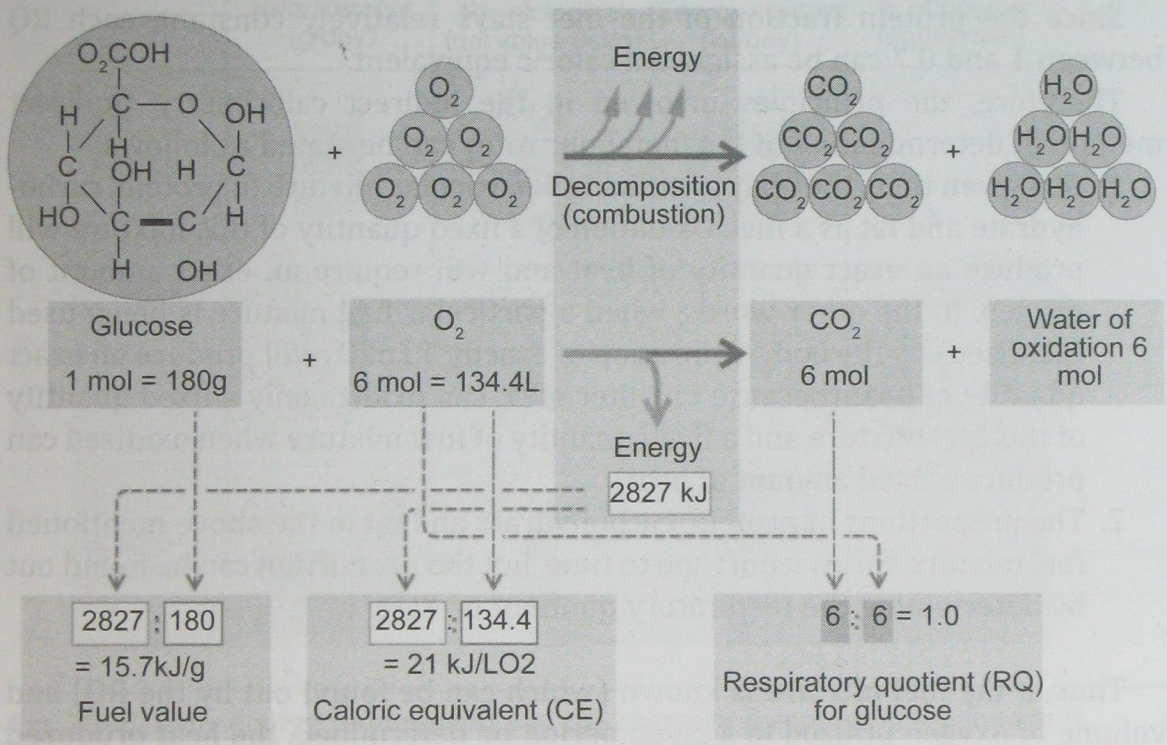


Fig. 29–6. The scheme of the glucose oxidation with the fuel value, caloric equivalent and respiratory quotient indexes.

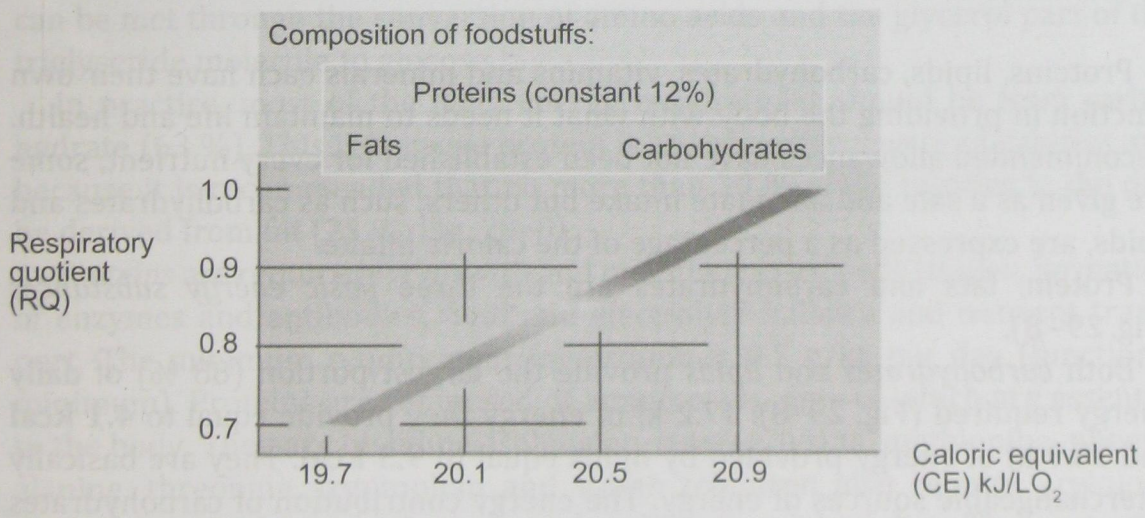


Fig. 29–7. Respiratory quotient and caloric equivalent of the foodstuff.

Since the protein fraction of the diet stays relatively constant, each RQ between 1 and 0.7 can be assigned a caloric equivalent.

Therefore, the principles involved in the indirect calorimetry (indirect method of determination of the metabolic rate) can be stated as follows:

1. At a given time, the body is utilizing a particular mixture of protein, carbohydrate and fat as a fuel. Oxidation of a fixed quantity of this mixture will produce an exact quantity of heat and will require an exact amount of oxygen. In the other words, when a particular fuel mixture is being used (oxidized) by the body, utilization of exactly 1 l of O_2 will produce an exact quantity of heat (because one liter of O_2 can oxidize only a fixed quantity of this fuel mixture and a fixed quantity of fuel mixture when oxidized can produce a fixed amount of heat).
2. The proportions of protein, carbohydrate and fat in the above mentioned fuel mixture varies from time to time, but the proportion can be found out by determining the respiratory quotient or RQ.

Thus, if the fuel mixture is known (which can be found out by the RQ) and volume of oxygen utilized in a given period be determined, the heat produced by the body in that given period can be found out. If now the surface area of the body is determined, the metabolic rate becomes known.

9. NUTRITION

Proteins, lipids, carbohydrates, vitamins and minerals each have their own function in providing the body with what it needs to maintain life and health. Recommended allowances have not been established for every nutrient; some are given as a safe and adequate intake but others, such as carbohydrates and lipids, are expressed as a percentage of the calorie intake.

Protein, fats and carbohydrates are the three *basic energy substances* (Fig. 29–8).

Both *carbohydrates* and *lipids* provide the largest portion (88 %) of daily energy required (Fig. 29–8). 17.2 kJ of energy they provide equal to **4.1 kcal** and 38.9 kJ of energy provided by lipids equal to **9.3 kcal**. They are basically interchangeable sources of energy. The energy contribution of carbohydrates can fall to about 10 % before metabolic disturbances occur. Dietary carbohydrates are composed of simple sugars, complex carbohydrates and undigested carbohydrates.

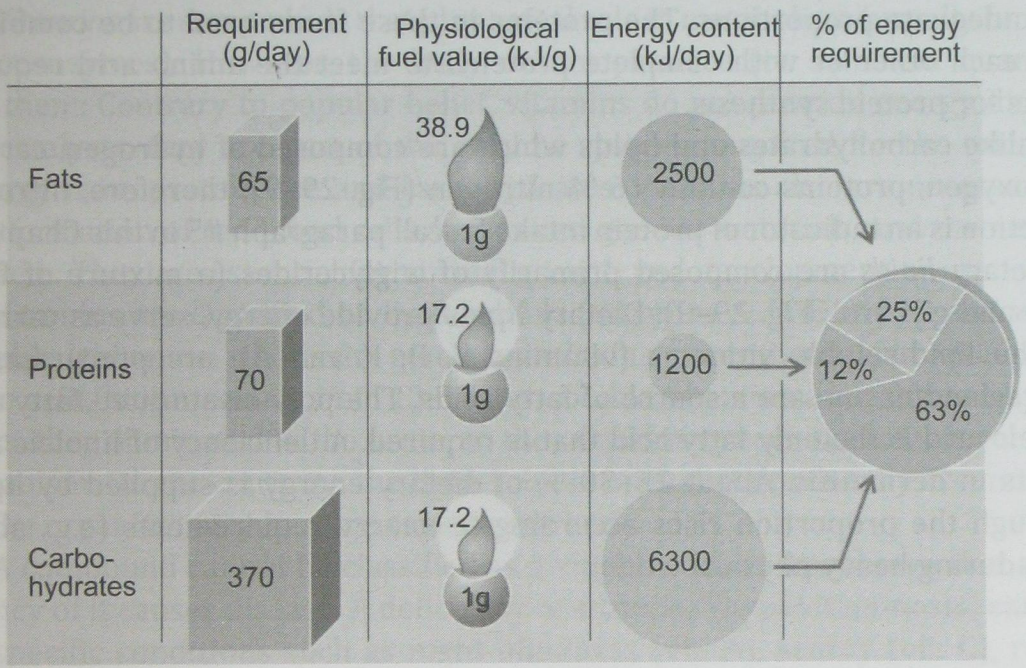


Fig. 29–8. Daily energy requirement and the energetical content of the foodstuff.

There is no specific dietary requirement for carbohydrates. All of the energy requirements can be met by dietary fats and proteins. Although some tissues, such as the nervous system, require glucose as an energy source, this need can be met through the conversion of amino acids and the glycerol part of the triglyceride molecule to glucose.

In practice, most of the daily energy requirement should be from carbohydrate (63 %). This is because protein is an expensive source of calories and because it is recommended that no more than 30 % of the calories in the diet be derived from fat (25 %, Fig. 28–8).

Proteins are required for growth and maintenance of body tissues, formation of enzymes and antibodies, fluid and electrolyte balance and nutrient transport. The minimum requirement for protein is 0.5 g/kg per day (functional minimum). Proteins are composed of amino acids, nine of which are essential to the body. They are histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine (children also require arginine). These aminoacids are in the meat, fish, milk, eggs and poultry. The content of most vegetable proteins is only about 50 % of animal protein. Dried peas and beans, nuts, seeds and grains contain all the essential amino acids but in less

than adequate proportions. The proteins in these foods need to be combined with each other or with complete proteins to meet the amino acid requirements for protein synthesis.

Unlike carbohydrates and lipids which are composed of hydrogen, carbon and oxygen, proteins contain 16 % nitrogen (Fig. 29–9); therefore, nitrogen excretion is an indicator of protein intake (recall paragraph #5 in this Chapter).

Dietary *lipids* are composed primarily of triglycerides (a mixture of fatty acids and glycerol; Fig. 29–9). Dietary lipids provide energy, serve as carriers for the lipid-soluble vitamins (vitamins E, D, K and A), are precursors of prostaglandins and are a source of fatty acids. The polyunsaturated fatty acid linoleic acid is the only fatty acid that is required. A deficiency of linoleic acid results in dermatitis. About 25–30 % of dietary energy is supplied by lipids although the proportion rises according to energy requirements (e.g., about 40 % during heavy physical work).

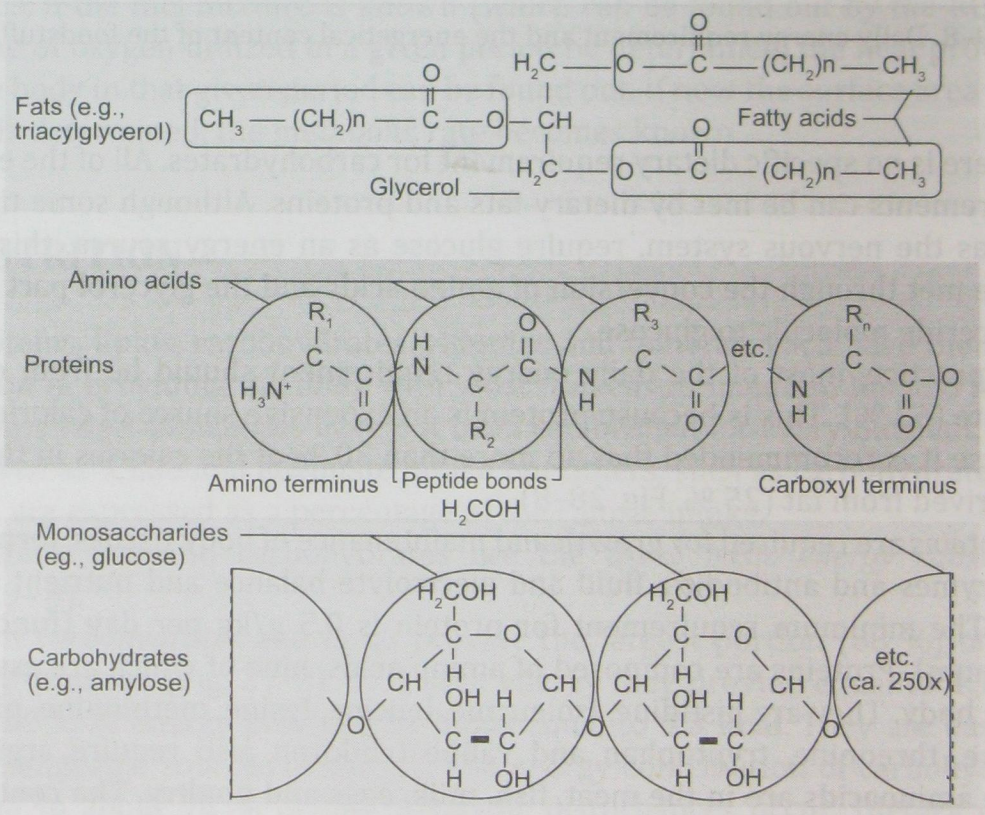


Fig. 29–9. Chemical structure of the lipids, proteins and carbohydrates.

Vitamins are compounds that play a vital role in metabolism (usually function as coenzymes). However, the body cannot produce (or sufficient quantities of) them. Contrary to popular belief, vitamins do not provide energy directly. As catalysts, they are part of the enzyme systems required for the release of energy from protein, fat, and carbohydrates. Vitamins also are necessary for the formation of red blood cells, hormones, genetic materials, and the nervous system. They are essential for normal growth and development.

Vitamins are classified on lipid-soluble (vitamins A, D, E and K) and water-soluble (thiamine, riboflavin, niacin, pyridoxine, pantothenic acid, vitamin B_{12}, folic acid, biotin and vitamin C). Water-soluble vitamins are excreted by kidneys, they eliminated in the urine, so they could not reach the toxic level in the organism. On the contrary, the lipid-soluble ones that are usually accumulated in the organism may become toxic to the body.

A compound cannot be classified as a vitamin unless it is shown that a deficiency of it causes disease. A deficiency of vitamins (hypovitaminosis) can lead to specific conditions such as night blindness (vit. A), scurvy (vit. C), rickets (vit. D), anemia (vit. B_{12}, folic acid) and coagulation disorders (vit. K). An excessive intake of certain vitamins like vitamin A and D, on the other hand, can be toxic *(hypervitaminosis)*.

An adequate intake of *minerals* (inorganic compounds), especially calcium, phosphorus, sodium, clorine, potassium etc. is essential for proper body function. These minerals are present in relatively large amounts in the body — that's why they are called *macrominerals*. They are involved in acid-base balance and in the maintenance of osmotic pressure in body compartments. Minerals are components of vitamins, hormones and enzymes. They maintain normal hemoglobin levels, play a role in nervous system function and are involved in muscle contraction and skeletal development and maintenance.

The rest of the minerals, especially iron (10–20 mg/day), iodine (0.15 mg/ day), As, F, Cu and others are also essential but because of the presence in the relatively small concentrations they define as *trace elements.* The normal diet provides sufficient quantities of them but excessive intake has toxic effects.

10. DIETARY STANDARDS FOR MAN

The *daily allowance of protein* that will satisfy the requirements of a human adult engaged in easy work under normal physiological conditions is average between 80 and 100 g. It must be the *protein optimum.*

The diet should include 70–80 g of *fats*.

The *carbohydrate* content of the daily diet should be between 400 and 500 g. The diet must also contain carbohydrates, mineral salts, and an adequate quantity of vitamins.

If person takes a meal *3 times a day*, breakfast should contain 30 % of kilocalories, dinner (lunch) — 50 % and supper 20 %.

Assimilability of food averages 95 % with animal food, 80 % with vegetable food, and 82–90 % with a mixed diet. In practice 90 % is usually adopted as the standard for calculations.

11. REGULATION OF ENERGY EXCHANGE PROCESSES AND NUTRITION

Stability of the weight and body composition over time requires that energy intake matches energy utilization. One person should eat the kinds and amounts of foodstuffs that provide adequate amount of nutrients including vitamins and minerals and calories. It should be mentioned that approximately 27 % of the energy ingested normally reaches the cells, and much of this energy is ultimately converted to heat which is generated as a result of muscle activity, protein metabolism and various organs and tissues of the body activities. Excess of energy intake is stored mainly as fat, whereas it deficit causes loss of total body mass until energy expenditure eventually equals energy intake or death occurs. Important that multiple factors — environmental, cultural, genetic and psychological — all influence food intake and energy expenditure. However, powerful physiological control systems are well-known to regulate hunger and food intake.

The sensation of *hunger* is associated with several sensory perceptions, such as the rhythmic contractions of the stomach and that "empty feeling" in the stomach that stimulates a person to seek food. A person's *appetite* is the desire for a particular type of food. It is useful in helping the person determine the type of food that is eaten. Satiety is the feeling of fullness or decreased desire for food.

The hypothalamus contains centers for hunger and satiety. The ventromedial nuclei of the hypothalamus serve as the *satiety center*. This center is believed to give a sense of nutritional satisfaction that inhibits the feeding center. Electrical stimulation of this region can cause complete satiety, and even in the presence of highly appetizing food the animal refuses to eat *(aphagia)*. Ven-

tromedial part of the hypothalamus destruction (injury) results in the appetite increasing — *hyperphagia* and obesity in experimental animals.

Lateral hypothalamus destruction induces the opposite process — *hypophagia* and weight loss, muscle weakness and decreased metabolism. This part of the hypothalamus is considered to be a *feeding center* which electrical stimulation in an experimental animals induce hyperphagia.

The paraventricular, dorsomedial and arcuate nuclei of the hypothalamus also play a major role in regulating food intake. For example, lesions of the paraventricular nuclei often cause excessive eating, whereas lesions of the dorsomedial nuclei usually depress eating behavior. Hence, the arcuate nuclei are established as the sites in the hypothalamus where multiple hormones released from the gastrointestinal tract and adipose tissue converge to regulate food intake as well as energy expenditure.

The hypothalamus receives neural input from the gastrointestinal tract that provides information about stomach filling, chemical signals from the blood about the nutrients in food and input from the cerebral cortex regarding the smell, sight and taste of the food (Fig. 29–10). It happens because stretch receptors in the stomach activate sensory afferent path-ways in the vagus nerve and inhibit food intake. Peptide YY (PYY), cholecystokinin (CCK) and insulin are those gastrointestinal hormones that are released by the ingestion of food and suppress further feeding. Ghrelin is released by the stomach, especially during fasting, and stimulates appetite. Leptin is a hormone produced in increasing amounts by fat cells as they increase in size; it inhibits food intake.

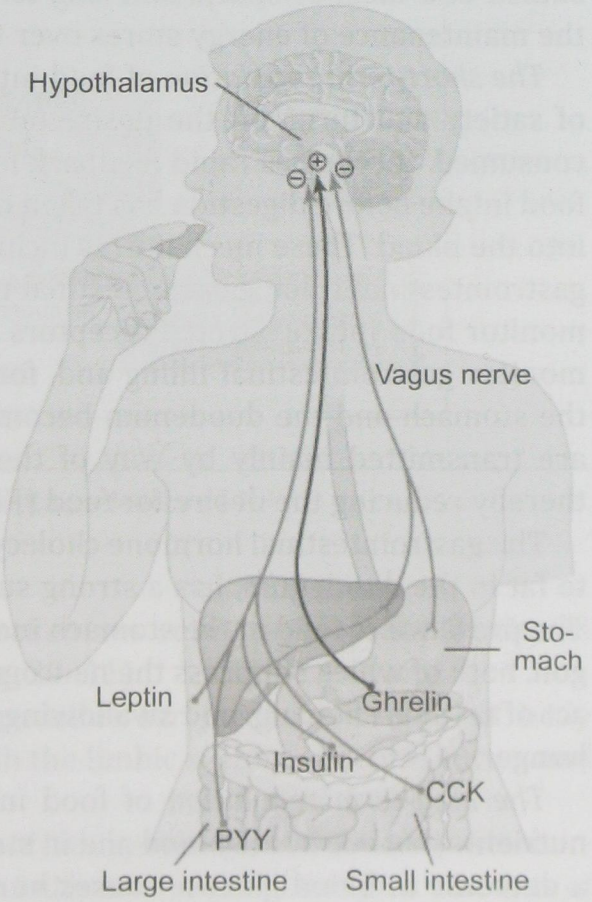


Fig. 29–10. Feedback mechanisms for food intake control (explanation see in the text).

The next question aroused — is hypothalamus the only one center responsible for food intake regulation? Experiments gave an answer — if the brain is sectioned below the hypothalamus but above the mesencephalon, the animal can still perform the basic mechanical features of the feeding process. It can salivate, lick its lips, chew food and swallow. Therefore, the actual mechanics of feeding are controlled by *centers in the brainstem.* The function of the other centers in feeding, then, is to control the quantity of food intake and to excite these centers of feeding mechanics to activity.

Neural centers located higher than the hypothalamus (f. i., amygdala and prefrontal cortex which are closely placed with the hypothalamus) also play important roles in the control of feeding, particularly in the control of appetite.

Therefore, being complex, *the control of food intake* can be divided into *short-term regulation*, which is concerned with the amount of food that is consumed at a meal or snack, and *long-term regulation* which is concerned with the maintenance of energy stores over time.

The short-term regulation of food intake provides a person with the feeling of satiety and turns off the desire for eating when adequate food has been consumed. It requires rapid feedback mechanisms that signal the adequacy of food intake before digestion has taken place and nutrients have been absorbed into the blood. These mechanisms include receptors that monitor filling of the gastrointestinal tract, gastrointestinal tract hormones and oral receptors that monitor food intake. Stretch receptors in the gastrointestinal tract constantly monitor gastrointestinal filling and, for instance, in case of overeating, when the stomach and the duodenum become distended, stretch inhibitory signals are transmitted mainly by way of the vagi to suppress the feeding center, thereby reducing the desire for food (Fig. 29–10).

The gastrointestinal hormone cholecystokinin, which is released in response to fat in the duodenum, has a strong suppressant effect on the feeding center. The presence of food in the stomach increases the release of insulin and glucagon, both of which suppress the neurogenic feeding signals from the brain. The act of tasting, chewing, and swallowing also appears to suppress the feeling of hunger.

The long-term regulation of food intake is determined by the amount of nutrients that are in the blood and in storage sites. It has long been known that a decrease in blood glucose causes hunger. In contrast, an increase in breakdown products of lipids such as ketoacids produces a decrease in appetite. The ketogenic weight-loss diet relies on the appetite suppressant effects of ketones in the blood.

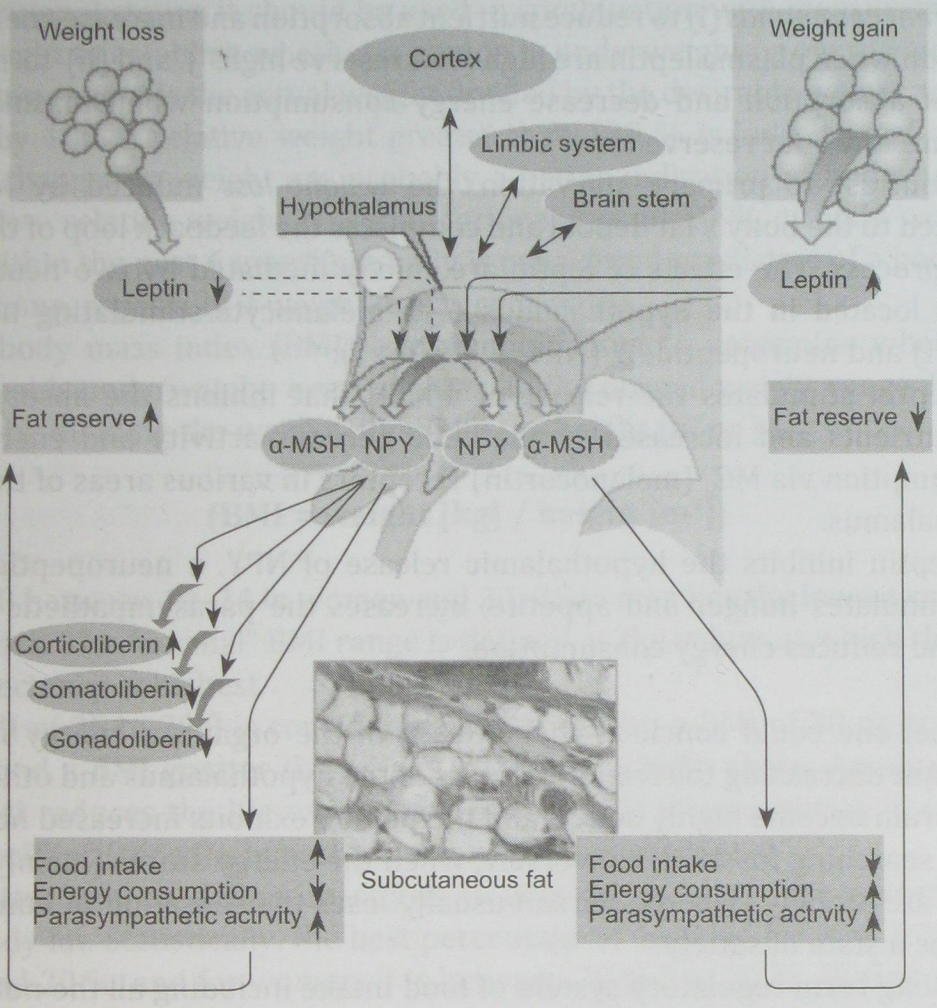


Fig. 29–11. Regulation of body weight by leptin, a-MSH and NPY.

Briefly, the following *regulatory mechanisms* serve to keep the fat depots and thus the body weight constant (Fig. 29–11).

The hypothalamus, the center responsible for feedback control of body weight, keeps neuronal projections with the limbic system, cerebral cortex and brainstem. It sends and receives:

Afferent messages concerning the size of fat depots. *Leptin*, a 16-kDa proteohormone produced by fat cells, is the main indicator for this. Increased amounts of leptin are released from the adipocytes when fat stores are increased. The stimulation of leptin receptors in the hypothalamus produces a decrease in appetite and food intake as well as an increase in metabolic rate and energy consumption.

Efferent commands (i) to reduce nutrient absorption and increase energy consumption when plasma leptin are high ("fat reserve high!"), and (ii) to increase nutrient absorption and decrease energy consumption when plasma leptin levels are low ("fat reserve low!").

In contrast to primary starvation, the *weight loss* induced by leptin is restricted to the body's fat depots and completes the feedback loop of the regulatory process. The effects of leptin are chiefly mediated by two neurotransmitters located in the hypothalamus — α-melanocyte-stimulating hormone (α- MSH) and neuropeptide Y (NPY; Fig. 29–11):

1. Leptin stimulates the release of a-MSH that inhibits the absorption of nutrients and increases sympathetic nervous activity and energy consumption via MC_4 (melanocortin) receptors in various areas of the hypothalamus.
2. Leptin inhibits the hypothalamic release of NPY, a neuropeptides that stimulates hunger and appetite, increases the parasympathetic activity and reduces energy consumption.

Hence, one could conclude that in case of the organism energy reserves significant decreasing the feeding centers of the hypothalamus and other areas of the brain become highly active, and the person exhibits increased hunger as well as searching for food; conversely, when the energy stores (mainly the fat stores) are already rich, the person usually loses the sensation of hunger and develops a state of satiety.

The long-term regulatory system of food intake including all the nutritional energy feedback mechanisms helps to preserve constant nutrients supplies in the tissues preventing them from becoming too low or too high. The short-term regulatory stimuli serve two other purposes. First, they tend to make the person eat smaller quantities at each eating session, thus allowing food to pass through the gastrointestinal tract at a steadier pace, so that its digestive and absorptive mechanisms can work at optimal rates rather than becoming periodically overburdened. Second, they help in preventing the person from eating amounts at each meal that would be too much for the metabolic storage systems once all the food has been absorbed.

Anthropometric measurements provide a means for assessing body composition, particularly fat stores and skeletal muscle mass. This is done by measuring height, weight, body circumferences etc. These measurements commonly are used to determine growth patterns in children and appropriateness of current weight in adults. Body weight is the most frequently used method of assess-

ing nutritional status; it should be used in combination with measurements of body height to establish whether a person is underweight or overweight.

Relative weight is the actual weight divided by the desirable weight and multiplied by 100. A relative weight greater than 120 % is indicative of obesity. Recent changes in weight are probably a better indication of undernutrition than a low relative weight. An unintentional loss of 10 % of body weight or more within the past 6 months usually is considered predictive of a poor clinical outcome, especially if weight loss is continuing.

The body mass index (BMI) is commonly used to determine whether an individual is underweight, overweight or in the normal weight range. It is calculated by dividing the weight (in kg) by the height (in m^2)

(BMI = weight [kg] / height [m^2])

A BMI between 19–24 in women and 20–25 in men has the lowest statistical health risk. The "normal" BMI range is defined as the values at which the mean life expectancy is highest.

A BMI of 25 to 29.9 is considered as overweight; a BMI of 30 or greater as obese; and a BMI greater than 40 as very or morbidly obese. An abnormally high BMI reduces the life expectancy since this is often associated with diabetes mellitus (type II), hypertension and cardiac disease.

A better way to define obesity is to actually measure the percentage of total body fat. Statistically, the best percentage of body fat for men is between 12 % and 20 %, and for women, it is between 20 % and 30 %. During physical training, body fat usually decreases and lean body mass increases.

Chapter 30. Temperature and its regulation

1. THE PRINCIPLE OF POIKILOTHERMY AND HOMOYOTHERMY

The body temperature of humans remains relatively constant despite changes in the environmental temperature. Such constancy defines as *homoyothermy* applies only to the *core temperature* (appr. 37 ± 5 °C) of the body. Within this range, there are individual differences and diurnal variations (Fig. 30–1). Important to note that *core temperature* reaches its highest point in late afternoon and evening and lowest point in the early morning hours (Fig. 30–2).

The opposite situation means that some cold-blood animals (f. i., frog, reptiles and others) live with the real *poikilothermy* that means that body temperature closely depends upon the surrounding temperature. Hence, human extremities and skin ("shell") exhibit *poikilothermy*, i.e., their temperature varies to some extent with environmental temperature.

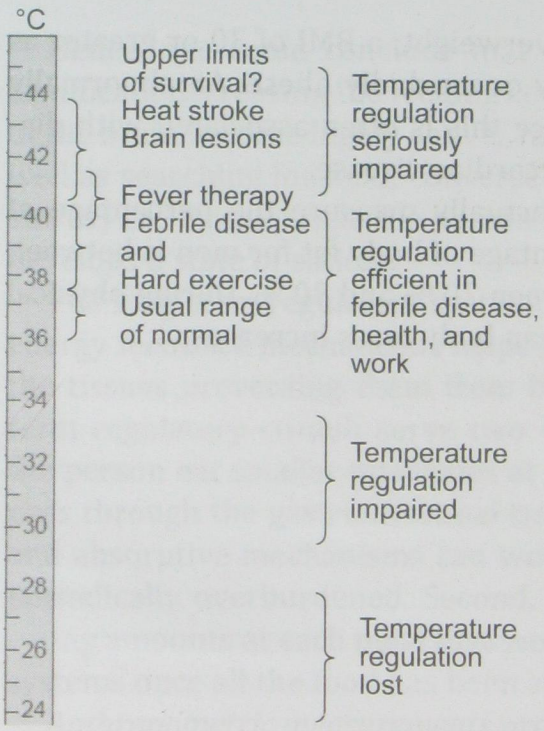


Fig. 30–1. Body temperatures in normal regulatory conditions in compensatory and non-compensatory states.

To maintain a constant core temperature, the body must balance the *amount of heat it produces and absorbs* with the *amount it loses*; this equation defines as *thermoregulation*.

Body temperature reflects the difference between heat production and heat loss. Therefore, human organism can be divided into two parts: *organs of heat production — core that* includes body and visceral organs and *organs of heat loss — shell,* includes skin and extremities. The temperature of the core remains almost exactly constant (±0,5 ºC) in

state of rest. The shell temperature, in contrast to the core temperature, rises and falls with the temperature of the surrounding.

The body temperature of humans is commonly estimated by measuring it in the axilla, where it is 36,5 °C to 36,9 °C in a healthy individual.

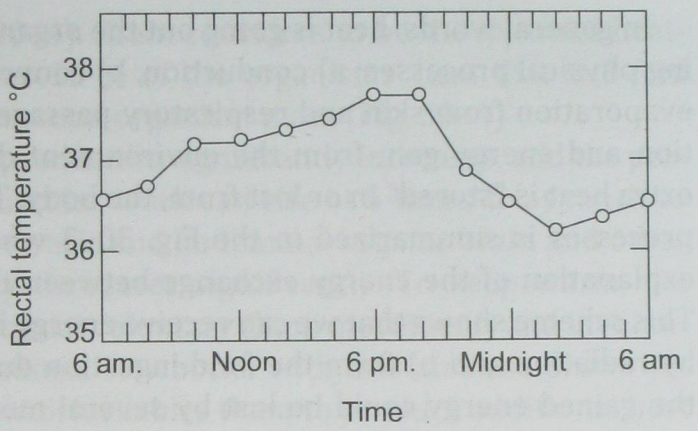


Fig. 30–2. Diurnal variations of the body temperature.

In clinical practice, particularly in pediatrics, rectal temperature is often measured; it is higher than the axillary and averages between 37,2 °C and 37,5 °C in a healthy subject, in the mouth temperature is 37,0–37,1 °C.

Thermoregulation is vitally important for the living mammals at least for the following reasons: a) the proteins structure and function could be failed in both extremely high and low temperature; b) all biochemical processes in the body are affected by changes in temperature; c) nucleinic acids structure and function and d) lipids physical-chemical condition and function are of the great dependence on the organisms' temperature.

2. THE MECHANISMS OF THERMOREGULATION

Heat is produced in all tissues of the body but is lost to the environment only from tissues in contact with the environment — predominantly from the skin and, to a lesser degree, from the respiratory tract. We'll discuss it in details later but now the important items are the following.

All the biochemical processes inside the organisms that use energy for their manifestation finish by heat production. There are two possible further routes of this heat produced inside the organism: the first one — such heat is the adverse reaction of the organisms' living processes and, the second, this heat represents the main mechanism of heat transfer. Important to remember the physical law concerning the direct dependence between the temperature and biological processes rate.

In general words, heat is going out the organism with the help of the following physical processes: a) conduction, b) convection, c) radiation and d) water evaporation from skin and respiratory passages. If the sum of energy production and energy gain from the environment does not equal energy loss, the extra heat is "stored" in or lost from, the body. This relationship between these processes is summarized in the Fig. 30–3 where you can see the schematic explanation of the energy exchange between the organism and environment. This scheme shows that we can receive energy by two main routes: a) from sun by radiation and b) from the food ingestion during the meal. Total amount of the gained energy could be lost by several mechanisms we mentioned above (conduction, convection, radiation and evaporation). Additionally, humans heat losses might be possible by conduction to the ground through their feet and by its conversion into mechanical work during the muscular work, e. g., when a person is walking uphill.

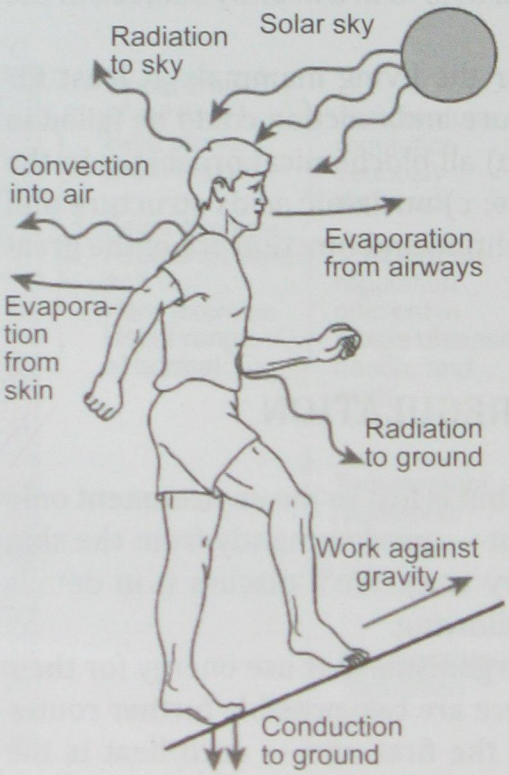


Fig. 30–3. The schematical explanation of energy transfer between the organism and environment.

3. HEAT PRODUCTION

For the *normal thermal homeostasis* total amount of heat produced should be equal to the total amount of the heat lost using different mechanisms. The amount of heat produced is determined by *energy metabolism*. It is discussed in details in the previous lecture.

Interestingly that *at rest* maximal amount of total heat (appr. 56 % vs 100 % totally) produces in the deeper core tissues (i.e., muscles and viscera) which are insulated from the environment and protected against heat loss by the subcutaneous tissues and about 18 % in the muscles and skin (Fig. 30–4, fragment 2). Important that total heat producing efficacy of internal organs vs their mass is 5-fold higher in the thoracic and abdominal organs — i.e. 1.65 times comparing with the same in the

muscles and skin — 0.32 (Fig. 30–4). Brain considers to be the maximal heat-producing organ because it gives 16 % of the organisms' heat that is 8-fold greater pertaining its mass vs the total organism mass (Fig. 30–4).

During *physical exercise* heat production significantly increases, and the percentage of heat produced by muscular work can rise at 5 times to as much as 90 % (Fig. 30–4, fragm. 2, b). Thoracic and abdominal organs, brain and other organs heat production efficacy decreases significantly. To keep warm, the body may have to generate additional voluntary (limb movement) and involuntary (shivering) muscle contractions. Newborns with this aim use *brown fat* tissue biochemical degradation which enables them to produce additional heat without shivering.

Heat produced in the body is absorbed by the bloodstream and conveyed to the body surface. To retain the *internal flow of heat*, the temperature of the body surface must be lower than that of the body interior. The *blood supply to the skin* is the chief determinant of heat transport to the skin.

Heat loss occurs during the well-known for you physical processes — radiation, conduction, convection and evaporation.

Radiation (Fig. 30–5). Radiation involves the transfer of heat through the air or a vacuum. The amount of heat lost by radiation from the skin is mainly

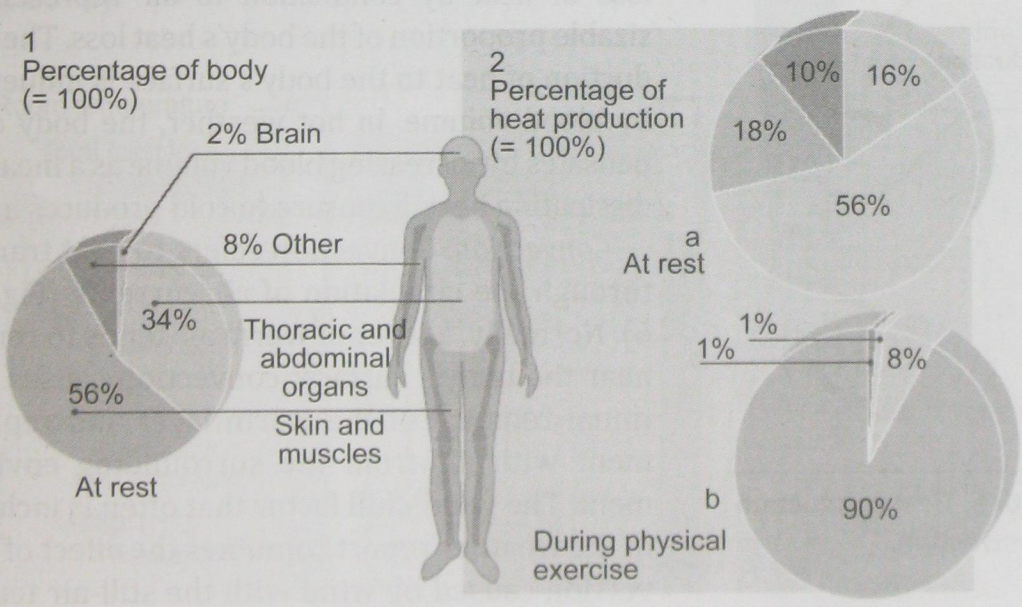


Fig. 30–4. The comparative contribution of organs heat production efficacy to their mass inside the organism at rest and during the physical exercise.

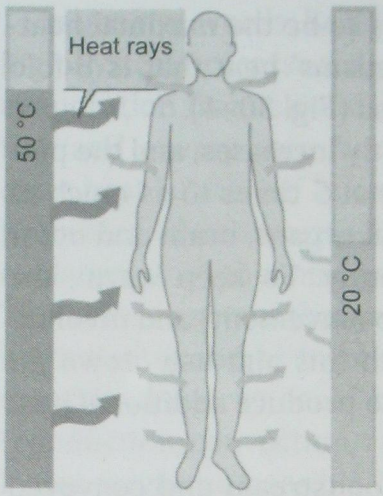


Fig. 30–5. Radiation.

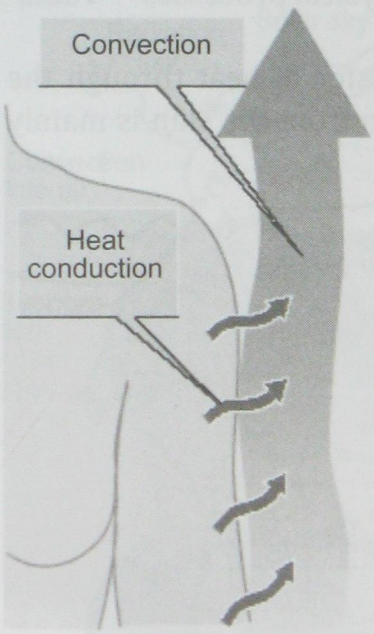


Fig. 30–6. Heat conduction and convection.

determined by the temperature of the environment (the 4th power of its absolute temperature).

Environmental temperature must be less than that of the body for heat loss to occur by radiation. Additionally, heat returns to the body from objects (sun) that are warmer than the skin. Heat radiates from the body into the environment when no radiating object is present (night sky). Heat radiation does not require the aid of any vehicle and is hardly affected by the air temperature. Therefore, the body loses heat to a cold wall (despite warm air in between) and absorbs radiation from the sun or an infrared radiator without air or cold air, respectively, in between.

Conduction. Conduction involves the direct transfer of heat from one molecule to another. Blood carries or conducts heat from the inner core of the body to the skin surface (Fig. 30–6). Normally, only a small amount of body heat is lost through conduction to a cooler surface. However, loss of heat by conduction to air represents a sizable proportion of the body's heat loss. The conduction of heat to the body's surface is influenced by blood volume. In hot weather, the body compensates by increasing blood volume as a means of dissipating heat. Exposure to cold produces a cold

Convection. Convection refers to heat transfer through the circulation of air currents (Fig. 30–6). Normally, a layer of warm air tends to remain near the body's surface; convection causes continual removal of the warm layer and replacement with air from the surrounding environment. The wind-chill factor that often is included in the weather report combines the effect of convection caused by wind with the still-air temperature.

Resuming, the last two physical processes involve the heat transfer from the skin to cooler air or a cooler object (e.g. sitting on rock) in contact with the

body (*conduction*). The amount of heat lost by conduction to air increases greatly when the warmed air moves away from the body by natural *convection* (heated air rises) or forced convection (wind).

Evaporation. The above mentioned mechanisms alone are unable to maintain adequate temperature homeostasis at high environmental temperatures or during hard (and prolonged) physical activity. Evaporation is the means by which the body manages with the additional heat. The water lost by evaporation reaches the skin surface by diffusion (*insensible perspiration*) and by neuron-activated *sweat glands* (Fig. 30–7). About 0.58 calories of heat are lost for each gram of water evaporating and thereby cooling the skin.

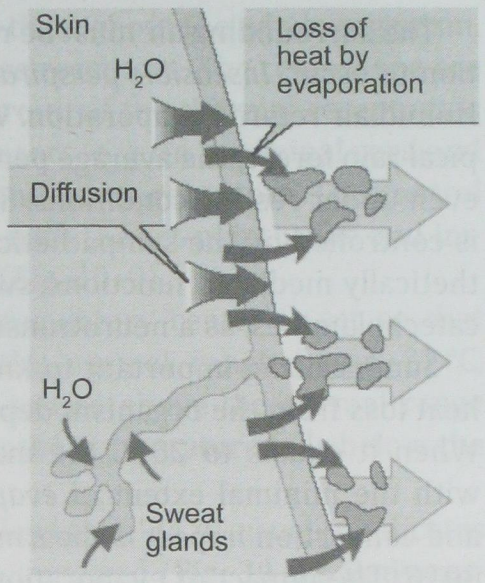


Fig. 30–7. Evaporation.

At temperatures above or equal to 36°C heat loss occurs by evaporation only (Fig. 30–8).

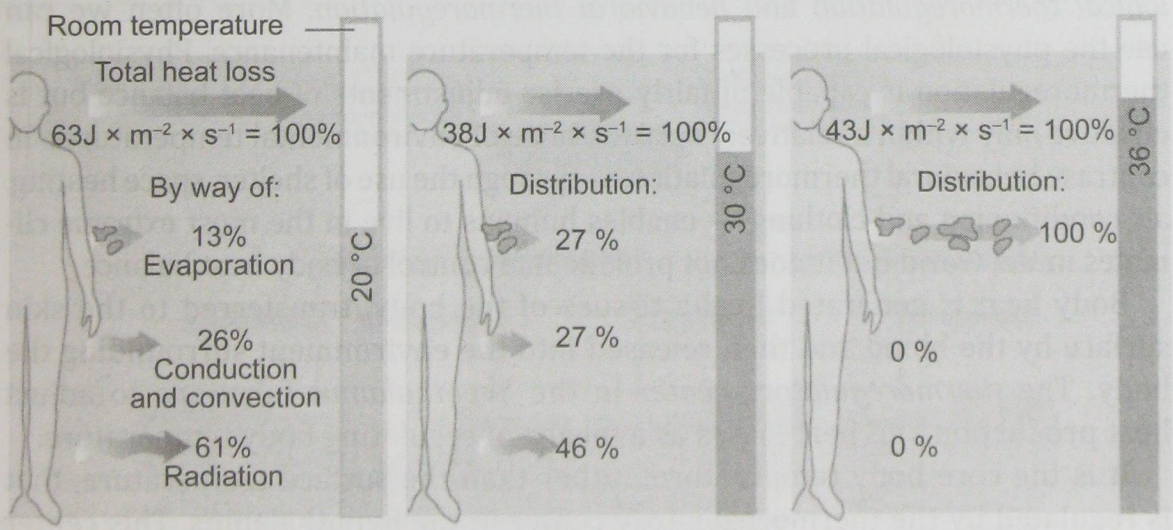


Fig. 30–8. The scheme of unclosed human heat production in case of different temperature.

The surrounding air must be relatively dry in order for heat loss by evaporation to occur. *Insensible perspiration* losses are greatest in a dry environment. Humid air retards evaporation. When the air is extremely humid (e.g., in a tropical rain forest) the average person cannot tolerate temperatures above 33 °C, even under resting conditions. Sweating occurs through the sweat glands and is controlled by the sympathetic nervous system. In contrast to other sympathetically mediated functions, sweating relies on acetylcholine, rather that the catecholamines, as a neurotransmitter.

Summarizing important to know that the mechanism that is the leading in heat loss from the organism depends only by the environmental temperature. When it equals to 20 °C the maximal amount of heat releases by *radiation* with the minimal extent of *evaporation*. Radiation together with conduction and convection impact is the same in case of room temperature equal to appr. 30 °C. The impact of evaporation increases with the environmental temperature rising with the practical absence of radiation and conduction/convection when temperature reaches 36 °C (Fig. 30–8). Any condition that prevents evaporative heat losses causes the body temperature to rise.

4. THE SYSTEM OF THERMOREGULATION

Humans have two distinct subsystems for regulating body temperature: *physiological thermoregulation* and *behavioral thermoregulation*. More often we can use the physiological processes for the temperature maintenance. Physiological thermoregulation is capable of fairly precise adjustments of heat balance but is effective only within a relatively narrow range of environmental temperatures. In contrast, behavioral thermoregulation — through the use of shelter, space heating, air conditioning and clothing — enables humans to live in the most extreme climates in the world but it does not provide fine control of body heat balance.

Body heat is generated in the tissues of the body, transferred to the skin surface by the blood and then released into the environment surrounding the body. The *thermoregulatory center* in the *hypothalamus* functions to adjust heat production and heat losses as a means of regulating body temperature.

It is the core body temperature, rather than the surface temperature, that is regulated by the thermoregulatory center in the hypothalamus. This center integrates input from cold and warm thermal receptors located throughout the body and generates output responses that conserve body heat or increase its dissipation.

The thermal receptors are diffusely located throughout the whole organism. One could classify the two wide groups of these receptors. The *central* one are located inside the hypothalamus. The other group of the thermal receptors is the involve the *peripherally* located thermosensors, e. g., in the spinal cord and skin. Those in the skin are divided into *cold* and *warm* receptors. The control center compares the actual core temperature with the set-point value and initiates measures to counteract any deviations (Fig. 30–9).

The *thermostatic set point* of the thermoregulatory center is set so that the core temperature is regulated within the normal range (appr. 37 °C, Fig. 30–10). When body temperature begins to rise above the normal range heat-dissipating behaviors are initiated; when the temperature falls below the normal range, heat production is increased.

A core temperature greater than 41°C or less than 34 °C usually indicates that the body's ability to thermoregulation is impaired (Fig. 30–1). Body responses that produce, conserve and dissipate heat are described in Table 30–1.

Core temperature increasing often could be normalized using the vessels diameter rising. Mainly the vessels of the skin participated in these processes thus human body at warm temperature has red colour.

Moreover, arterio-venous anastomoses (AV shunts) open in the periphery, especially in the fingers. The blood volume transported per unit time then not only conveys more heat but also reduces the countercurrent exchange of heat between the arteries and their accompanying veins (Fig. 30–11).

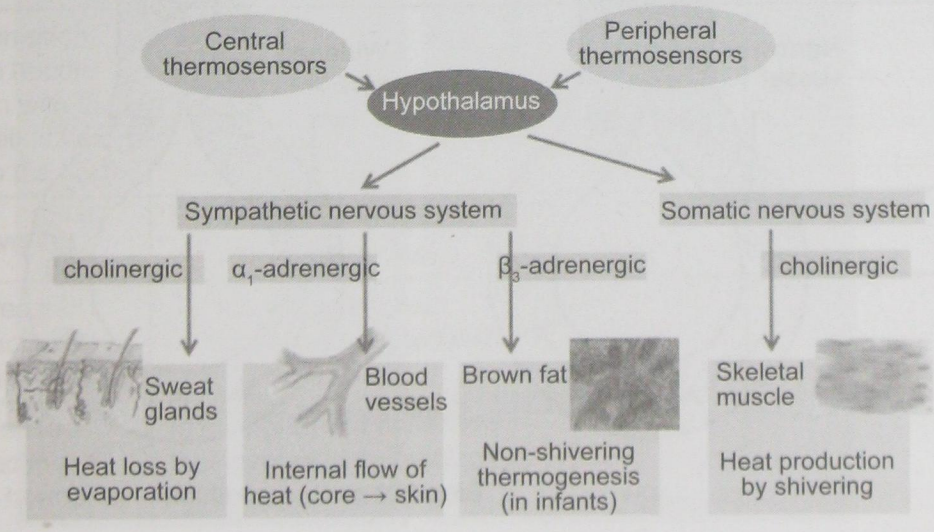


Fig. 30–9. The temperature zones of the body.

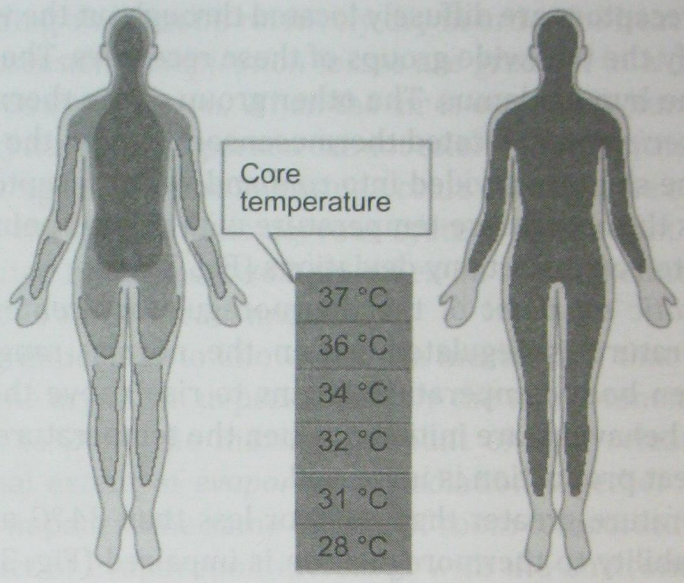


Fig. 30–10. Vessels participation in the temperature regulation.

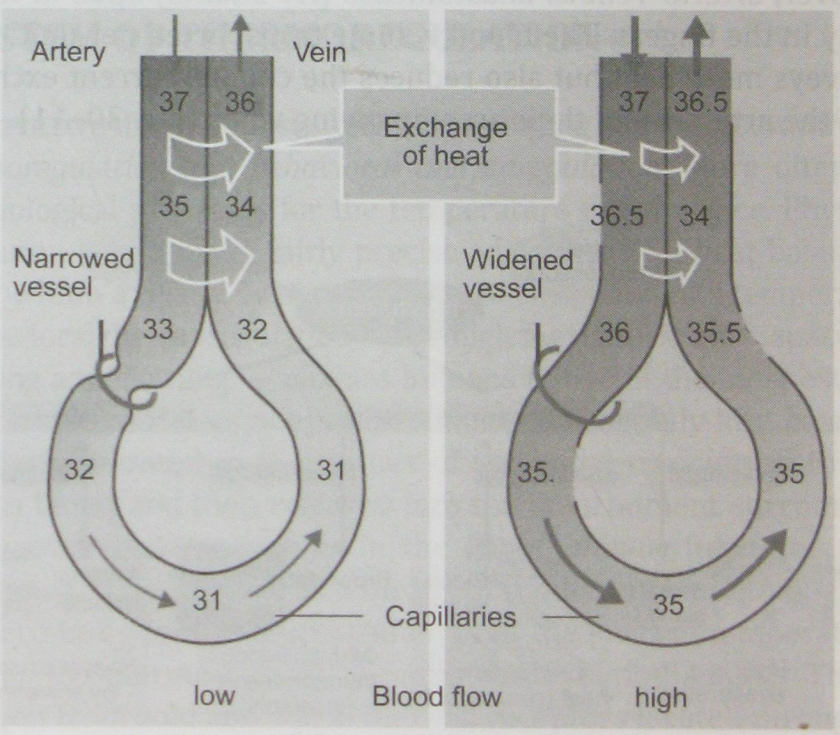


Fig. 30–11. Vessels participation in the temperature regulation.

These AV shunts are much like the radiators in a heating system. When the shunts are open, body heat is freely dissipated to the skin and surrounding environment; when the shunts are closed, heat is retained in the body. The blood flow in the AV shunts is controlled almost exclusively by the sympathetic nervous system in response to changes in core temperature and environmental temperature.

Table 30–1. Heat production and heat loss responses used in body temperature regulation

Heat production		Heat loss	
Body response	**Mechanism of action**	**Body response**	**Mechanism of action**
Vasoconstriction of the superficial blood vessels	Confines blood flow to the inner core of the body, with the skin and subcutaneous tissues acting as insulation to prevent loss of core heat	Dilatation of the superficial blood vessels	Delivers blood containing core heat to the periphery where it is dissipated through radiation, conduction, and convection
Contraction of the pilomotor muscles that surround the hairs on the skin	Reduces the heat loss surface of the skin	Sweating	Increases heat loss through evaporation
Assumption of the huddle position with the extremities held close to the body	Reduces the area for heat loss		
Shivering	Increases heat production by the muscles		
Increased production of epinephrine	Increases the heat production associated with metabolism		
Increased production of thyroid hormone	Is a long-term mechanism that increases metabolism and heat production		

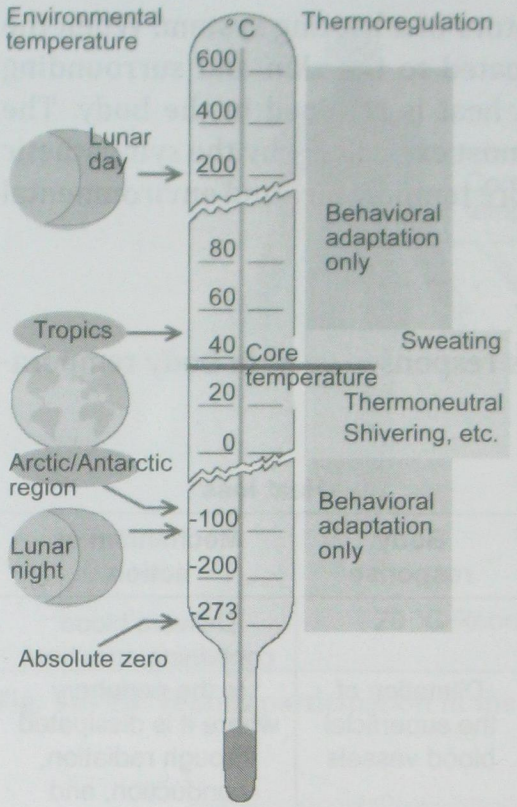


Fig. 30–12. The ratio between the body's comfortable temperature and ambient temperature.

Central warm sensors emit the signals that activate the sweat glands. (In this case, the thermosensors of the skin do not detect warmth because their environment is cooler than the core temperature). The efferent nerve fibers to the sweat glands are cholinergic fibers of the sympathetic nervous system (Fig. 30–9).

The range of ambient temperatures between the sweating and shivering thresholds is known as the *thermoneutral zone*. It lies between 27 °C and 32 °C in the nearly unclothed test subject. The only thermoregulatory measure necessary within this range is variation of blood flow to the skin. The narrow range of this zone shows the thermoregulatory importance of behavior — one has to understand the importance of the *behavioural methods of thermoregulation* (see above). To remember, it involves choosing the appropriate clothing, seeking shade, heating or cooling our dwellings, etc. Behavioral adaptation is the chief factor in survival at extreme ambient temperatures (Fig. 30–12).

The *thermoneutral zone* is subjectively perceived as the *comfort zone*. 95 % of all subjects wearing normal office attire and performing normal office activities perceive an indoor climate with the following conditions to be comfortable: ambient and radiant (wall) temperature appr. 23 °C, wind velocity less 0.1 m/s and relative humidity appr. 50 %. A resting, unclothed subject feels comfortable at the air at about 28 °C and appr. 31 °C to 36 °C in water depending on the thickness of subcutaneous fat (heat isolator).

To finalize, important to know (remembering about the vessels role in thermoregulation) that spinal cord injuries that transect the cord at Th_6 or above can seriously impair temperature regulation because the hypothalamus can no longer control skin blood flow or sweating.

5. HYPOTHALAMIC CONTROL OF BODY TEMPERATURE CHANGES

Temperature receptors of the skin and of the hypothalamus (which "read" the core temperature) relay information by *tractus spinothalamicus* to the anterior hypothalamus.

The anterior hypothalamus compares the core temperature to the set-point temperature.

Heat-loss mechanisms — response to heat. The response is orchestrated by the *anterior hypothalamus.* Increases in temperature cause a decrease in sympathetic tone to skin blood vessels, increasing blood flow through arterioles and increasing arterio-venous shunting of blood to venous plexus near the surface of the skin. Shunting of warm blood to the surface of the skin *increases heat loss by radiation and convection.* Vasodilatation also increases due to acetylcholine-sympathetic nerves influence (see Fig. 30–11).

Heat loss by *evaporation* increases substantially. It depends upon the activity of sweet glands, which are under sympathetic muscarinic control.

Decrease in heat production will take place as a result of posterior hypothalamic centres inhibition. Such mechanisms as shivering and chemical thermoregulation that cause excess heat production are strongly inhibited.

Heat exhaustion is caused by excessive sweating, which causes decreased blood volume and as a result, decreased arterial blood pressure and syncope.

Heat stroke occurs when body temperature increases to the point of tissue damage. The normal response to increased ambient temperature (sweating) is impaired, and core temperature will rise further.

Temperature-increasing mechanisms –response to cold. At cold temperature thermoregulatory reactions are quite opposite.

Increasing of posterior hypothalamus tone causes vasoconstriction of skin blood vessels due to sympathetic nerves (see Fig. 30–11). Heat production will increase as a result of sympathetic nervous system excitation and endocrine mechanisms. Adrenaline and noradrenaline produce a rapid but short-lived increase in heat production. Thyroid hormones produce a slowly developing but prolonged increase. As a result of sympathetic activation pilomotor reaction of skin hair takes place.

A source of considerable heat, particularly in infants, is breaking down of brown fat.

Shivering. When body temperature falls below a critical, primary motor center of shivering of posterior hypothalamus is activated. It then transmits signals that cause shivering down to the brain stem, spinal cord and finally, to the anterior motor neurons. These signals increase the tone of the skeletal muscles throughout the body. When the tone rises above a certain critical level, shivering begins. This probably results from feedback oscillations of the muscle spindle-stretch reflex mechanism. During maximum shivering, body heat production can rise to four to five times normal.

Unit XI PHYSIOLOGY OF KIDNEYS AND ACID-BASE BALANCE

Chapter 31. Structural organization of excreting system. Kidneys and their role in urine formation

Excretory processes are the final link in metabolism within the organism; their immediate result is elimination of breakdown products. The excretory organs in man are the kidneys, sweat glands, lungs, and intestine.

The lungs are excretory organs because carbon dioxide, water, and certain volatile substances, e.g., the ether and chloroform vapours of anesthetics and alcohol vapour from intoxication are eliminated through them. The intestine excretes salts of heavy metals.

1. KIDNEYS FUNCTIONS

The kidney is a regulatory organ essential to maintaining the homeostasis of body fluids. This is accomplished by continuous monitoring and regulation of both the composition and volume of all body fluids. Maintaining homeostasis — wide definition that means regulating the balance of solutes, electrolytes, pH, volume and pressure of the body fluids. The regulatory functions of the kidney involve *monitoring* and *modifying* the composition and volume of body fluids resulting in the selective excretion of excess fluid volume, solutes, electrolytes, acid, toxins etc (which are the composites of the urine). To accomplish this, the kidneys balance precisely the intake, production, excretion and consumption of many organic and inorganic compounds. This balancing requires that the kidneys maintain the volume and composition of body fluids by the conservation and excretion of water and solutes to meet bodily requirements.

In general, kidneys function are wide and numerous. For this we'll list below their main functions with the brief explanation. Hence, kidneys are responsible for:

- *water and electrolyte intake and metabolism*;
- *excretion of water and solutes.* Important to stress that kidneys excrete appr. 1.5 l of hypertonic urine in 24 hrs. They eliminate the waste products of metabolism including urea (the main nitrogen-containing end-product of protein metabolism in humans), uric acid (an end-product of purine metabolism) and creatinine (an end-product of muscle metabolism). The kidneys remove many drugs (e.g., penicillin) and foreign or toxic compounds;
- *the volume of the extracellular fluid regulation* by controlling Na^+ and water excretion;
- *the osmotic pressure (osmolality) of the body fluids regulation* by daily excreting of osmotically dilute or concentrated urine;
- *blood plasma ions concentration regulation* including Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, bicarbonate, phosphate and sulfate;
- *the formation and release of renin* which is a major component of the renin-angiotensin-aldosterone mechanism allows the kidneys *to regulate blood pressure*;
- *the formation and release of renal erythropoietic factor* (erythropoietin) increases the number of the circulating erythrocytes;
- *vitamin D activation*;
- *gluconeogenesis.* The kidneys acquire the important ability to synthesize and secrete glucose produced from noncarbohydrate sources (glutamine) only in unusual circumstances such as prolonged starvation, chronic respiratory acidosis etc.;
- *acid-base balance regulation* by excreting H^+ when there is acid excess or HCO_3^- in case of base excess. Additionally it should be mentioned that the kidneys synthesize ammonia, which plays a role in acid-base balance;

Kidneys are the place of several polypeptide hormones including insulin, glucagon and parathyroid hormone biochemical degradation. They synthesize substances that affect renal blood flow and Na^+ excretion including arachidonic acid derivatives (prostaglandins, thromboxane A_2) and kallikrein (a proteolytic enzyme that results in the production of kinins).

The main function of kidneys, however, is the *formation of urine.*

Such a tremendous variety of the kidneys functions shows their vital importance for the human organism and explains the wide spectrum of pathological events starting in case of the kidneys failure.

3. THE PECULIARITIES OF THE RENAL BLOOD SUPPLY

Each kidney is supplied by a single renal artery that arises on either side of the aorta. The kidneys receive appr. 20-25 % of the total resting cardiac output or about 1.20–1.25 l blood/min. This is required to maintain a high glomerular filtration rate.

Over a wide range of renal arterial pressures (i. E., 80-180 mm Hg) the renal blood flow and glomerular filtration rate remain quite constant. This intrinsic phenomenon observed in the renal capillaries also occurs in the capillaries of muscles and is defined as *autoregulation*. Due to *autoregulation of renal blood flow* (the phenomenon of Baylis) only slight changes in renal plasma flow and glomerular filtration rate occur (even in a denervated kidney) when the systemic blood pressure fluctuates between 80 and about 180 mm Hg (Fig. 31–1). Therefore, for autoregulation to occur, the resistance to blood flow through the kidneys must be varied in direct proportion to the arterial pressure.

Additionally to autoregulation phenomenon kidneys blood flow could be regulated by the neuronal and humoral mechanisms. Briefly, the kidney is richly innervated by the sympathetic nervous system which increased activity causes constriction of both afferent and efferent arterioles and thus a decrease in renal blood flow (the details see later).

Several humoral substances, including angiotensin II, antidiuretic hormone and endothelins, produce vasoconstriction of renal vessels. The endothelins are a group of peptides released from damaged endothelial cells in the kidney and other tissues responsible for renal blood flow reduction.

Other substances such as dopamine, nitric oxide (NO) and

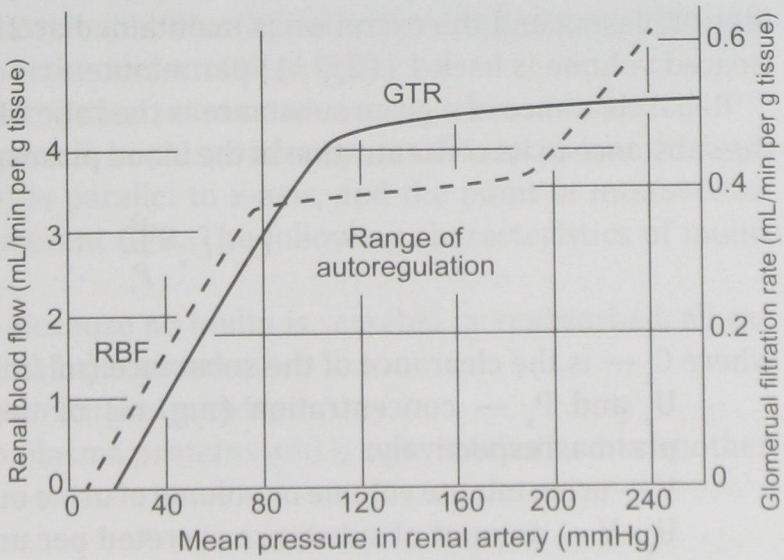


Fig. 31–1. The phenomenon of renal blood flow (RBF) and glomerular filtration rate (GFR) autoregulation.

prostaglandins produce vasodilation. NO, a vasodilator produced by the vascular endothelium, appears to be important in preventing excessive vasoconstriction of renal blood vessels and allowing normal excretion of Na^+ and H_2O. Prostaglandins are a group of mediators of cell function that are produced locally and exert their effects locally. Although prostaglandins do not appear to be of major importance in regulating renal blood flow and GFR under normal conditions, they may protect the kidneys against the vasoconstricting effects of sympathetic stimulation and angiotensin II.

4. RENAL CLEARANCE

Clearance is a measure of the volume of plasma completely freed of a given substance per minute by the kidneys. It is the *efficiency* with which the plasma is cleared of a given substance.

Clearance concept is very important for the clinical practice that why some examples are reasonable. For instance, if the concentration of the interested substance is 4 molecules per liter and its excretion is 4 molecules per minute, the volume of plasma cleared of interested substance is 1 l (4/4=1) per minute.

If the excretion of interested substance decreases to 2 molecules per minute, the volume cleared of this substance equals to 0.5 l (2/4=0.5) per minute.

If the concentration of interested substance decreases to 2 molecules per liter of plasma and the excretion is maintained at 2 molecules per minute, the cleared volume is back 1 l (2/2=1) per minute.

Renal clearance of a given substance is the ratio of the renal excretion rate of the substance to its concentration in the blood plasma. This ratio is expressed as:

$$C_x = U_x \cdot \frac{\dot{V}}{P_x}$$

where C_x — is the clearance of the substance (ml/min);

U_x and P_x — concentration (mg/ml) of the substance in urine and plasma, respectively;

V — urine minute volume or volume of urine output per minute (ml/min);

$U_x \cdot V$ — amount of substance excreted per unit time (mg/min) .

Clearances of several substances plotted against their plasma concentrations are given on the Fig. 31–2. Units of plasma concen-trations vary over a

wide range so that values on the abscissa are relative only. Note that the ordinates is interrupted and para-amino-hippuric acid (PAH) clearance at low plasma concentration is greater than 600 ml/min.

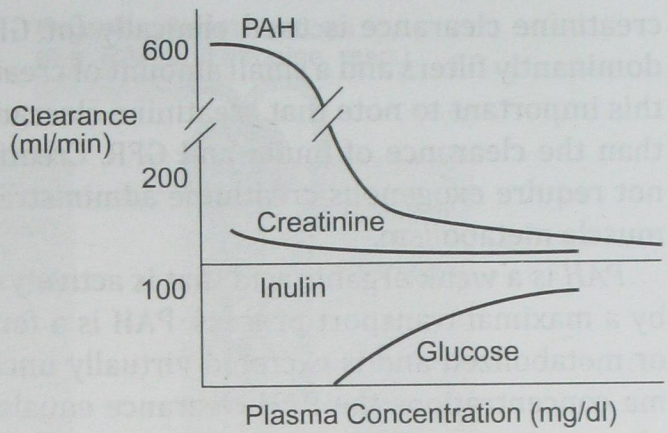


Fig. 31–2. Clearances some characteristic substances.

Explaining this, important to stress that as the *glucose* plasma levels rises it will appear in the urine once some nephrons' carriers are saturated. At low glucose plasma levels its clearance equals to zero. The plasma level at which glucose firstly appear in the urine is called the plasma (or renal) threshold and at this point glucose will have a positive clearance. As the plasma levels rises further the clearance will increase and approach to that of inulin. But the clearance will never equal the same of inulin because some glucose will always be reabsorbed. In case of glucose reabsorption complete inhibition the clearance of glucose would be equal to glomerular filtration rate (GFR).

Inulin, a fructo-polysaccharide that does not occur naturally in the body, can be used in test for renal function determining. Inulin produces a line parallel to x-axis (Fig. 31–2) and intersection point on the y-axis represents GRF. If GFR increases the line shifts upward, correspondently, if GFR decreases this line shifts down. In any cases it is parallel to x-axis, and the point of intersection with y-axis will always represent GFR. The following characteristics of inulin account for this quality:

- inulin is only filtered. Because no inulin is secreted or reabsorbed, all excreted inulin must come from plasma;
- inulin is biologically inert and nontoxic;
- inulin is not bound to plasma proteins and is not metabolized to another substance.

Although inulin clearance can be used to measure GFR, other substances can measure GFR. The most commonly used agent is *creatinine* additionally with mannitol, sorbitol, sucrose and radioiodine-labeled Hypaque. Endogenous

creatinine clearance is used clinically for GFR determination because it predominantly filters and a small amount of creatinine is secreted in humans. With this important to note that creatinine clearance will always be slightly greater than the clearance of inulin and GFR. Creatinine clearance determination do not require exogenous creatinine administration as creatinine is a product of muscle metabolism.

PAH is a weak organic acid that is actively secreted into the proximal tubule by a maximal transport process. PAH is a foreign substance that is not stored or metabolized and is excreted virtually unchanged in the urine. At low plasma concentrations the PAH clearance equals renal plasma flow (RPF). As the plasma concentrations rises some PAH will appear in the renal venous plasma and the clearance will be less than renal plasma flow. As the plasma levels rises further the clearance approaches but never equals GFR because some PAH is always secreted. In case of PAH secretion complete suppression the clearance would equal GFR. However, the clearance of PAH is referred to as effective renal plasma flow because some plasma peruses the renal capsule. This flow (appr. 10 %) is not cleared of PAH. Thus, PAH clearance is only 90 % of the true renal plasma flow.

Clearance concept could be used also to tubular function determination. Especially, if inulin is used for the GFR determination.

1. If clearance of concrete substance (C_x) equals to inulin clearance (C_{in}) it indicates that the amount of the substance excreted per unit time equals the amount of the substance filtered in the same time. Thus, clearance ratio of 1 indicates that the substance is neither reabsorbed nor secreted and, therefore, is *only filtered*. Substances with the clearance ratios close to 1 include mannitol, sorbitol, vitamin B_{12} and sucrose
2. If clearance of C_x less than C_{in} it indicates that clearance ratio less than 1. This demonstrates that the amount of the substance excreted per unit time is less than the amount filtered per unit time. Thus, the substance following filtration undergoes *reabsorption* (Fig. 31–3, 1). Substances with the clearance ratios less than 1 include glucose, xylose and fructose.
3. If clearance of C_x exceeds than C_{in} it indicates that clearance ratio greater than 1. This demonstrates that the amount of the substance excreted per unit time is greater than the amount filtered per unit time. Therefore, the substance is cleared by filtration and *secretion* (Fig. 31–3, 2). Substances with the clearance ratios greater than 1 include PAH, phenol red, certain penicillins, iodide-containing roentgen-contrasting solutions.

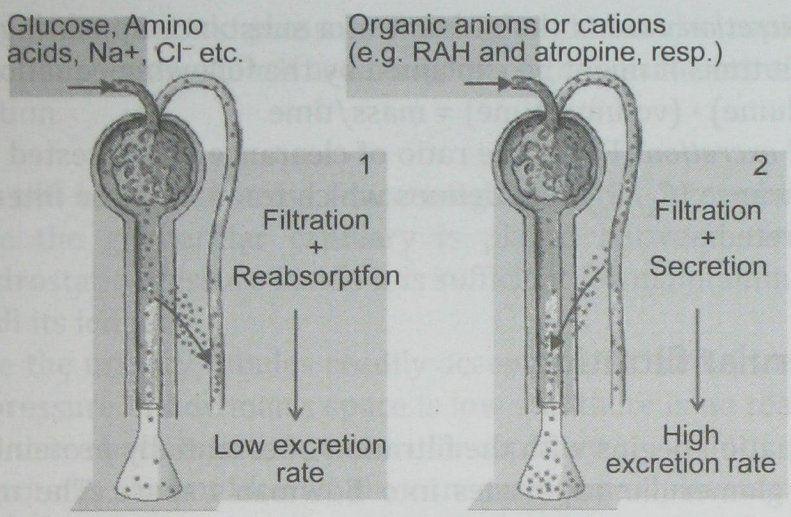


Fig. 31–3. Clearance indexes less (left side) and greater (right side) than inulin clearance.

5. THE MECHANISMS OF THE URINE FORMATION

Urine formation involves three processes: a) the *filtration of blood* by the glomerulus to form primary urine, b) the *tubular reabsorption* of electrolytes and nutrients needed to maintain the constancy of the internal environment and c) waste materials eliminating (*tubular secretion*).

Firstly, we have to introduce to you some important definitions concerning the glomerular filtration processes.

Renal blood flow (RBF) — amount of blood entering the kidneys per unit time (volume/time), equal to 1200 ml/min.

Renal plasma flow (RPF) — amount of plasma entering kidneys per unit time (volume/time). It is equal to the renal blood flow minus the red cell volume of the blood hematocrit (~45 %). Thus, RPF = RBF — (1-Hct) = 660 ml/min.

Glomerular filtration rate (GFR) is the amount of plasma being filtered by the glomeruli per unit time (volume/time). In other words, it's the amount of glomerular urine generated by unit time — equal to 120 ml/min.

Creatinine clearance — amount of plasma that is cleared of creatinine by the formation of urine per unit time (volume/time). For several reasons detailed above the clearance of creatinine is very close to the GFR.

Filtration fraction — the fraction of the renal plasma flow that gets filtered. About 1/5 of the plasma entering the kidneys becomes glomerular urine.

Urinary excretion rate is the amount of a substance appearing in the final urine per unit time. It might be explained by the following equation

(mass/volume) · (volume/time) = mass/time.

Fractional excretion (FE) is the ratio of clearance of interested substance X to inulin clearance (C_X/C_{In}) and defines which fraction of the filtered quantity of X was excreted.

5.1. Glomerular filtration

Urine formation begins with the filtration of essentially protein-free plasma through the glomerular capillaries into Bowman's space. The movement of fluid through the glomerular capillaries is determined by the same factors (i.e., capillary filtration pressure, colloidal osmotic pressure and capillary permeability) that affect fluid movement through other capillaries in the body.

The filtration barrier consists of three layers: endothelial cells of the capillaries with perforations about 0,1 micron in diameter; thin basement membrane, which consists of the net of the collagen fibers and has negative charge; next layer is the endothelium of the capsule. Cells of the capsule have pseudopodia and they are known as podocytes. They are the third layer of the filter.

The *glomerular filtrate* defined also as *primary urine* has a chemical composition similar to plasma but it contains almost no proteins because large molecules do not readily cross the glomerular wall. Molecules with a radius of less than 1.8 nm (molecular mass less than 10 kDa) can freely pass through the filter while those with a radius of radius greater than 4.4 nm (molecular mass greater than 80 kDa) normally cannot pass through it. Negatively charged particles are less permeable than neutral substances of equal radius because negative charges on the wall of the glomerular filter repel the ions. When small molecules are bound to plasma proteins the bound fraction is practically nonfilterable. Primary urine has not also erythrocytes because of their large diameter. Approximately 125 ml of filtrate is formed each minute for both kidneys per 1.73m^2 of body surface area. This is the index of GFR which has a little sexual difference (female GFR index equal to 110 ml/min). Easy to calculate the daily amount of the primary urine produced — 150–180 liters.

The location of the glomerulus between two arterioles allows for maintenance of a high-pressure filtration system. The capillary filtration pressure (*hydrostatic pressure*, appr. 60 mm Hg) in the glomerulus is approximately two to three times higher than that of other capillary beds in the body. Needless

to say, the high hydrostatic pressure inside the glomerular capillary is very favorable for filtration inasmuch as formation of glomerular urine is the result of ultrafiltration.

There are two important items concerning the forces determining the GFR:

1) because the glomerular capillary is placed between two arterioles, the hydrostatic pressure inside it is sufficient for maintaining blood flow along all its length;
2) because the urinary tubules readily accept the filtered urine, the hydrostatic pressure in Bowman's space is low and there is no reabsorption of filtered urine back into the glomerular capillary;
3) since proteins are not filtered, the concentration of plasma proteins (and oncotic pressure) increases as blood moves along the glomerular capillary length and urine gets formed.

Therefore, fluid movement is proportional to the membrane permeability and to the balance between hydrostatic and osmotic (Oncotic) forces across the glomerular membrane. When hydrostatic exceeds Oncotic pressure, filtration occurs. Conversely, when oncotic pressure exceeds hydrostatic pressure there is no filtration and reabsorption occurs.

Effective filtration pressure (EFP) is the main driving force for water and solutes transport across the glomerular membrane. EFP is a function of two variables: the *hydrostatic pressure gradient* driving fluid out of the glomerular capillary and into Bowman's capsule and the *oncotic pressure gradient* bringing fluid into the glomerular capillary. Hence, EFP could be calculated as

$$EFP = (P_{glom} - P_{cap}) - P_{onc}$$

where $(P_{glom} - P_{cap})$ — the outward forces promoting filtration (so-called transmural pressure)

P_{onc} — the inward forces opposing filtration.

In the normal conditions EFP equals to 25 mm Hg [(65 — 15) — 15 = 25].

The *filtration coefficient* (FC) of the glomerular membrane is a function of the capillary surface area and the membrane permeability. Using EFP and FC the GFR might be expressed as

$$GFR = FC \cdot EFP$$

The filtration pressure and the GFR are regulated by the following determinants. Firstly, the *phenomenon of the arterial pressure autoregulation* allows GFR and RBF to remain quite constant in a wide range of the renal arterial pressure (from 80 till 180 mm Hg). For details read part 3 of this Chapter.

The second mechanism involves the afferent and efferent arterioles constriction and relaxation. Constriction of the efferent arteriole increases resistance to outflow from the glomeruli and increases the glomerular pressure and the GFR. Constriction of the afferent arteriole causes a reduction in the renal blood flow, glomerular filtration pressure and GFR. The afferent and the efferent arterioles are innervated by the sympathetic nervous system and also are sensitive to vasoactive hormones such as angiotensin II.

During periods of strong sympathetic stimulation, such as occurs during shock, increased sympathetic tone in both the afferent and efferent arterioles additionally to the autoregulatory response to hypotension results in the renal arterioles constriction that causes a marked increase of renal vascular resistance together with RBF and GFR decrease. Consequently, urine output can fall almost to zero.

5.2. Renal tubular transport: reabsorption and secretion

Being filtered through the Bowman's capsule, the primary urine (or glomerular filtrate) moves into the tubular segments of the nephron. In its movement through the lumen of the tubular segments, the glomerular filtrate is changed considerably by the tubular transport of water and solutes. Tubular transport can result in reabsorption of substances from the tubular fluid into the blood or secretion of substances from the blood into the tubular fluid.

Segments of the renal tubule are adapted to reabsorb or secrete specific substances using particular modes of transport. The basic mechanisms of transport across the tubular epithelial cell membrane are similar to those of other cell membranes in the body and include active and passive transport mechanisms. Apart from H_2O, many inorganic ions (e.g., Na^+, Cl^-, K^+, Ca^{2+} and Mg^{2+}) and organic substances (e.g., HCO_3^-, D-glucose, L-amino acids, urate, lactate, vitamin C, peptides and proteins) are also subject to *tubular reabsorption* (Fig. 31–4).

Water and urea are passively absorbed along concentration gradients (Fig. 31–4, 2,3). Sodium, potassium, chloride, calcium and phosphate ions as well as urate, glucose and amino acids are reabsorbed using *primary or secondary active* transport mechanisms to move across the tubular membrane (Fig. 31–4, 1).

Endogenous products of metabolism such as potassium, urate, glucuronides, hippurates and sulfates and foreign substances (e.g., penicillin, diuretics and PAH) are appeared into the tubular fluids by means of the *transcellular secretion* (Fig. 31–4, 4). Transcellular transport implies that two cellular membranes — luminal and basolateral — must be crossed, usually by two different mechanisms.

Other substances such as ammonia (NH_3) and H^+ are firstly produced by tubule cells before they enter the tubule by *cellular secretion* — NH_3 enters the tubule lumen by *passive* transport (Fig. 31–4, 5) while H^+ ions are secreted by *active* transport (Fig. 31–4, 6).

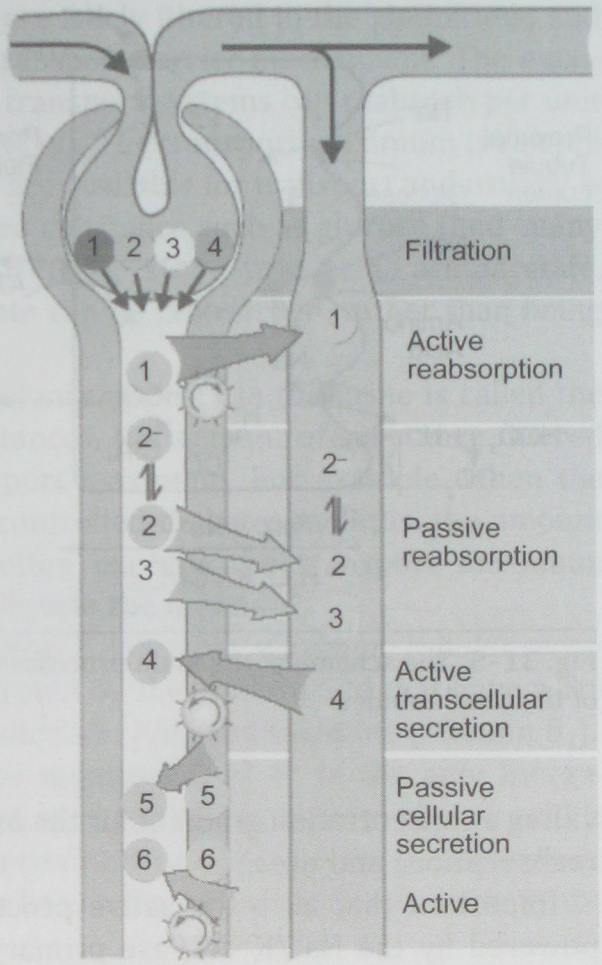


Fig. 31–4. The idealized scheme of the renal tubular transport.

Tubular secretion is very intensive process — under normal conditions only appr. 1 ml of the 125 ml of glomerular filtrate that is formed each minute is excreted in the urine. The other 124 ml is reabsorbed in the tubules. This means that the average output of urine is approximately 60 ml/hour.

Although tubular transport occurs throughout the renal tubule, most of it occurs in the *proximal tubule*. The fluid entering the proximal tubule (primary urine) is the isotonic filtrate. Approximately 65 % of all reabsorptive and secretory processes that occur in the tubular system take place in the proximal tubule.

Approximately 2/3 of the filtered Na^+ is reabsorbed in the proximal tubule by primary and secondary active transport (Fig. 31–5). Electrolytes such as K^+ and Cl^- are 65 % to 80 % reabsorbed. Two thirds amount of water follows Na^+

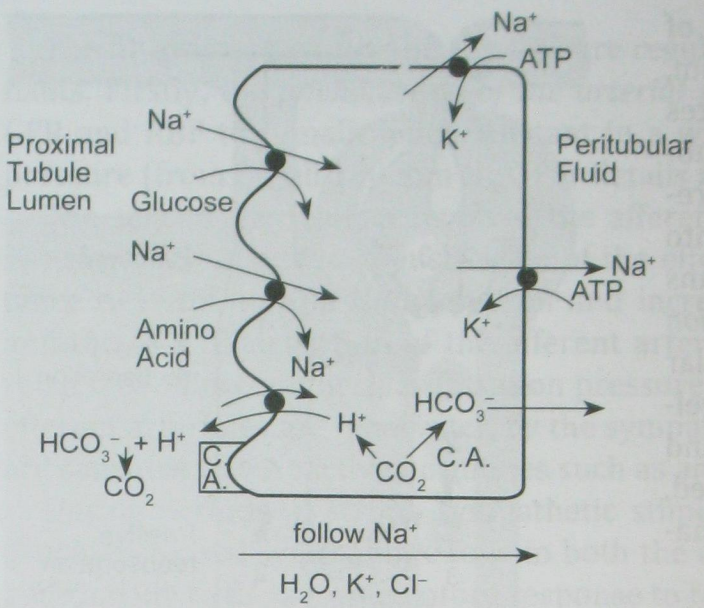


Fig. 31–5. The scheme of the main processes of the renal tubules.

also. About 80-90 % of the HCO_3^- in reabsorbed indirectly in the proximal tubules, the small amoun of these ions that leaves the proximal tubules is normally reabsorbed in the subsequent segments.

All carbohydrates, proteins, peptides, aminoacids and ketones are reabsorbed using secondary active transport (requires luminal sodium, linked to sodium reabsorption). As these solutes move into the tubular cells, their concentration in the tubular lumen decreases, providing a concentration gradient for the osmotic reabsorption of water (primary reabsorption) and urea.

Important that all o the active processes illustrated in the Fig. 31–5 are powered by the Na^+/K^+-ATPase primary active pump. This pump located in the proximal tubule basal and basolateral membrane is directly or indirectly responsible for most of the water and electrolyte reabsorption in the nephron. Thus, it represents the most energy-demanding process of the nephron.

The proximal tubule epithelium is highly permeable to water. Maximal amount of water (appr. 75-85 %) from the total volume of the primary urine (appr. 150-180 l/day) reabsorbs here. Such water reabsorption defines as *obligate* (i.e., obvious) or *primary* reabsorption. The following forces are determined the *passive mechanism* of obligate water reabsorption: a) *blood hydrostatic pressure* inside the glomerulus; b) the *osmotic pressure gradient* which primarily is established by the reabsorption of Na^+ and Cl^-; c) the special setting of the renal nephrons' descendent and ascendent tubules (the countercurency of the descendent and ascendent loops of Henle) makes more difficult it movement toward ascendent part of the loop Henle. Thus, the *hydrostatic pressure of the primary urine* increases that makes a certain impact on water passive reabsorption.

Many substances, such as glucose, are freely filtered in the glomerulus and reabsorbed by energy-dependent cotransport carrier mechanisms. The maximum amount of substance that these transport systems can reabsorb per unit time is called the *transport maximum* (Tm). The transport maximum is related to the number of carrier proteins that are available for transport and usually is sufficient to ensure that all of a filtered substance such as glucose (and many monosaccharides), phosphate ion (HPO_4^{2-}), SO_4^{2-}, uric acid, acetoacetate, β- hydroxybutyrate and α-ketoglutarate can be reabsorbed, rather than being eliminated in the urine.

The plasma level at which the substance appears in the urine is called the *renal threshold*. Under some circumstances, the amount of substance filtered in the glomerulus exceeds the transport maximum. For example, when the blood glucose level is elevated in uncontrolled diabetes mellitus, the amount that is filtered in the glomerulus often exceeds the transport maximum (appr. 320 mg/ min) and glucose spills into the urine.

The highest attainable rate of secretion is called the *maximum tubular secretory capacity*. Substances that are secreted by the kidneys and that have a Tm include penicillin, certain diuretics, salicylate, PAH and thiamine (vitamin B_1).

Two peculiarities of Tm should be mentioned: a) K^+ is the only inorganic cation that is both reabsorbed and secreted by the kidney; b) uric acid is the only organic substance to be both reabsorbed and secreted by the kidney; c) Na^+ reabsorption along the nephron has no Tm as well as K^+ secretion by the distal tubules.

The loop of Henle is divided into three segments: the thin descending segment, the thin ascending segment and the thick ascending segment. The loop of Henle, taken as whole, always reabsorbs more sodium and chloride than water. This is in contrast to the proximal tubule, which reabsorbs sodium and water in equal proportions.

The thin descending limb is highly permeable to water and moderately permeable to urea, sodium and other ions. As the urine filtrate moves down the descending limb, water moves out of the filtrate into the surrounding interstitium. Thus, the osmolality of the tubular fluid that reaches its highest point at the elbow of the loop of Henle (**1200** mOsm/kg of H_2O, hypertonic concentration, Fig. 31–6).

The ascending limb of the loop of Henle, contradictory to the descending limb, is impermeable to water. In this segment, solutes are reabsorbed but water cannot follow; as a result, the tubular filtrate becomes more and more dilute, often reaching an osmolality of **100 mOsm/kg** of H_2O as it enters the

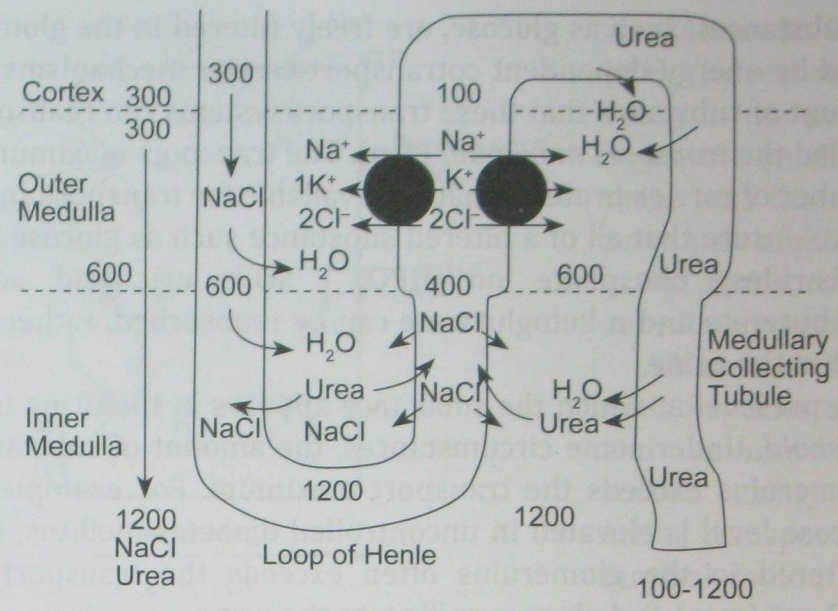


Fig. 31–6. Three limb model of the countercurrent multiplier system of the juxtamedullary nephron. Numbers inside the tubules represent the osmolality of the tubular fluid.

distal convoluted tubule. Here tubular fluid osmolality becomes to be minimal again after the initial portion of the thin segment of the loop Henle (Fig. 31–6). This allows for excretion of free water from the body. For this reason, it is often called the *diluting segment*.

The thick segment of the loop of Henle begins in the ascending limb where the epithelial cells become thickened. The beginning of the thick ascending limb marks the border between the outer and inner medulla; thus, the thick ascending limb is found only in the cortex and outer medulla (Fig. 31–6).

As with the thin ascending limb, this segment is relatively impermeable to water. The epithelium of the thick segment has a Na^+-K^+-$2Cl^-$ — the cotransport system that moves these ions out of the urine filtrate into the surrounding interstitium and peritubular capillaries (Fig. 31–6). Approximately 20 % to 25 % of the filtered load of sodium, potassium and chloride is reabsorbed in the thick loop of Henle. This transport system is selectively blocked by diuretic agents known as loop diuretics.

In approximately one fifth of the juxtamedullary nephrons, the loops of Henle and special hairpin-shaped capillaries called the *vasa recta* descend into the medullary portion of the kidney. A countercurrent mechanism controls

water and solute movement so that water is kept out of the area surrounding the tubule and sodium and urea are retained (Fig. 31–6).

The term *countercurrent* refers to a flow of fluids in opposite directions in adjacent structures. In this case, there is an exchange of solutes between the adjacent descending and ascending loops of Henle and between the ascending and descending sections of the vasa recta. Because of these exchange processes, a high concentration of osmotically active particles (approximately 1200 mOsm/ kg of H_2O) collects in the interstitium of the kidney medulla. The presence of these osmotically active particles in the interstitium surrounding the medullary collecting tubules facilitates the ADH-mediated reabsorption of water.

The kidney forms a dilute urine in the absence of ADH. In the presence of ADH the kidney forms a concentrated urine via the functioning of the *countercurrent multipliers* (i. e., the loop of Henle and the collecting duct) and the *countercurrent exchangers* (i. e., the vasa recta). It should be mentioned that regardless ADH presence the tubular fluid osmolarity usually increases progressively from the corticomedullary junction to the papillary tip (Fig. 31–6).

Distal Convoluted Tubule. Before reaching the collecting duct which is the next major component of the countercurrent multiplication the tubular fluid must pass through the distal convoluted tubule where the role of ADH again is apparent. Like the thick ascending loop of Henle, the distal convoluted tubule is relatively impermeable to water, and reabsorption of sodium chloride from this segment further dilutes the tubular fluid. Sodium reabsorption occurs through a Na^+-Cl^- — cotransport mechanism. Approximately 10 % of filtered sodium chloride is reabsorbed in this section of the tubule.

The role of ADH in this portion of the nephron influences on tubular fluid osmolarity. In case of its absence (during *diuresis*) the tubular fluid remain hypotonic (hypoosmotic) pertaining to blood plasma. In the presence of ADH (during *antidiuresis*) the Hypotonic fluid entering the distal convoluted tubules becomes isoosmotic. This is because ADH increases the water permeability of the cells of the late distal convoluted tubule and collecting duct. All the ADH (and others hormones) effects of the renal tubular transport and urine formation will be discussed in the next Chapter.

Distal Tubule and Cortical Collecting Tubule. Processes carried out in this section determine the final urine characteristics. ADH controls the final water and urea reabsorption, aldosterone — the final NaCl reabsorption and K^+ secretion. The late distal tubule and the cortical collecting tubule constitute the site where aldosterone exerts its action on Na^+ and K^+ reabsorption. Although responsible for only 2 % to 5 % of sodium chloride reabsorption, this

site is largely responsible for determining the final Na^+ concentration of the urine. The late distal tubule with the cortical collecting tubule also is the major site for regulation of K^+ excretion by the kidney. A more detailed description of these hormones actions will be discussed next in the next Chapter.

This section of the nephron is also responsible for the final acidification of the urine.

The cortical and upper medullary collecting ducts are relatively impermeable to water, urea and NaCl. Therefore, in case of ADH absence the dilute tubular fluid entering the collecting duct from the distal convoluted tubule remains Hypotonic and is excreted as a higher volume, hypotonic urine.

The secretory function of the tubules. Tubular secretion results from an activity of the epithelial cells that is associated with intensive metabolic processes. It is a transport of products from the blood and interstitial fluid into the tubuli liquid.

Owing to the secretory function of the tubules substances that do not undergo glomerular filtration are eliminated from the organism. The secretory mechanism may also take part in the elimination of urea when the level of the latter increases in the blood. Tubular secretion would seem to serve as a reserve mechanism of urine formation ensuring elimination of breakdown products.

The secretory function of the tubules may be measured by injection into the blood substances that are mainly excreted from the organism through tubular secretion such as diodrast, the sodium salt of the PAH and others. Antibiotics are eliminated from the organism quickly due to the tubular secretion.

These substances (diodrast and PAH) are completely cleared from the blood when it flows through the kidneys for the very first time, which explains why they are only present in the arterial blood flowing to the kidneys and are absent from the venous blood flowing away from them. These substances are used for determination of the RBF volume (see paragraphs #4, 5 in this Chapter).

6. CONTROL OF KIDNEY ACTIVITY

6.1. Humoral control

ADH produced in the supra-optical nucleus of hypothalamus *intensifies the reabsorption of water in the distal part of nephron and decreases the diuresis.*

It raises the permeability of the walls of the collecting tubules so that water passes from the urine into the tissue fluid of the renal medulla and into the blood.

The increase of the permeability of the tubules occurs under the influence of an enzyme *hyaluronidase*, which depolymerizes hyaluronic acid, a constituent of the intercellular matter of the tubular walls. As a result, the tubules become porous and permeable to water. The inhibitors of hyaluronidase (heparin, ascorbinic acid) are antagonists of the ADH and significantly increase elimination of water in the urine.

Deficiency of the posterior pituitary lobe cases *diabetes insipidius.* Decreased production of ADH decreases permeability of the distal parts of the nephron walls to water and great amount of water are passed in urine. In such cases the daily output of water may be as high as 20 or 25 l.

The mineralocorticoids secreted by the adrenal cortex (*aldosterone*, *desoxycorticosterone*) act upon the tubular epithelium and promote the absorption of sodium into the blood by the active transport. It is known that *aldosterone activates synthesis of the transport proteins for sodium.* Mineralocorticoids increase the renal reabsorption for Na^+ and secretion for K^+ by Na^+-K^+-pump mechanism. They also increase renal secretion for H^+.

Disease of the adrenals or their resection abolishes this mechanism and gives rise to marked loss of sodium in the urine and to severe disturbance in the organism.

Adrenaline, the hormone, secreted by the adrenal medulla, also affects urinary secretion. Increasing of the *adrenaline causes constriction of the efferent arterial vessels (vas efferens),* and thus leads to an increase of the glomerular filtration pressure. As a result amount of urine will increase. Large doses of adrenaline (at pain, stress) also constrict the afferent vessels, which reduces blood flow to the glomeruli and suppresses diuresis.

The *thyroid and parathyroid (PTH) hormones* also influense renal activity. The hormone of the thyroid gland disturbs the binding of water and salts by the tissues, causing them to pass into the blood, and causes an increase in urine production. At the same time, it intensifies all types of metabolism, protein metabolism in particular (causing an increase in the formation of its end products), which also leads to an increase in urinary secretion.

PTH increases renal Ca^{2+} reabsorption in distal tubules and inhibits renal phosphate reabsorption in proximal tubule and therefore causes phosphateuric effect.

6.2. The nervous control of renal function

Constriction of the renal vessels is observed when the sympathetic nerves innervated the kidneys are stimulated. If the afferent arteriole is constricted, filtration pressure falls and formation of filtrate is reduced. But if the efferent arterioles are narrowed, pressure in the glomeruli rises and filtration is increased.

At strong sympathetic activation (pain) as reflex as humoral mechanisms participate in the reaction and cause stoppage of urine production (*pain anuria*). Cortical control of the urine formation by conditioned reflexes is possible too.

Chapter 32. Regulation of water-electrolyte balance

1. BODY FLUID HOMEOSTASIS

Life cannot exist without water. Being the initial and final product of countless biochemical reactions, water serves as a solvent, transport vehicle, heat buffer and coolant and has a variety of other functions. Water is present in cells as *intracellular fluid* (ICF) and surrounds them as *extracellular fluid* (ECF). It provides a constant environment (*internal milieu*) for cells of the body similar to that of the primordial sea surrounding the first unicellular organisms.

Total body water (TBW) averages about 60 % of body weight in young adult men and about 50 % of body weight in young adult women. Newborns have a low percentage of body weight as water because of a relatively large ECF volume and little fat. Adult women have relatively less water than men because, on average, they have more subcutaneous fat and less muscle mass. As people age, they tend to lose muscle and add adipose tissue; hence, water content declines with age (Fig. 32–1).

Distribution of TBW is as follows: a) muscle — 50 %, b) skin — 20 %, c) other organs — 20 %, d) blood — 10 %.

About one-third of the TBW is in the extracellular fluid compartment, which contains all the fluids outside the cells, including that in the interstitial or tis-

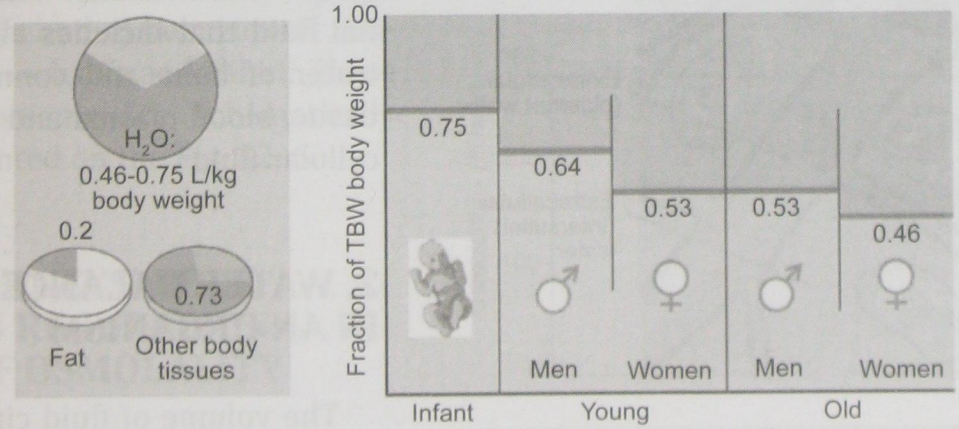


Fig. 32–1. Total body water content in an organism.

sue spaces and blood vessels. The remaining two-thirds of the TBW in healthy adults is in the ICF compartment which consists of fluid contained within all of the billions of cells in the body (Fig. 32–2).

The ECF compartment includes several subcompartments: a) *plasma* that is the liquid part of the blood — represents appr. 20-25 % of the ECF; b) *interstitial fluid* that is the liquid between cells (including lymph) — represents appr. 75-80 % of the ECF and c) *transcellular fluid* — represents appr. 1 % of the ECF.

Interstitial fluid. is contained in the gel-like extracellular matrix of glycoproteins and proteoglycans. This fluid, derived from blood plasma, provides nutrients and regulatory molecules to the cells. The extracellular environment is supported by collagen and elastin protein fibers, which also form the basal lamina below epithelial membranes (Fig. 32–3).

A third, usually minor, subdivision of the ECF compartment — the *transcellular compartment* — includes the cerebrospinal fluid and fluid contained in the various body spaces such as the peritoneal, pleural, and pericardial cavities, the joint spaces, and the gastrointestinal tract. Normally, only appr. 1 % of ECF is in the transcellular space but this amount can increase significantly in pathological conditions such as ascites, in which large amounts of fluid are sequestered in the peritoneal cavity.

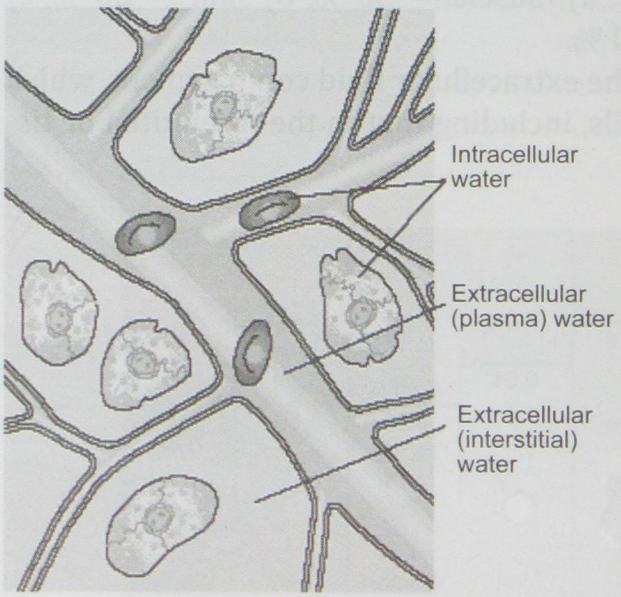


Fig. 32–2. The scheme of the of body water distribution in an organism.

Distribution of the TBW in an average 70-kg man is shown on Fig. 32–4. To resume, it should be mentioned that ICF compartment is not subdivided but the ECF consists of the following three subdivisions: interstitial fluid that includes also the water of bone and connective tissue, blood plasma and transcellular fluid

2. WATER BALANCE IN AN ORGANISM

The volume of fluid circulating in the body remains relatively constant when the *water*

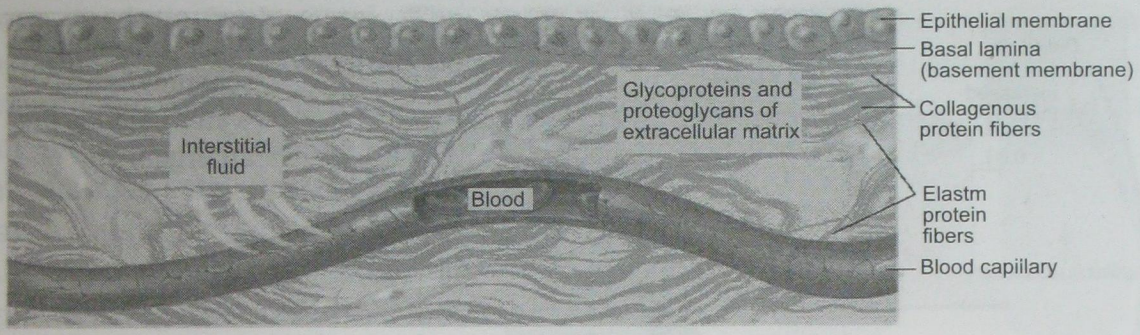


Fig. 32–3. The extracellular environment.

balance is properly regulated. The average *fluid intake* of appr. 2.5 l per day is supplied by beverages, solid foods and metabolic oxidation. The fluid intake must be high enough to counteract *water losses* due to urination, respiration, perspiration, and defecation (Fig. 32–5). The mean daily H_2O turnover is 2.5 l/70 kg (1/30-th the body weight) in adults and 0.7 liters/7 kg (1/10th the body weight) in infants. The water balance of infants is therefore more susceptible to disturbance.

Therefore, people normally stay in a stable water balance that means that water input should be equal to water output and vice versa. There are three regulatory mechanism of water balance constancy: a) hormonal regulation by means of antidiuretic hormone (ADH) activity, b) kidneys functional activity — i. e. excretion of water by the kidneys and c) behavioural activities — habit and thirst. These mechanisms are schematically presented on the Fig. 32–6.

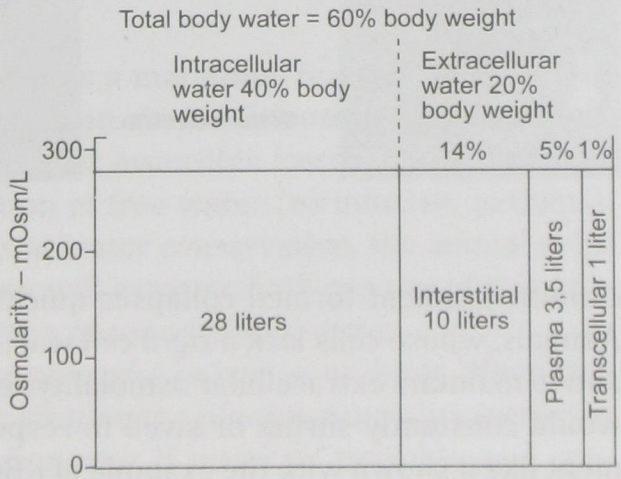


Fig. 32–4. The approximate size of body compartments in a 70-kg male adult.

3. KIDNEYS' ROLE IN REGULATION OF OSMOLALITY

Cell membranes have high water permeability and any

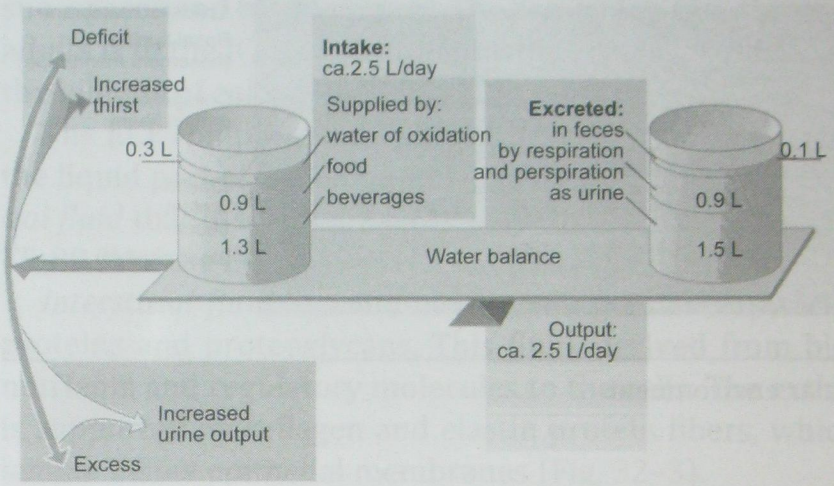


Fig. 32–5. The scheme of the water balance in an organism.

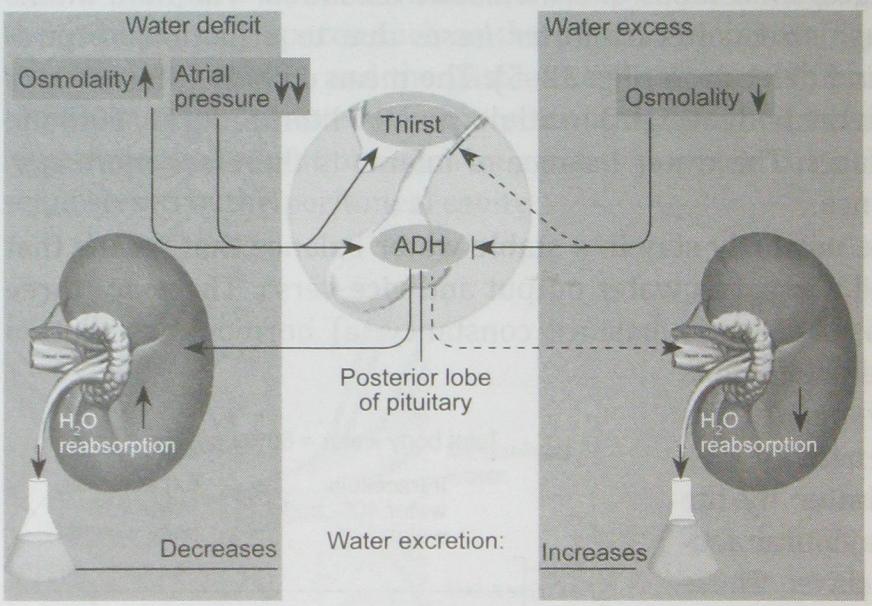


Fig. 32–6. The scheme of water balance regulation in case of different water volumes.

osmotic gradient formed collapses quickly as water moves in or out of cells. Animals, whose cells lack a rigid cell wall, must be able to regulate cell volume and to maintain extracellular osmolality relatively constant. If they did not, cells would constantly shrink or swell in response to a changing osmotic environment like it shown with the example of RBC size change in solutions of different tonicity (Fig. 32–7).

If we lived in an idealized isotonic environment there would be no need for these protective mechanisms. A long time ago, however, our vertebrate ances-

tors lived in fresh water. The osmolality of the ECF could be maintained only if there were an efficient and rapidly acting mechanism to excrete the large quantities of free water absorbed through the skin and gills. During this time the kidneys became filtering rather than secreting organs, so that large volumes of water could be promptly eliminated. Life in fresh water was dull and vertebrates moved onto land. To venture away from fresh water they now had to possess mechanisms to conserve water, especially those in hot and dry climates. Mammals developed the ability to concentrate the urine and in doing so, freed themselves to explore the planet. In short, the mammalian kidneys can both excrete huge volumes of free water as well as conserve water by forming concentrated urine. The details of these elegant mechanisms will be discussed below.

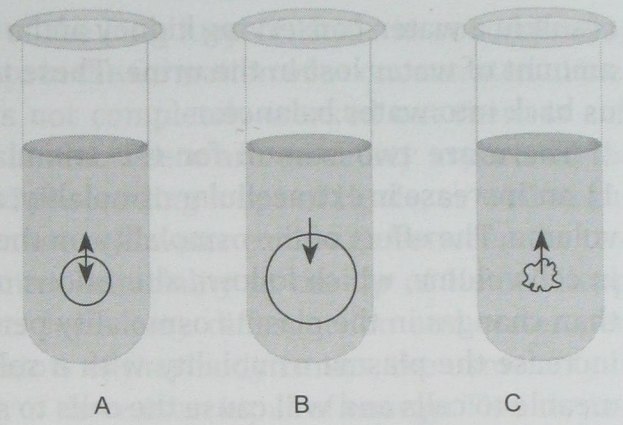


Fig. 32–7. RBC size changes in different solutions.

3.1. Conservation of water

When an animal lives in deserts or runs a marathon, there is a substantial loss of water via sweating. Even at rest in an air-conditioned room, water will be lost via the skin and lungs, the so-called insensible losses. Replacement of these losses can only occur by ingestion of free water (or infusion, perhaps). No matter how efficient the kidney is at water conservation, the animal must drink water to replace losses. Thus, we will examine both aspects of this process; thirst (and drinking) and formation of concentrated urine.

Thirst: We are all well aware of the powerful sensation of thirst. When free water is lost either via sweating or respiration the plasma osmolality increases a bit which turns on the thirst mechanisms. If water is available and consumed, the water ingested would replace the water lost. In addition, the same physiological responses which initiate thirst also stimulate the release of ADH. The thirst will make us drink free water to replenish the losses and ADH will

result in a water-conserving kidney and a concentrated urine, thus limiting the amount of water lost in the urine. These two responses, acting together, bring us back into water balance.

There are two stimuli for the stimulation of thirst and release of ADH: 1) an increase in extracellular osmolality, and 2) a reduction in the extracellular volume. The effect of the osmolality on the hypothalamus results from changes in cell volume, which follows alterations of the extracellular osmolality, rather than changes in the plasma osmolality per se. Thirst is greatly stimulated if we increase the plasma osmolality with a solute such as sodium which is impermeable to cells and will cause the cells to shrink. However, we increase plasma osmolality with a freely permeable solute such as urea or glucose, which will not cause the cells to shrink, the drinking response which follows is minimal compared to that produced by the solutes which are impermeable.

The other major stimulus for thirst (and release of ADH) is extracellular volume depletion. Even if the extracellular volume is reduced iso-osmotically (e.g. by hemorrhage), there will follow a release of ADH and an increase in thirst This latter response is likely mediated by angiotensin II.

Forming a Concentrated Urine. Each day the kidney filters 180 l of water. The potential for losing volume via the kidney is thus immense. If a mammal needs to conserve water, it must remove as much as water free urine as possible. Unfortunately, because urine is formed by filtration, there will be a significant number of osmotically active soluble compounds in the urine which will raise the osmolality of the tubular fluid as water is removed along the nephron. For the kidney to maximally conserve water, it has to remove water from the tubule leaving the osmotically active particles. As water is extracted and the osmolality of the tubular fluid rises, an indication of successful removal of water, the urine is said to become "concentrated." It is "concentrated" with respect to the plasma by virtue of the fact that the osmolality is higher than the plasma.

The kidney forms a concentrated urine by establishing a hyperosmotic medullary interstitium. As tubular fluid enters the medulla via the collecting tubule, water moves from lumen to the interstitium, across the tubule, because of the steep transtubular osmotic gradient. Water moves across the tubule only if ADH is present to make the epithelial cells permeable to water.

The secret of how the kidney establishes the hypertonic medulla lies within the structure of the kidney. Some years ago it was noted that the lengths of the long loops of Henle are correlated to the ability of the kidney to concentrate the urine. Table 1 demonstrates that animals with long loops of Henle form a urine much more concentrated than those with short loops.

What is the special significance of the long loop of Henle in concentrating the urine and how do they participate in the formation of a concentrated urine? Although the precise mechanism is not completely worked out there is sufficient data to establish a working model. There are three key aspects to the mechanism. First, the ascending and descending loops of Henle, as well as the vasa recta, are juxtaposed so that solute and water exchange between the two segments results in a very high lumen osmolality at the tip of Henle's loop. Second, the different loop segments have very different ion and water permeabilities. Third, given the constraints of thermodynamics, the osmolality of the fluid will rise above plasma only if there is an input of energy. That is, passive processes alone can not establish the medulla as hypertonic to the plasma.

It is easiest to explain the urinary concentrating mechanisms by beginning at the point at which there is input of energy, the thick ascending limb (see Chapter # 31, Fig. 31–12). This segment actively transports sodium (and chloride) out of the lumen. Because this segment is virtually impermeable to water, the urine which enters the cortex (distal tubule) is diluted with respect to plasma. For this reason this segment is also called the diluting segment. As the dilute fluid passes through the cortical collecting tubule, water is extracted so that at the start of the medullary collecting tubule, the urine is isoosmotic. If the urine osmolality increased from 50 mOsm at the end of the diluting segment to 300 mOsm at the beginning of the collecting tubule, then 5/6 of the fluid would have been reabsorbed. Given the -30 1/day of fluid which passes by the thick ascending limb, this amounts to ~25 liters of water which, without the presence of ADH, would be lost in the urine. Thus, the action of the thick ascending limb results indirectly in the reabsorption of ~25 liter of water per day.

Because both the distal convoluted tubule and cortical collecting tubule are impermeable to urea, the concentration of urea rises as water is extracted (and based on the example cited above, to the six fold). When the fluid reaches the medulla, however, urea passively diffuses out of this segment (Fig. 31-12), which is permeable to the solute and fills the interstitium with an osmotically active solute (urea). The increased osmolality draws water out of the descending loop of Henle which is permeable to water and relatively impermeable to sodium. As water is extracted the sodium concentration increases dramatically as fluid heads towards the tip of the medulla. After the turn in the loop sodium starts to diffuse out of the thin ascending limb which is permeable to sodium and impermeable to water. The result is that the medulla is rich in urea (600 mOsm) and NaCl (600 mOsm).

In sum, there are three critical phases of urine concentration:

1) dilution of the urine at the thick ascending limb by absorption of NaCl with water left behind,
2) diffusion of urea out of the collecting tubule into the medulla
3) diffusion of sodium out of the thin ascending limb into the medullary interstitium. Once the gradient is established, water will be drawn out of the medullary collecting duct under the influence of ADH.

You may wonder how the kidney maintains the gradient, which is obviously critical if the urine is to be concentrated. Won't the water reabsorbed from the collecting tubule simply wash it away? First, most of the free water is reabsorbed in the cortex, not the medulla (see above). Second, the vasa recta which also form a hairpin turn, serve to carry water out of the medulla without disturbing the gradient. Solutes entering the descending vasa recta diffuse out of the ascending vasa recta, maintaining the osmotic gradient. It is thus the passive recycling across the vasa recta which makes it a countercurrent exchanger. The loop of Henle's is a countercurrent multiplier because an active process is combined with passive process to step up the gradient at the medullary tip.

Obviously, if any pieces of the puzzle are missing or damaged, the urine cannot be concentrated. For example, furosemide is a powerful "diuretic" drug which acts by inhibiting sodium transport in the thick ascending limb. As you would predict, because this drug prevents the establishment of the medullary osmotic gradient the urine can not be concentrated.

Advantages to Forming a Concentrated Urine: It is fun to make a few calculations to demonstrate the "effectiveness" of concentrating urine.

The ability to concentrate maximally our urine above the plasma osmolality allows us to go 8-12 hrs without drinking water. In other words, we can sleep through the night without getting up to urinate. Were we unable to concentrate the urine above isotonicity, we would have to drink an extra three liters a day. One would have to get up in the middle of the night to drink ~1 quart of fluid. This would be easy with the omnipresent and ubiquitous water fountain, but try doing this in a desert.

3.2. Dilution of the urine

The kidney is able to excrete 20-30 l of free water a day. In order to excrete free water the kidney must generate free water from the filtrate, which, as you

remember, has 140 mM Na as well as other solutes. The diluting segment (thick ascending limb) reabsorbs Na(Cl) and lowers the tubular fluid osmolality to ~50 mOsm (from ~280), the minimum osmolality which can be achieved. The very dilute urine will be excreted only if it can escape the effects of ADH. That is, only if the distal segments of the tubule are completely impermeable to water will the urine remain dilute. Thus, ADH must be completely suppressed in order for the kidney to make a maximally dilute urine. Suppression of ADH release follows swelling of the hypothalamic cells after a reduction in osmolality.

The degree to which the urine osmolality falls indicates "how well" the kidney is able to dilute the urine but does not indicate "how much" water has been excreted. For example, one liter of a 50 mosm/liter urine has less "free" water in it than 3 l of a 75 mosm/liter urine. We thus need some quantitative measure, and the simplest is to calculate the free water clearance (free water clearance" is not a "clearance" as described in the section on GFR. It is a term that quantifies the amount of solute free water that is either excreted or conserved by the kidneys). That is, how much additional water is in the urine. One trick to understand this relationship is to ask yourself, "how much free water would be needed to dilute isotonic urine (280 mosm/liter) down to an osmolality of 70 mosm/liter?" You would have to dilute the fluid 1:4, adding 3 parts free water.

There are a number of factors which influence the free water clearance. Obviously, for the thick ascending limb to dilute the urine, there must be filtrate to dilute! If the GFR is low or if most of the filtered fluid is reabsorbed in the proximal segments, there will be little fluid to dilute. Thus, even though the urine osmolality may be as low as 50 mOsm, the free water clearance is greatly reduced.

4. ANTIDIURETIC HORMONE

Antidiuretic hormone [also known as arginine vasopressin (AVP)] synthesized by the nerve cells located in the supraoptic and paraventricular nuclei of the anterior hypothalamus (see Chapter # 9). The hormone travels by axoplasmic flow down the hypothalamic-neurohypophyseal tract and is stored in vesicles in nerve terminals in the median eminence and, mostly, the neurohypophysis. When the cells are brought to threshold, the ADH-containing vesicles release their contents into the interstitial fluid surrounding the nerve terminals, and hormone diffuses into nearby capillaries. The hormone is carried by the blood stream to its target tissue inducing its physiological effects.

In case of increased net water losses (hypovolemia) due, for example, to sweating, urination or respiration the ECF becomes hypertonic (Fig. 32–6, left panel). This follows by osmolality rise on 1 % or more (≥3 mOsm/kg H_2O) that is sufficient to stimulate the secretion of ADH from the posterior lobe of the pituitary gland. ADH decreases urinary H_2O excretion.

One should add that the likewise hypertonic cerebrospinal fluid stimulates central osmosensors in the hypothalamus, which trigger *hyperosmotic thirst*. The perception of *thirst* results in an urge to replenish the body's water reserves. Peripheral osmosensors in the portal vein region and vagal afferent neurons warn the hypothalamus of water shifts in the gastrointestinal tract.

In case of hypotonic fluid absorption the ECF osmolality decreases (Fig. 32– 6, right panel). This signal inhibits the ADH secretion resulting in water diuresis and normalization of plasma osmolality within less than 1 hr.

Concluding, we have to stress that there are many factors that could influence the ADH release, including pain, trauma, emotional stress, nausea, fainting, most anesthetics, nicotine, morphine and angiotensin II. These conditions or agents produce a decline in final urine output and more concentrated urine.

Ethanol and atrial sodium-uretic peptide inhibit ADH release leading to the excretion of a large volume of dilute urine.

The main factor controlling ADH release under ordinary circumstances is a change in plasma osmolality. When plasma osmolality rises, neurons called osmoreceptor cells, located in the anterior hypothalamus, shrink. This stimulates the nearby neurons in the paraventricular and supraoptic nuclei to release ADH and plasma ADH concentration rises. The result is the formation of osmotically concentrated urine (Fig. 32–6, left panel).

When plasma osmolality falls in response to the addition of excess water, the osmoreceptor cells swell, ADH release is inhibited and plasma ADH levels fall (Fig. 32–6, right panel). In this situation, the collecting ducts express their intrinsically low water permeability, less water is reabsorbed, a dilute urine is excreted, and plasma osmolality can be restored to normal by elimination of the excess water.

A second important factor controlling ADH release is the blood volume — more precisely, the effective arterial blood volume. A decreased blood volume (hypovolemia) stimulates ADH release (Fig. 32–6, left panel) whereas an increased blood volume inhibits ADH release (Fig. 32–6, right panel).

The question aroused: which receptors are responsible for ADH release? They are the receptors for blood volume — *stretch receptors* located in the left atrium of the heart and in the pulmonary veins within the pericardium.

More stretch results in more impulses transmitted to the brain via vagal afferents and inhibition of ADH release.

Arterial baroreceptors in the carotid sinuses and aortic arch also reflexly change ADH release; a fall in pressure at these sites stimulates ADH release.

One indirect mechanism of ADH released is on when a *decrease in renal bloodflow* stimulates renin release which leads to increased angiotensin II production. Angiotensin II in turn stimulates ADH release by acting on the brain.

ADH secretion determines also by blood plasma Na^+ levels With this aspects one should mentioned the following.

In case of the salt deficit hyposodiumemia occurs in the presence of a primarily normal H_2O content of the body. That will follow by both blood osmolality and ADH secretion decrease. The ECF volume, plasma volume and blood pressure consequently decrease. This, in turn, activates the renin–angiotensin system which triggers hypovolemic thirst by secreting angiotensin-II and induces Na^+ retention by secreting aldosterone. The retention of Na^+ increases plasma osmolality leading to secretion of ADH and, ultimately, to the retention of water. The additional intake of fluids in response to thirst also helps to normalize the ECF volume (Fig. 32–8, left panel).

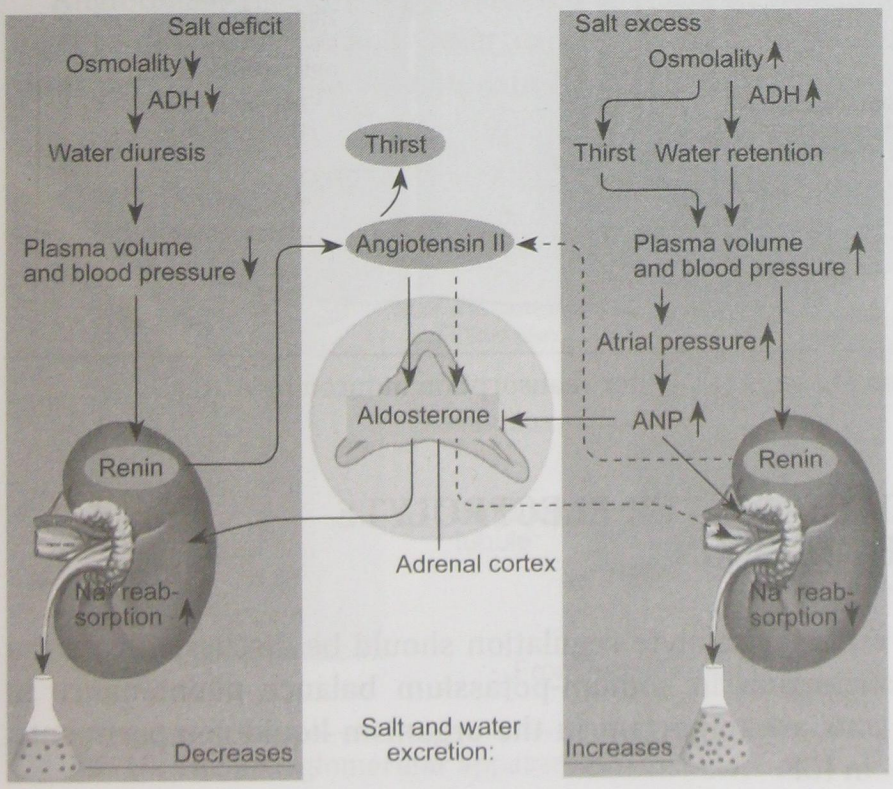


Fig. 32–8. The scheme of water balance regulation in case of different blood plasma sodium levels.

In case of salt excess in the presence of a normal H_2O volume plasma osmolality (thirst) increases that leads to ADH secretion. Thus, the ECF volume rises and renin-angiotensin system activity is curbed. The additional secretion of atrial sodium-uretic peptide leads to increased excretion of NaCl and H_2O and, consequently, to normalization of the ECF volume (Fig. 32–8, right panel).

The last one important question is that ADH mechanism of action connected with increased water reabsorption by kidneys tubes (distal segments — so- called second or active water reabsorption; Fig. 32–9) with the maximum possible antidiuresis.

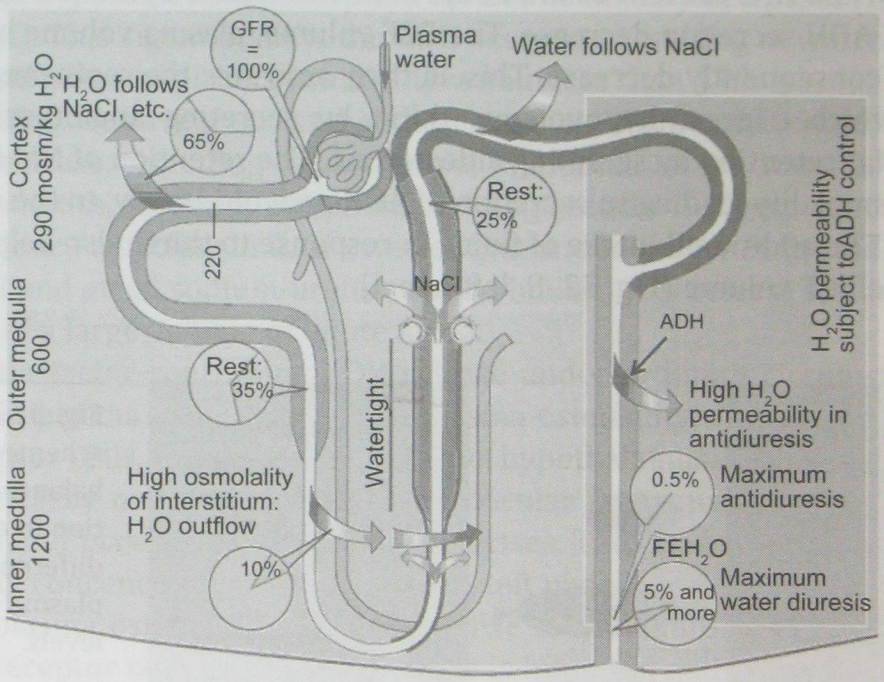


Fig. 32–9. The scheme of secondary water reabsorption induced by ADH.

5. KIDNEYS' ROLE IN WATER-ELECTROLYTE BALANCE REGULATION

Kidneys role in water-electrolyte regulation should be discussed from the position of its participation in sodium-potassium balance maintenance in an organism. Both ions are important in the organism liquid compartments: Na^+ — in ECF, K^+ — in ICF.

Sodium cation with its accompanying anions Cl^- and HCO_3^- mainly determines the ECF osmolality. Potassium has multiply important effects in the body. The kidneys play a dominant role in regulating Na^+ and K^+ balance. These regulatory mechanisms we're going to consider below.

5.1. The Renin-Angiotensin-Aldosterone Mechanism

The renin-angiotensin-aldosterone mechanism plays an important part in the short-term and long-term regulation of blood pressure (see Chapter # 21).

Briefly, renin is synthesized and stored in the juxtaglomerular cells of the kidney (Fig. 32–10).

If the mean renal blood pressure acutely drops below 90 mm Hg or so, renal baroreceptors will trigger the release of renin, thereby increasing the systemic plasma renin concentration. Renin is a peptidase that catalyzes the cleavage of angiotensin-I from the renin substrate angiotensinogen (released from the liver). Angiotensin-converting enzyme (ACE) produced in the lung, etc. cleaves two amino acids from angiotensin-I to produce angiotensin-II approx. 30– 60 min after the drop in blood pressure (Fig. 32–11).

Angiotensin-II is a potent vasoconstrictor and it acts directly on the kidneys to decrease salt and water excretion. Both mechanisms have relatively short periods of action. Angiotensin-II also stimulates aldosterone secretion by

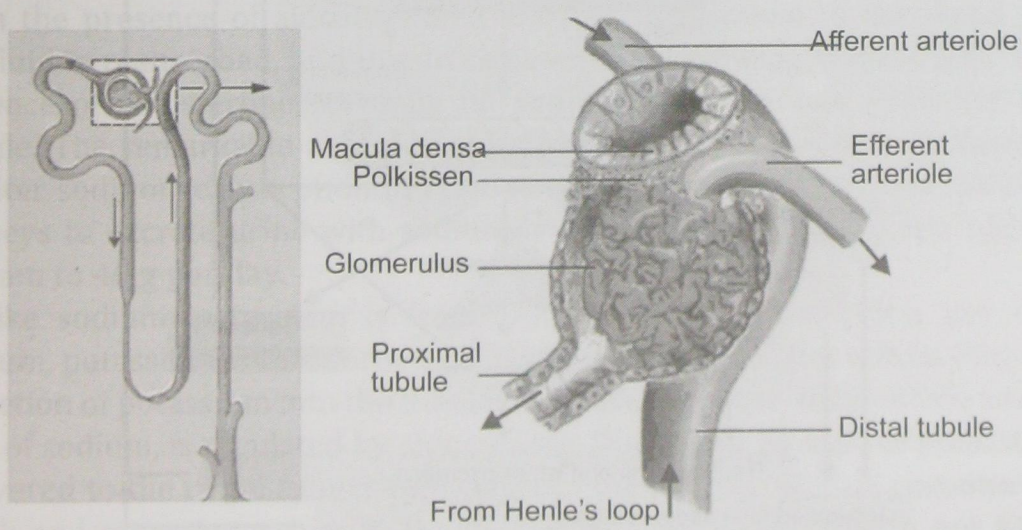


Fig. 32–10. The juxtaglomerular apparatus.

the adrenal gland. Both hormones directly (fast action) or indirectly (delayed action) lead to a renewed increase in arterial blood pressure (Fig. 32–11), and renin release therefore decreases to normal levels. Moreover, both hormones inhibit renin release (negative feedback).

If the mean blood pressure is decreased in only one kidney (e.g., due to stenosis of the affected renal artery), the affected kidney will also start to release more renin which, in this case, will lead to renal hypertension in the rest of the circulation.

Angiotensin-II induces the following fast or delayed effects:

- *Vessels:* Angiotensin-II has potent vasoconstrictive and hypertensive action, which (via endothelin) takes effect in the arterioles (*fast action*);
- *CNS*: Angiotensin-II takes effect in the hypothalamus, resulting in vasoconstriction through the circulatory "center" (*rapid action*). It also

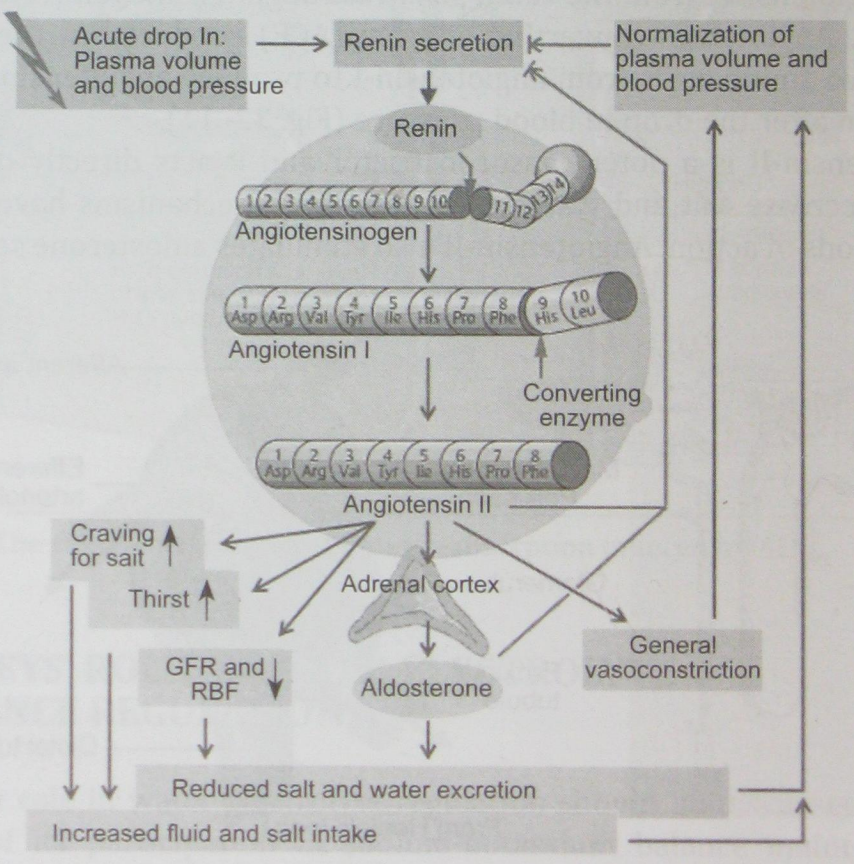


Fig. 32–11. The scheme of the renin-angiotensin-aldosterone system.

increases ADH secretion in the hypothalamus which stimulates thirst and a craving for salt (*delayed action*);

- *Kidney*: Angiotensin-II plays an important role in renal circulation and GFR regulation by constricting of the afferent and/or efferent arteriole (*delayed action*). It directly stimulates Na^+ reabsorption in the proximal tubule (*delayed action*);
- *Adrenal gland*: Angiotensin-II stimulates aldosterone synthesis in the adrenal cortex (*delayed action*) and leads to the release of epinephrine in the adrenal medulla (*fast action*).

5.2. Aldosterone

Both sodium and potassium ions reabsorption in the distal tubule and collecting duct are highly variable and depend on the presence of aldosterone that is the main mineralocorticoid hormone synthesized and secreted by the zona glomerulosa of the adrenal cortex (see Chapter # 11). As with other steroid hormones, aldosterone is not stored but is synthesized as needed. The principal function of aldosterone is to regulate Na^+ and K^+ transport in the kidney, gut, and other organs (Fig. 32–12).

Aldosterone secretion increases in response to (a) drops in blood volume and blood pressure (mediated by angiotensin-II) and (b) hyperkalemia (Fig. 32–12). Aldosterone synthesis is inhibited by atrial sodium-uretic peptide.

In the presence of aldosterone, almost all the sodium in the distal tubular fluid is reabsorbed, and the urine essentially becomes sodium free. In the absence of aldosterone, virtually no sodium is reabsorbed from the distal tubule. The remarkable ability of the distal tubular and collecting duct cells to alter sodium reabsorption in relation to changes in aldosterone allows the kidneys to excrete urine with sodium levels that range from a few tenths of a gram to 40 g per day.

Like sodium, potassium is freely filtered in the glomerulus, but unlike sodium, potassium is reabsorbed from and secreted into the tubular fluid. The secretion of potassium into the tubular fluid occurs in the distal tubule and, like that of sodium, is regulated by aldosterone. Only appr. 70 mEq of potassium is delivered to the distal tubule each day, but the average person consumes this much and more potassium in the diet. Excess potassium that is not filtered in the glomerulus and delivered to the collecting tubule therefore must be secreted into the tubular fluid for elimination from the body. In the absence of

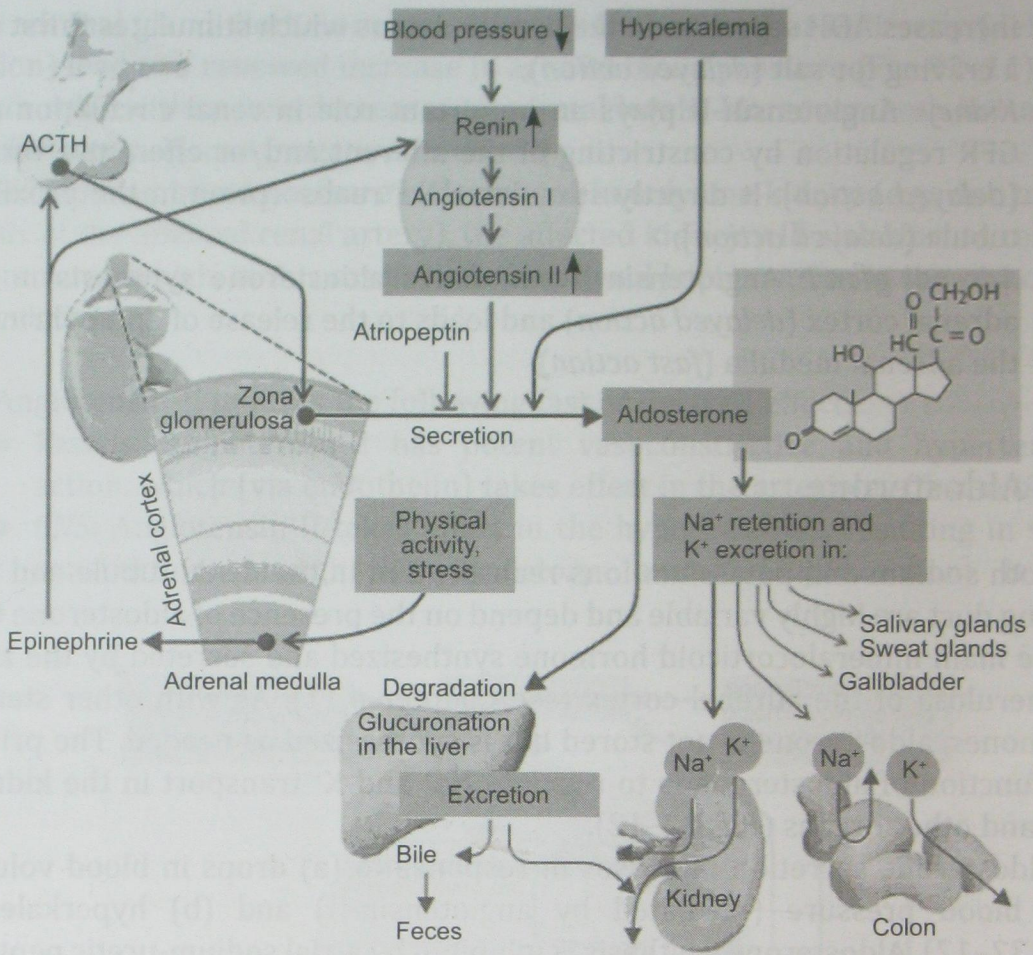


Fig. 32–12. The scheme of aldosterone synthesis, effects and metabolism.

aldosterone (as in Addison's disease), potassium secretion becomes minimal. In these circumstances, potassium reabsorption exceeds secretion, and blood levels of potassium increase.

5.3. Atrial Sodium-uretic Peptide

Atrial sodiumuretic peptide (ASP) discovered in 1981, is a hormone believed to have an important role in salt and water excretion by the kidney. It is synthesized in muscle cells of the atria of the heart and released when the atria are stretched (Fig. 32–13).

The actions of ASP include vasodilation of the afferent and efferent arterioles, which results in an increase in renal blood flow and glomerular filtration rate. ASP also inhibits sodium reabsorption from the collecting tubules through its inhibition of aldosterone secretion and through direct action on the tubular cells. It also inhibits ADH release from the posterior pituitary gland, thereby increasing excretion of water by the kidneys. ASP directly inhibits aldosterone secretion by the adrenal cortex; it also indirectly inhibits aldosterone secretion by diminishing renal renin release.

The actions of ASP are, in many respects, just the opposite of those of the renin-angiotensin-aldosterone system; it promotes salt and water loss by the kidneys and lowers blood pressure.

ASP also has vasodilator properties. Whether these effects are sufficient to produce long-term changes in blood pressure is uncertain.

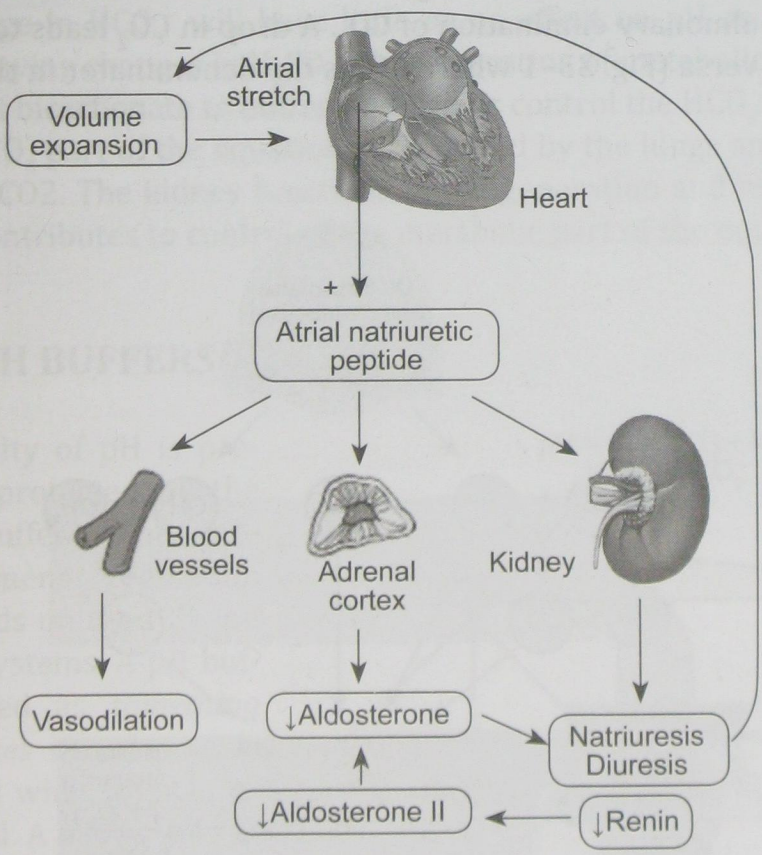


Fig. 32–13. Atrial sodium-uretic peptide and its physiologic effects.

Chapter 33. Acid-base balance

1. CALCULATION OF pH

Changes in the pH of the blood are mainly due to changes in the following factors (Fig. 33–1):

- *H^+ ions*. Direct uptake in foodstuffs (f. i., vinegar) or by metabolism, or removal from the blood (e. g., by the kidney);
- *OH^- ions*. Uptake in foodstuffs containing (basic) salts of weak acids, especially in primarily vegetarian diet;
- *CO_2*. Its concentration can change due to alterations in metabolic production or pulmonary elimination of CO_2. A drop in CO_2 leads to a rise in pH and vice versa (Fig. 33–1 where CO_2 is the denominator in the equation);

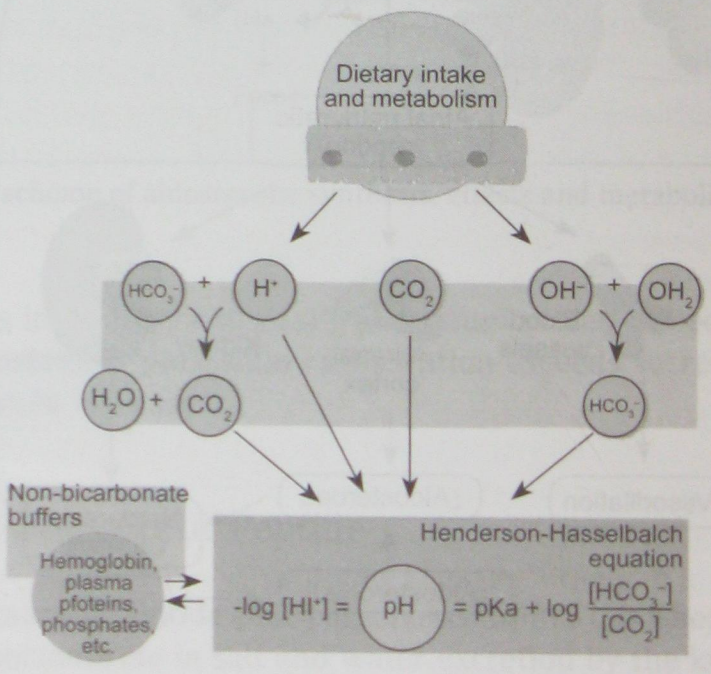


Fig. 33–1. Factors affecting blood pH.

- *HCO_3^- ions.* It can be eliminated directly from the blood by the kidney or gut (in diarrhea). A rise or fall in HCO_3^- will lead to a corresponding rise or fall in pH (Fig. 33–1 where HCO_3^- is the numerator in the equation)

The pH is calculated with the Henderson-Hasselbalch equation using the dissociation constant for the bicarbonate buffer system (equal to 6.1) and the HCO_3^- to H_2CO_3 ratio.

$$pH = 6.1 + log \frac{HCO_3^-}{H_2CO_3}$$

With HCO_3^- equal to 24mmol/l and H_2CO_3 [CO_2] equal to 1.2mmol/l, the HCO_3^- to H_2CO_3 ratio (24/1.2) equals to 20. The log of 20 is 1.3. Thus, when the ratio is 20 to 1, the pH is within the normal range at 7.4 (Fig. 33–2). It might be that change in HCO_3^- will have little or no effect on pH as long as there are accompanying changes in H_2CO_3. The generation of metabolic acids and the availability of bicarbonate to buffer these acids control the HCO_3^- part of equation. The H_2CO_3 part of the equation is regulated by the lungs and their ability to eliminate CO2. The kidney functions in the generation and reabsorption of HCO_3^- and contributes to control of the metabolic part of the equation.

2. BODY PH BUFFERS

The stability of pH is predominantly protected by the action of buffers. The moment-by-moment regulation of pH depends on the ICF and ECF buffer systems. A pH buffer is defined as *something that minimizes the change in pH* produced when an acid or base is added. A buffer system consists of a weak acid and the base salt of that acid or of a weak base and its acid salt.

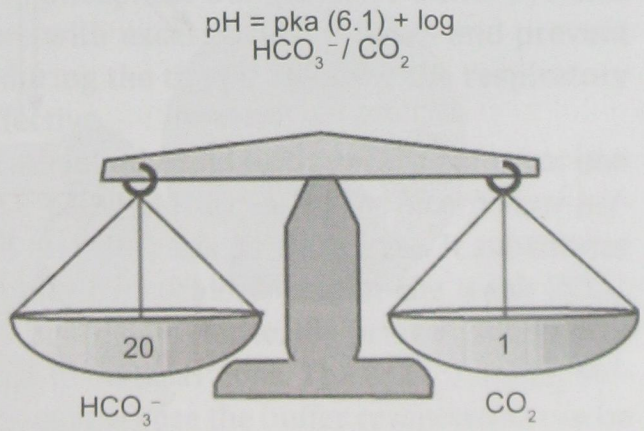


Fig. 33–2. The Henderson-Hasselbalch equation expressed as a simple manner.

In the process of preventing large changes in pH, the system trades a strong acid for a weak acid or a strong base for a weak base. Important that a buffer does not prevent a pH change.

Despite constant threats to acid-base homeostasis, a healthy person maintains a normal blood pH. The pH of body fluids is regulated by three major mechanisms:

1) ICF and ECF buffering systems which are the first line of defense of blood pH;
2) the lungs, which control the elimination of CO_2. The respiratory system is the second line of defense of blood pH;
3) the kidneys, which eliminate H^+ and regulate the elimination of HCO_3^- (Fig. 33–3). The kidneys are the third line of defense of blood pH with slower regulatory effect on blood pH level comparing to others buffering mechanisms in the body — full renal compensation may take from 1 till 3 days.

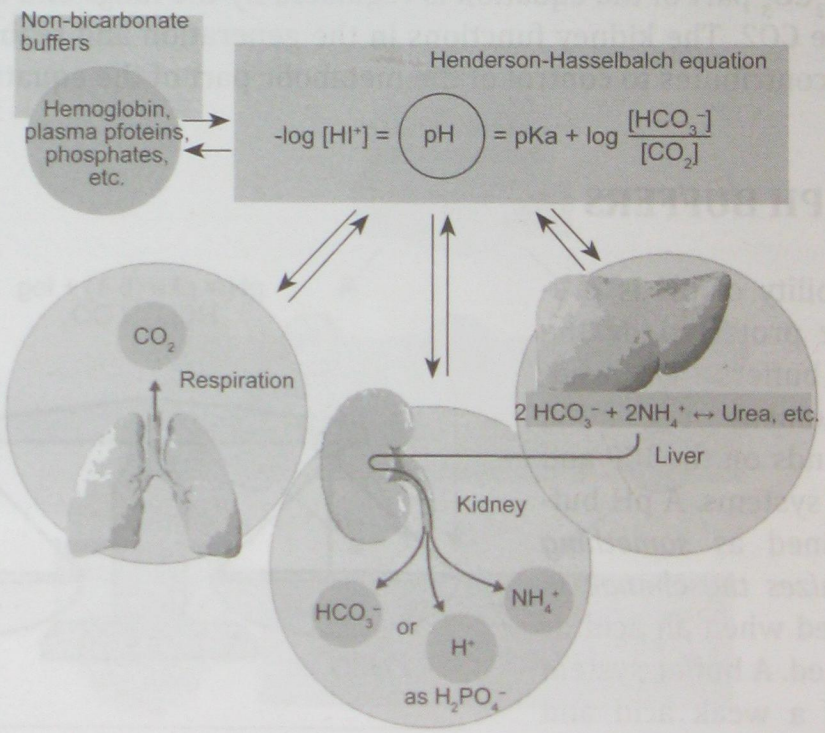


Fig. 33–3. Factors affecting blood pH.

There are two types of body buffer systems: *blood* and *tissue* buffers. It is important to note that blood buffers are not solely plasma buffers but that hemoglobin, bicarbonate and phosphate are found in erythrocytes and act as important blood buffers (Table 33–1).

Table 33–1. Whole blood buffers

Buffer type	Buffering capacity of the whole blood (%)
Bicarbonate:	
Plasma	35
Erythrocyte	18
Total bicarbonate	53
Non-bicarbonate:	
Hemoglobin and oxyhemoglobin	35
Plasma proteins	7
Organic phosphate	3
Inorganic phosphate	2
Total non-bicarbonate:	47

The four major blood buffer systems are (i) the HCO_3^- buffer system, (ii) hemoglobin (iii) proteins and (iv) phosphate buffer. These buffer systems are immediately available to combine with excess acids or bases and prevent large changes in pH from occurring during the time it takes for the respiratory and renal mechanisms to become effective.

One important buffer for blood and other body fluids is the bicarbonate/car bon dioxide (HCO_3^-/CO_2) buffer system: $CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$. *The bicarbonate buffer system* uses H_2CO_3 as its weak acid and HCO_3^- as its weak base. It substitutes the weak H_2CO_3 for a strong acid such as hydrochloric acid or the weak HCO_3^- base for a strong base such as sodium hydroxide. Hence, the primary function of the HCO_3^-/CO_2 buffer system in blood is to buffer H^+ ions. The HCO_3^-/ H_2CO_3 buffer system is a particularly efficient system because the buffer components can be readily added or removed from the body. Metabolism provides an ample supply of CO_2, which can replace any H_2CO_3 that is lost when excess base is added, and CO_2 can be readily eliminated when excess acid is added. Likewise, the kidney can

form new HCO_3^- when excess acid is added, and it can excrete HCO_3^- when excess base is added. However, this system is especially important because the concentrations of the two buffer components can be modified largely independent of each other: CO_2 by respiration and HCO_3^- by the liver and kidney.

The mechanism of bicarbonate buffer activity might be seen using the Henderson–Hasselbalch equation, where HCO_3^- = 24 mmol/l and CO_2=1.2 mmol/l (Fig. 33–4, part 1). Therefore, HCO_3^-/CO_2 ratio equals to (24/1.2) 20. Given 1.3 as log 20 and pKa=6.1, a pH equals to 7.4 (Fig. 33–4, part 2). If HCO_3^- drops to 10 and CO_2 decreases to 0.5 mmol/l the ratio of the two variables will not change and the pH will remain constant.

When added to a buffered solution, H^+ ions combine with the buffer base (HCO_3^- in this case), resulting in the formation of buffer acid ($HCO_3^- + H^+ \rightarrow CO_2 + H_2O$). In a *closed system* from which CO_2 cannot escape (Fig. 33–4, part 3), the amount of buffer acid formed (CO_2) equals the amount of buffer base consumed (HCO_3^-). The inverse holds true for the addition of hydroxide ions ($OH^- + CO_2 \rightarrow HCO_3^-$). After addition of 2 mmol/l of H^+, the abovementioned baseline ratio HCO_3^-/CO_2 of 24/1.2 (Fig. 33–4, part 2) changes to 22/3.2, making the pH fall

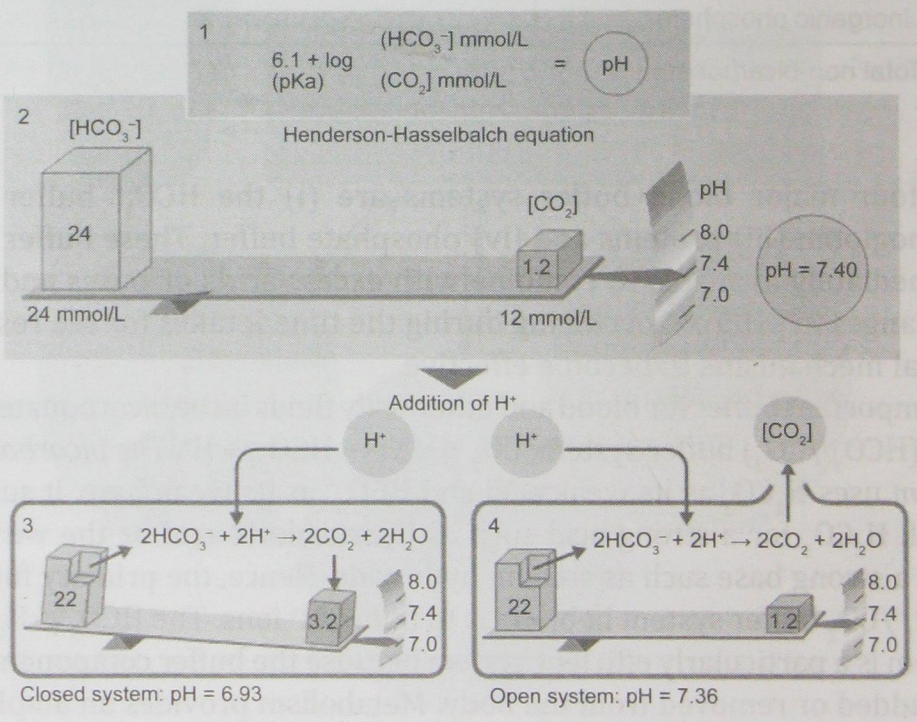


Fig. 33–4. Bicarbonate buffers in closed and open systems.

to 6.93 (Fig. 33–4, part 3). Thus, the buffer capacity of the HCO_3^-/CO_2 buffer at pH 7.4 is very low in a closed system, for which the pKa of 6.1 is too far from the target pH of 7.4.

If, however, the additionally produced CO2 is eliminated from the system (*open system*; Fig. 33–4, part 4), only the HCO_3^- will change when the same amount of H^+ is added (2 mmol/l). The corresponding decrease in the HCO_3^-/CO_2 ratio (22/1.2) and pH (7.36) is much less than in a closed system.

In the body, bicarbonate buffering occurs in an open system in which the partial pressure (PCO_2) and hence the concentration of carbon dioxide in plasma ($CO_2 = \alpha PCO_2$) are regulated by respiration (Fig. 33–5).

The lungs normally eliminate as much CO_2 as produced by metabolism (appr. 15000–20000 mmol/day), while the alveolar PCO_2 re mains constant. Since the plasma PCO_2 adapts to the alveolar PCO_2 during each respiratory cycle, the arterial PCO_2 also remains constant. An increased supply of H^+ in the periphery leads to an increase in the PCO_2 of venous blood [$H^+ + HCO_3^- \rightarrow CO_2 + H_2O$] (Fig. 33–5, part 1). The lungs eliminate the additional CO_2 so quickly that the arterial PCO_2 remains practically unchanged despite the addition of H^+ (open system).

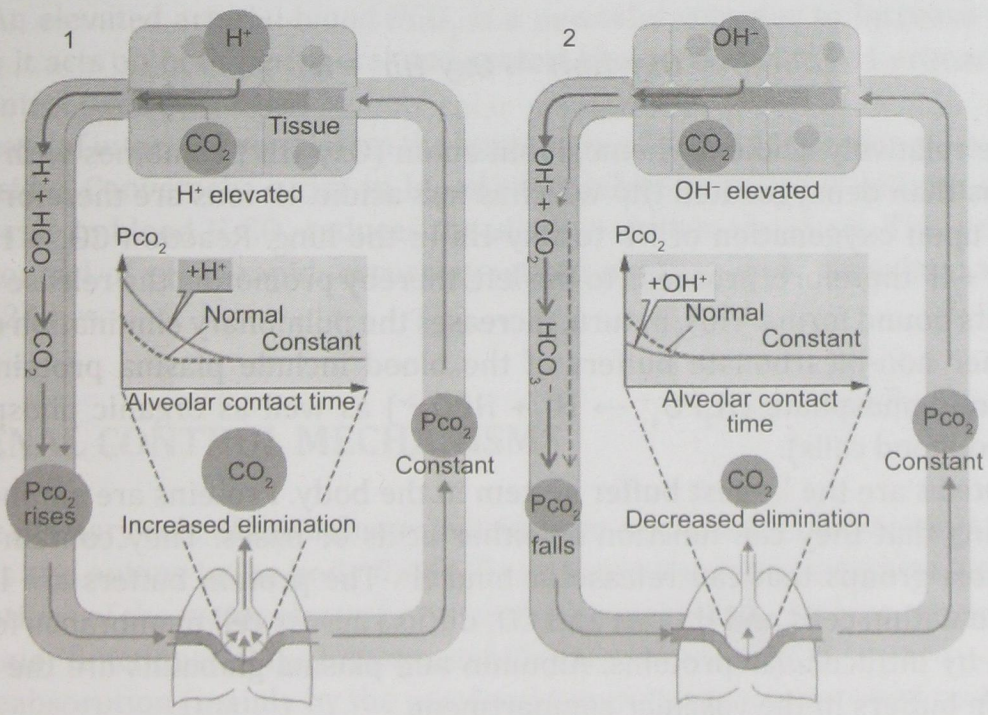


Fig. 33–5. Bicarbonate buffers in blood (open system).

An increased supply of OH^- ions in the periphery has basically similar effects. Since $OH^- + CO_2 \rightarrow HCO_3^-$, it $[HCO_3^-]$ increases and the venous PCO_2 becomes smaller than normal. Because the rate of CO_2 elimination is also reduced, the arterial PCO_2 also does not change in the illustrated example (Fig. 33–5, part 2).

At a pH of 7.4, the open HCO_3^-/CO_2 buffer system makes up about 2/3 of the buffer capacity of the blood when the PCO_2 remains constant at 5.33 kPa. Mainly intracellular non-bicarbonate buffers provide the remaining buffer capacity.

The blood contains not only the HCO_3^-/CO_2 buffer but also non-bicarbonate buffers (Table 33–1).

Hemoglobin in red blood cells which is the second most important buffer in blood is a non-bicarbonate buffer. Although hemoglobin is found intracellularly, it is more conventionally regarded as extracellular and, therefore, part of the extracellular buffer system because: (i) hemoglobin is confined to the erythrocyte which is cellular component of the ECF; (ii) hemoglobin is readily available for the buffering of extracellular acids; (iii) hemoglobin is the primary non-bicarbonate buffer of the blood. Its biochemical activity id described in the following ways:

$$\boldsymbol{HbH \rightarrow Hb^- + H^+}$$

$$\boldsymbol{Oxy\text{-}HbH \rightarrow Oxy\text{-}Hb^- + H^+}$$

The relatively acidic oxyhemoglobin anion ($Oxy\text{-}Hb^-$) combines with fewer H^+ ions than deoxygenated Hb^- which is less acidic. H^+ ions are therefore liberated upon oxygenation of Hb to Oxy-Hb in the lung. Reaction $CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$ therefore, proceeds to the left, thereby promoting the release of CO_2 from its bound forms. This, in turn, increases the pulmonary elimination of CO_2.

Other non-bicarbonate buffers of the blood include plasma proteins and inorganic phosphate ($H_2PO_4^- \rightarrow H^+ + HPO_4^{2-}$) as well as organic phosphates (in red blood cells).

Proteins are the largest buffer system in the body. Proteins are amphoteric, meaning that they can function as either acids or bases. They contain many ionizable groups that can release or bind H^+. The protein buffers are largely located within cells and H^+ ions and CO_2 diffuse across cell membranes for buffering by intracellular proteins. Albumin and plasma globulins are the major protein buffers in the vascular compartment.

Intracellular organic and inorganic substances in various tissues also function as buffers.

The major buffer capacity of the body is not in the blood but in the H^+ acceptors found in the other tissues, principally, in the muscle and in bone. Hence, since skeletal muscle represents about half of the cellular mass, *most intracellular buffering presumably occurs in muscle.*

Bone also represents an important site for the buffering of acids and bases. The total body store of bone carbonate is appr. 50 times the amount of HCO_3^- in the ICF and ECF. *Bone carbonate* appears to be the predominant source of base for neutralizing noncarbonic acid excess *in the ECF.*

The role of bone buffers is even greater in chronic acid-base disorders. One consequence of bone buffering is the release of calcium from bone and increased renal excretion of calcium. In addition to causing demineralization of bone, it also predisposes to kidney stones.

3. RESPIRATORY CONTROL MECHANISMS

The respiratory system provides CO_2 elimination into the air and plays a major role in acid-base regulation. As discussed in Chapter #24, blood pH decrease stimulates ventilation primarily by acting on peripheral chemoreceptors. An elevated arterial blood PCO_2 is a powerful stimulus to increase ventilation; it acts on both peripheral and central chemoreceptors but primarily on the central ones.

In case of intensive ventilation the lungs blow off more CO_2, making the blood less acidic. Conversely, a rise in blood pH inhibits ventilation and the consequent rise in blood H_2CO_3 reduces the alkaline shift in blood pH. The respiratory control of pH is rapid, occurring within minutes and is maximal within 12 to 24 hrs.

4. RENAL CONTROL MECHANISMS

The primary role of the kidneys in acid-base regulation is to conserve major cations and anions in the body fluids. To maintain the total quantities and concentrations of the major electrolytes within normal limits, the kidneys perform two major functions: (i) the kidneys stabilize the standard HCO_3^- pool by obligatory reabsorption (mainly by the proximal tubule) and by controlled reabsorption of filtered HCO_3^- (by the distal and collecting tubules); (ii) the kidneys excrete a daily load of 50-100 mEq of metabolically produced non-carbonic acid.

Therefore, the kidneys regulate acid-base balance by excreting acidic or alkaline urine. Excreting an acidic urine reduces the amount of acid in the ECF and excreting an alkaline urine removes base from the ECF. The renal mechanisms for regulating acid-base balance cannot adjust the pH within minutes, as respiratory mechanisms can but they continue to function for days until the pH has returned to normal or near-normal range.

4.1. The kidneys filter HCO3$^-$ in the glomerulus and then reabsorb it in the tubules as a means of maintaining ECF levels. Bicarbonate ions do not readily cross the membranes of renal tubular cells; therefore, HCO_3^- that has been filtered in the glomerulus cannot be directly reabsorbed. Instead, the HCO_3^- ions are reabsorbed in a special process in which they first combine with H^+ ions that have been secreted into the tubular fluid to form H_2CO_3 which is converted to CO_2 and water (Fig. 33–6). The resulting CO_2 can readily cross the tubular membrane and enter the tubular cell, where it combines with water, under the influence of carbonic anhydrase, to generate a new H_2CO_3 molecule. The newly formed H_2CO_3, in turn, dissociates to form a HCO_3^- and a H^+ ion. The HCO_3^- that is formed moves out of the tubular cell into the bloodstream and the H^+ is secreted into the tubular fluid.

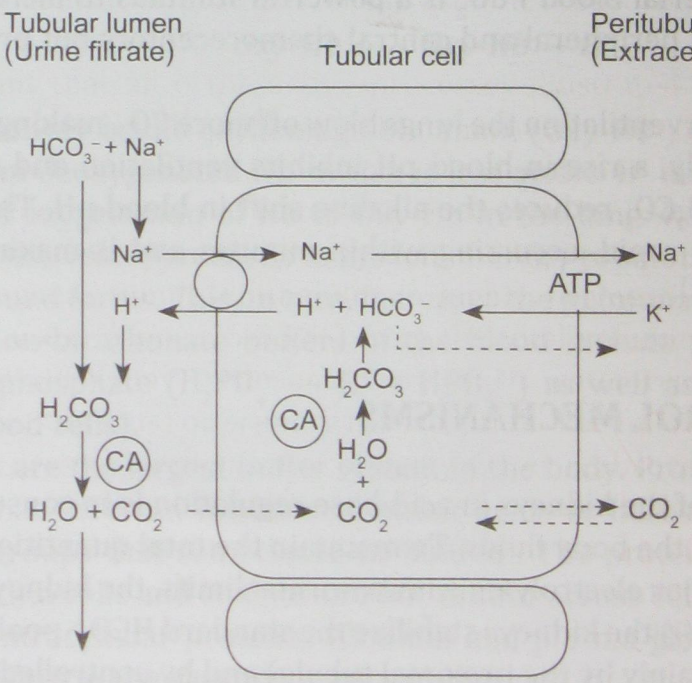


Fig. 33–6. H^+ ion secretion and HCO_3^- ion reabsorption in a renal tubular cell.

The HCO_3^- reabsorption system serves a dual purpose — it buffers the H^+ ions that have been secreted from the blood into the tubular fluid, and it conserves the HCO_3^- ions that have been filtered from the blood into the tubular fluid. Normally, only a few of the secreted H^+ ions remain in the tubular fluid because the secretion of H^+ ions is roughly equivalent to the number of HCO_3^- ions that are filtered in the glomerulus.

When the number of H^+ ions secreted into the tubular fluid exceeds the amount of filtered HCO_3^- ions, the urine filtrate becomes acidic.

The elimination of H^+ is accomplished through secretion into the urine filtrate by the cells in the renal tubules. Because extremely acidic urine would be damaging to structures in the urinary tract, the pH of the urine is maintained within a range from 4.5 to 8.0. This limits the number of unbuffered H^+ ions that can be excreted by the kidney. When the number of free H^+ ions secreted into the tubular fluid threatens to cause the urine to become too acidic, the H^+ ions must be carried in some other form. This is accomplished by combining the H^+ ions with intratubular buffers before being excreted in the urine.

4.2. There are two important intratubular buffer systems: the *phosphate buffer* system and the *ammonia buffer* system.

The phosphate buffer system uses HPO_4^{2-} and $H_2PO_4^-$ that are present in the tubular filtrate to buffer H^+. The combination of H^+ with HPO_4^{2-} to form HPO_4^- allows the kidneys to increase their secretion of H^+ ions. Because they are poorly reabsorbed, the phosphates become more concentrated as they move through the tubules. This system works best when the renal tubular fluid contains a high concentration of H^+ ions.

Another buffer system is *the ammonia buffer system.* Renal tubular cells are able to use the aminoacid glutamine to synthesize ammonia (NH_3) and secrete it into the tubular fluid. The H^+ ions then combine with the NH_3 to form an ammonium ion (NH_4^+). The NH_4^+ ions combine with Cl^- ions which are present in the tubular fluid, to form NH_4Cl, which is then excreted in the urine (Fig. 33–7).

4.3. Plasma potassium levels influence renal elimination of H^+ ions and vice versa. When plasma potassium levels fall, there is movement of K^+ ions from tubular cells into the plasma and a reciprocal movement of H^+ ions from the plasma into tubular cells. Potassium depletion also produces a reduction in Cl^- reabsorption in the distal tubule. The result is increased reabsorption of the filtered bicarbonate and development of metabolic alkalosis. An elevation in serum potassium has the opposite effect.

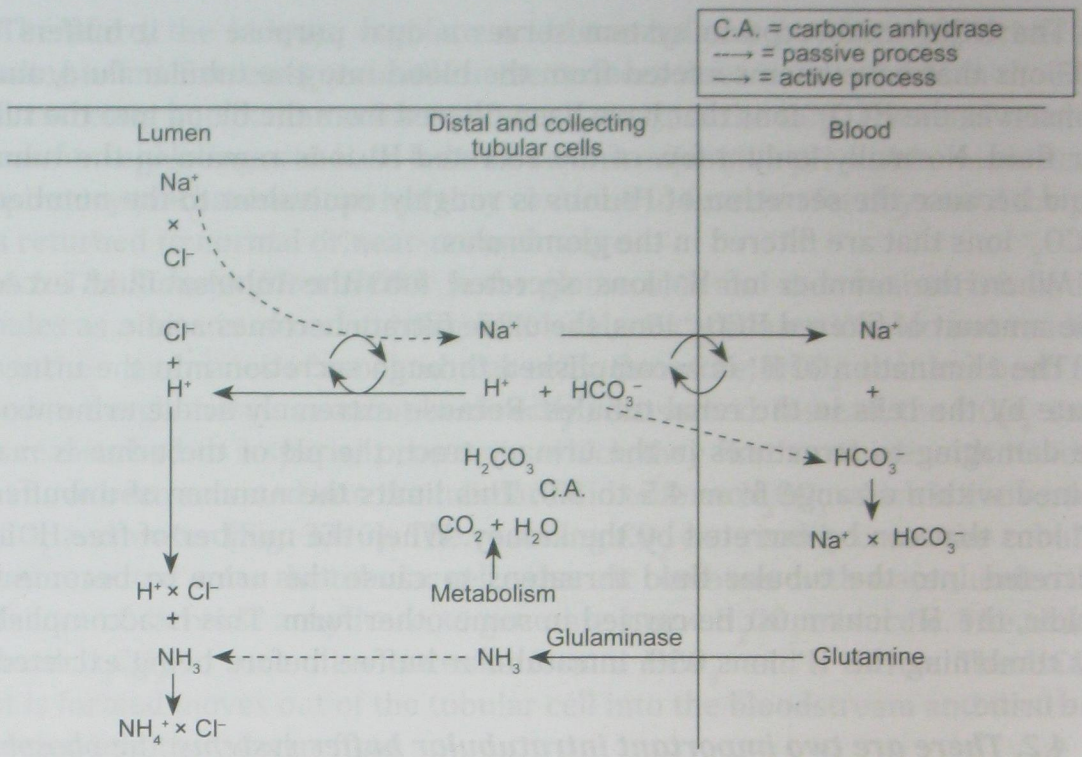


Fig. 33–7. The scheme of NH_4^+ excretion in the distal and collecting tubules.

4.4. Aldosterone also influences H^+ ion elimination by the kidney. It acts in the collecting duct to indirectly stimulate H^+ ion secretion, while increasing Na^+ ion reabsorption and K^+ secretion. Hyperaldosteronism tends to lead to a decrease in serum K^+ levels and increased pH and alkalosis caused by increased H^+ ion secretion. Hypoaldosteronism has the opposite effect. It leads to increased K^+ levels, decreased H^+ ion secretion and acidosis.

4.5. One of the mechanisms that the kidneys use in regulating the pH of the ECF is the *HCO_3^--ions conservation or elimination.* Appr. 90 % of the filtered HCO_3^- is reabsorbed into the proximal tubules as a result of the Na^+-H^+ exchange. Chloride is the most abundant anion in the ECF and can substitute for HCO_3^- when an anion shift is needed. As an example, serum HCO_3^- levels normally increase as HCl is secreted into the stomach after a heavy meal causing what is called the postprandial alkaline tide. Later, as the Cl^- ion is reabsorbed in the small intestine, the pH returns to normal.

Any secreted H^+ that combines with HCO_3^- in the lumen to bring about HCO_3^- reabsorption does not contribute to the urinary excretion of acid. Thus, the

majority of the secreted H+ is used to accomplish HCO_3^- reabsorption. Referring to this process as HCO_3^- reabsorption, then, seems inaccurate, since the HCO_3^- that appears in the peritubular capillaries is not the same HCO_3^- that was filtered. Moreover, most of the H^+ that were secreted into the lumen is not excreted in the urine but is incorporated into water and reabsorbed.

4.6. Urinary acid excretion. If the dietary protein intake is 70 g per day, a daily load of about 190 mmol of H^+ occurs after the amino acids of the protein have been metabolized. HCl (from arginine, lysine and histidine), H_2SO_4 (from methionine and cystine), H_3PO_4 and lactic acid are the main sources of H^+ ions. They are "fixed" acids which, unlike CO_2, are not eliminated by respiration.

Since appr. 130 mmol of H^+/day are used to break down organic anions (glutamate, aspartate, lactate, etc.), the net H^+ production is about 60 (from 40 till 80) mmol/day. Although the H^+ ions are buffered at their production site, they must be excreted to regenerate the buffers.

In extreme cases, the urinary pH can rise to about pH=8.0 (high HCO_3^- excretion) or fall to about pH=4.5. At a daily urine output of 1.5 l, the kidneys will excrete 1 % of the produced H^+ ions in their free form.

4.7. Titratable acids (80 % phosphate, 20 % uric acid, citric acid etc.) comprise a significant fraction (10–30 mmol/day) of H^+ excretion (Fig. 33–8, part 1). This amount of H^+ ions can be determined by titrating the urine with NaOH back to the plasma pH value, which is normally pH=7.4 (Fig. 33–8, part 2). Around 80 % of phosphate in the blood occurs in the form of HPO_4^{2-}

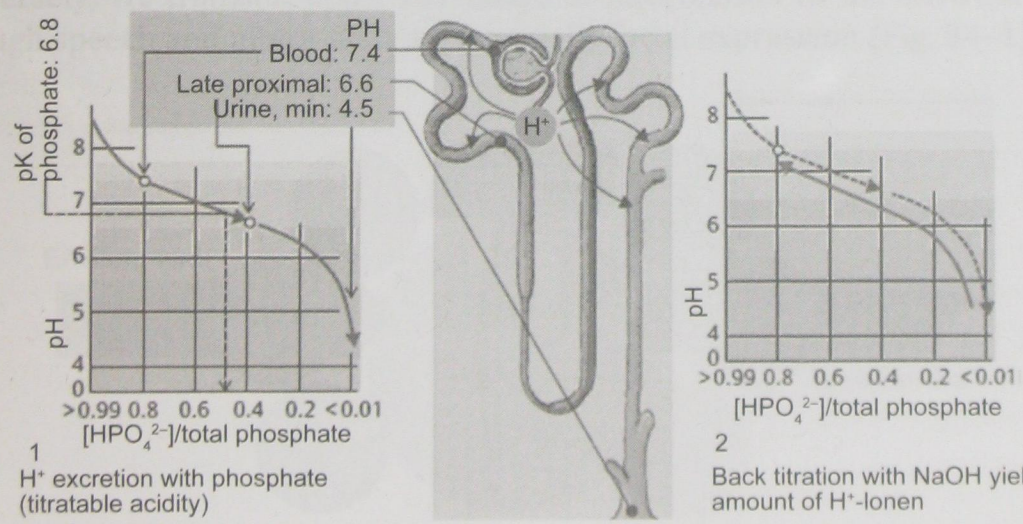


Fig. 33–8. The scheme of titratable acids excretion.

whereas about all phosphate in acidic urine occurs as $H_2PO_4^-$, i.e., the secreted H^+ ions are buffered by filtered HPO_4^{2-}. Non-reabsorbed phosphate (5–20 % of the filtered quantity) is therefore loaded with H^+ ions, about half of it in the proximal tubule (pH decreases from 7.4 to 6.6), and the rest in the collecting duct (pH goes from 6.6 to 4.5) (Fig. 33–8, part 1).

Unit XII SENSORY PHYSIOLOGY

Chapter 34. General properties of sensory systems. Proprioceptive, taste and olfactory sensory systems

1. GENERAL INFORMATION

The survival of any organism, human included, depends on having adequate information about the external environment, where food is to be found and where hazards abound.

With our senses we receive huge quantities of information from the surroundings (109 bits/s). Only a small portion of it is consciously perceived (101–102 bits/s); the rest is either subconsciously processed or not at all. Conversely, we transmit appr. 107 bits/s of information to the environment through speech and motor activity, especially facial expression (Fig. 34–1).

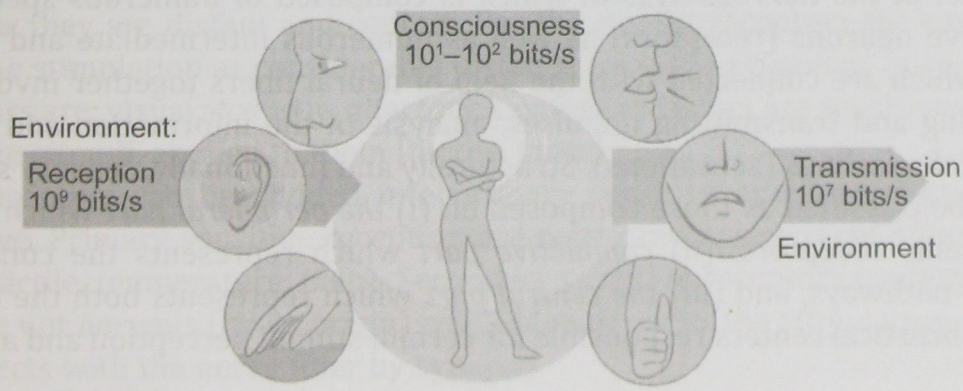


Fig. 34–1. Reception, perception and transmission of information.

Stimuli reach the body in different forms of energy, e.g., electromagnetic (visual stimuli) or mechanical energy (e.g., tactile stimuli). Various sensory receptors or sensors for these stimuli are located in the five "classic" sense organs (eye, ear, skin, tongue, nose) at the body surface as well as inside the body (e.g., propriosensors, vestibular organ). Conventionally, the sensory system is divided into (i) general senses, (ii) special senses (Table 34–1).

Table 34–1. Classification of senses

SENSES			
GENERAL			**SPECIAL**
Somatic		**Visceral** pain; baroreceptors	vision; audition; olfaction; gustation
Superficial touch-pressure; heat/ cold (thermal); pain; synthetic senses	**Deep** pain; proprio-ception		**N. B.** Vestibular apparatus (gravity detection + acceleration detection) although sometimes regarded as a special sense, has been discussed in motor section.

We will discuss further *sensory receptors* as *sensors* to distinguish them from binding sites for hormones and transmitters.

Firstly we will introduce the definition of *"sensory system"* that we mean as the part of the nervous system which is composed of numerous specialized sensitive neurons (receptors) as well as numerous intermediate and central cells which are connected with the help of neural fibers together involved in receiving and transmitting impulses, analysis of the information and formation of the feelings (sensations). Structurally and functionally, sensory systems could be classified as those composed of: (i) *the peripheral part* which mainly represents receptors; (ii) *conductive part* which represents the conductive neural pathways, and (iii) *the central part* which represents both the cortical and subcortical centers responsible for certain stimuli perception and analysis.

No satisfactory classification of senses exists. However, some existing classifications are given below. One form of classification is given in Table 34–1.

Senses may be classified also in another way. In this form, receptors, rather than the senses are classified. However, end result remains the same:

1. *Exteroceptors* are those receptors which sense the immediate external (surrounding) environment, e. g. cutaneous receptors for touch, pain, heat, cold etc. Thus the sense detected by exteroceptors may be called exteroception.
2. *Interoceptors* are those which detect any change in the internal environment (e. g., changes in blood pressure (baroreceptors), changes in O_2 tension / pH of blood (chemoreceptors) . Such senses may be called interoception.
3. *Proprioceptors* are receptors which detect the degree of tone /contraction of skeletal muscles (length detection), the degree of flexion /extension/ other movements of joints and thus the CNS is kept informed of the position of the body/ limb in space, posture and so on. Such senses may be called proprioception.
4. *Telereceptors* are receptors where the origin of the stimulus comes from a distance (f. e., vision, audition). This form of sense, thus, can be called telereception.

Alternative receptors classifications are available:

- according to their *physical nature* they are divided into: photoreceptors, mechanoreceptors, thermoreceptors, chemoreceptors, baroreceptors;
- according to the *character of sensations*, arising after their stimulation: visual, hearing, smell, taste, touch, sensation of heat and cold, posture and pain;
- according to the *distance between the source of stimulation and the receptor* they are distant and contact. For the distant receptors the source of the stimulation at considerable distance from the organism. Such receptors are: visual, acoustic, olfactory. Contact receptors are sensitive only to the stimuli, applied to them (tactile, taste);
- receptors can be divided into primary sensitive and secondary sensitive. Primary sensitive receptors are nerve endings, percept stimulation (tactile, temperature, pain). Secondary sensitive receptors contain specific not nervous (by its origin) cell, which percepts the stimulation. It connects with the nerve fiber by synapse.

2. RECEPTORS HISTOLOGY

Some receptors are encapsulated structures within which lies the beginning of the afferent nerve. The afferent nerve therefore begins from here. Examples are (i) Meissner's corpuscle and (ii) Pacinian corpuscle.

Some receptors are expanded structures which forms the beginning of an afferent sensory nerve. Examples are (i) Merkel's disc and (ii) Ruffini's end organ.

Some other receptors are free nerve endings. Pain receptor is the classical example.

Other receptors like rods and cones, cristae and maculae, carotid body chemoreceptors, muscle spindle etc. have specialized structures.

Receptors are usually classified as follows (Table 34–2).

3. SENSORY STIMULI CHARACTERISTICS

Each receptor is specialized to receive a particular type of 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, the retina can detect the presence of a single photon of light. However, a mechanical stimulus (e.g., rubbing the eyes) can produce a sensation of light.

The sensory system extracts four stimulatory elements: *modality, intensity, duration* and *localization*. Each type of sensor is specific for a unique or adequate stimulus that evokes specific sensory modalities such as *sight, sound, touch, vibration, temperature, pain, taste, smell* as well as the *body's position* and *movement* etc. Each modality has several submodalities, e.g., taste can be sweet or bitter, etc.

In secondary sensors (e.g., gustatory and auditory sensors), sensor and afferent fibers are separated by a synapse, whereas primary sensors (e.g., olfactory sensors and nocisensors) have their own afferent fibers. A stimulus induces a change in sensor potential (transduction) which results in depolarization of the sensor cell (Fig. 34–2, fragment 1) or hyperpolarization as in retinal sensors. The stronger the stimulus, the greater the amplitude of the sensor potential (Fig. 34–3, fragment 1). Once the sensor potential exceeds a certain threshold, it is transformed into an *action potential* (AP) [Fig. 34–2, fragment 1].

Table 34–2. Receptors classification

A. End organs (receptors) of special senses
Vision: rods and cones (eye). **Audition**: hair cells of organ of Corti (internal ear) . **Gustation**: gustatory cells of taste bud (tongue). **Olfaction**: olfactory neurons (nose).
B. Somasthetic and visceral sense receptors
Touch-pressure (tactile receptors): Mercel's disc; Meissner's corpuscle; Pacinian corpuscle; Ruffini's end organ; Hair end organ; Free nerve ending. **Pain (nociceptors):** free nerve ending. **Cold:** Cold receptors (Krause's end bulb in the yesteryears) Free nerve endings. **Warmth:** Free nerve endings. **Proprioception**: muscle spindle (muscle) -length detection of muscles; cristae and maculae of vestibular apparatus (internal ear) — gravity and acceleration detection. **Interoception**. chemoreceptors: Carotid body chemoreceptors — blood O_2/CO_2 tension, pH; Central chemoreceptor- detection of CO_2 (brainstem); glucoreceptors (hypothalamus). baroreceptors: carotid sinus baroreceptor (BP detection); afferent arteriole of renal artery (BP detection); left ventricular mechanoreceptors; receptors of Hering-Breuer reflex.

A transducer region and a spike generation region are contained within most receptor membranes. In some receptors (e.g., those in the eye or the ear) the transducer and spike generation functions are served by separate cells. The transducer region is responsible for converting the stimulus energy into an electrical signal [i.e., the receptor (generator) potential]. The mechanism used by the transducer region to produce the receptor potential varies depending

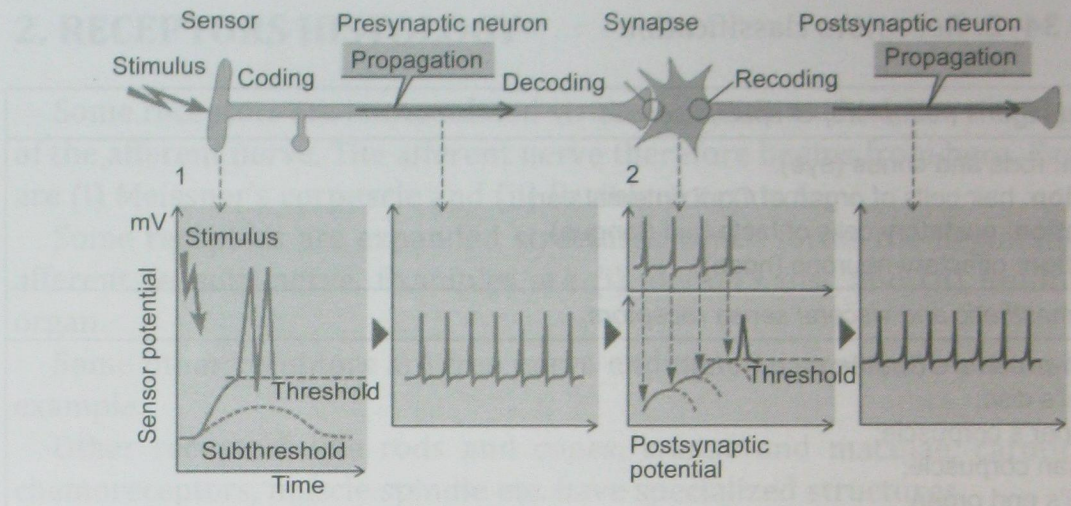


Fig. 34–2. The scheme of stimulus processing and information coding.

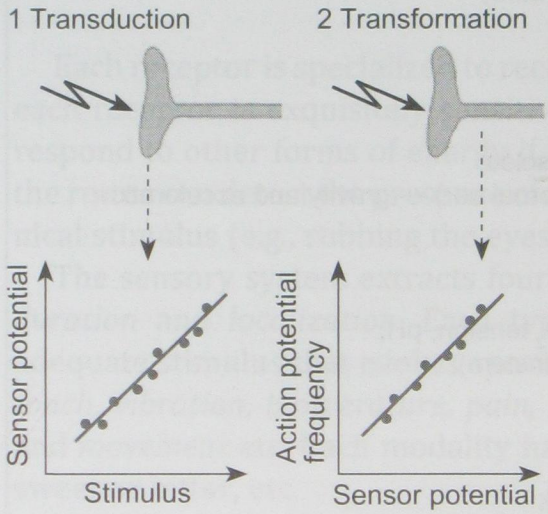


Fig. 34–3. Stimulus, sensor and action potential relationships.

on the type of receptor. The spike generator region of the receptor is responsible for converting the receptor potential into a train of action potentials.

A transducer region and a spike generation region are contained within most receptor membranes. In some receptors (e.g., those in the eye or the ear) the transducer and spike generation functions are served by separate cells. The transducer region is responsible for converting the stimulus energy into an electrical signal [i.e., the receptor (generator) potential]. The mechanism used by the transducer region to produce the receptor potential varies depending on the type of receptor. The spike generator region of the receptor is responsible for converting the receptor potential into a train of action potentials.

The receptor potential spreads passively from the transducer region to the spike generator region, if the spike generator is depolarized to threshold,

an action potential is generated. At the end of one action potential, the receptor potential causes the spike generator membrane to depolarize toward threshold. If threshold is reached again, another action potential is generated.

3.1. Coding of signals

The stimulus is encoded in AP frequency (imp/s = Hz), i.e., the higher the sensor potential, the higher the AP frequency (Fig. 34–3, fragment 2). This information is decoded at the next synapse: The higher the frequency of arriving APs, the higher the EPSP. New APs are fired by the postsynaptic neuron when the EPSP exceeds a certain threshold (Fig. 34–2, fragment 2).

Frequency coding of APs is a more reliable way of transmitting information over long distances than amplitude coding because the latter is much more susceptible to change (and falsification of its information content). At the synapse, however, the signal must be amplified or attenuated (by other neurons), which is better achieved by amplitude coding.

3.2. Adaptation

The phenomenon of adaptation is important in preventing "sensory overload," and it allows less important or unchanging environmental stimuli to be partially ignored. Adaptation causes the receptors' response to decrease despite the continued presence

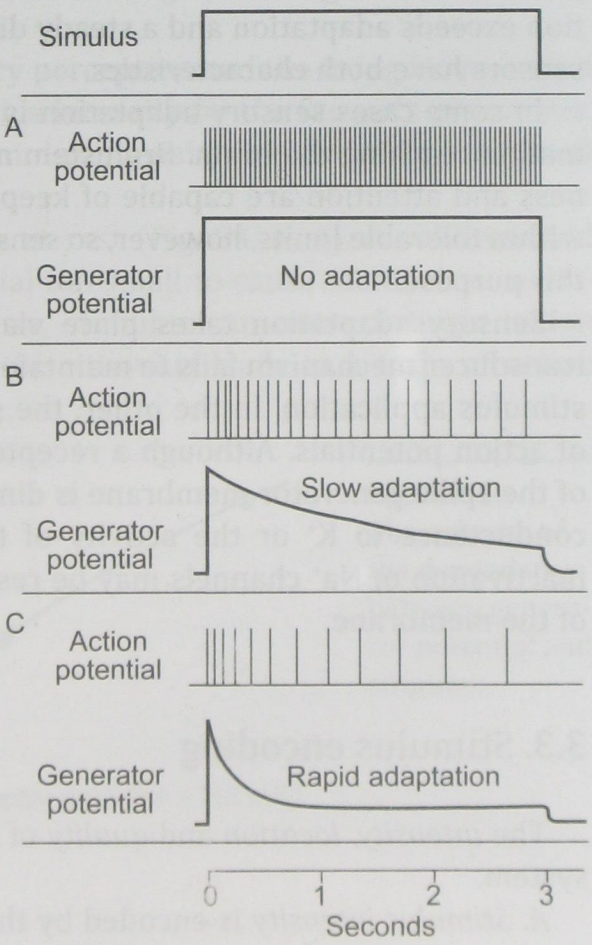


Fig. 34–4. The scheme of adaptation of slow- and fast-adapting receptors on dependence with their stimulation.

of a stimulus (Fig. 34–4, part A). Sensory adaptation functions primarily to encode information about the rate of stimulus application.

At constant stimulation, most sensors adapt, i.e., their potential decreases. The potential of *slowly adapting sensors* becomes proportional to stimulus intensity (P sensors or tonic sensors). In these types of receptors the discharge continues at a steady state as long as the stimulus is applied (Fig. 34–4, part B).

Fast-adapting sensors respond only at the onset and end of a stimulus (Fig. 34–4, part C). In these types of the receptors the discharge rate increases as the rate of stimulus application increases. They sense differential changes in the stimulus intensity (D sensors or phasic sensors). When a stimulus is applied too slowly, adaptation of the spike generator occurs before an action potential can be generated. When a stimulus is applied rapidly enough, excitation exceeds adaptation and a steady discharge of action potentials occurs. PD sensors have both characteristics.

In some cases sensory adaptation is used to decrease the amount of information reaching the brain. Brainstem mechanisms responsible for consciousness and attention are capable of keeping the information input to the brain within tolerable limits, however, so sensory adaptation is not often required for this purpose.

Sensory adaptation takes place via two major mechanisms. In one, the transducer mechanism fails to maintain a receptor potential despite continued stimulus application. In the other, the spike generator fails to sustain a train of action potentials. Although a receptor potential is present, the excitability of the spike generator membrane is diminished. An increase in the membrane conductance to K^+ or the activity of the electrogenic Na^+/K^+ pump, or the inactivation of Na^+ channels may be responsible for the decreased excitability of the membrane.

3.3. Stimulus encoding

The *intensity, location* and *quality* of a stimulus are encoded by the sensory system.

A. Stimulus intensity is encoded by the firing frequency of a sensory neuron that is proportional to the magnitude of the receptor potential. As the intensity of a stimulus increases, the magnitude of the receptor potential increases. As the intensity of a stimulus increases, the magnitude of the receptor potential increases. This relationship is expressed by the *Steven's power law function*

$$V = k \cdot I^n,$$

where V — is the magnitude of the receptor potential; k — is a constant; L — is the intensity of the stimulus and n — is a constant.

Fig. 34–5 is a plot of this relationship in which the maximal value of the receptor potential is about 50 mV. Assuming that the sensory system requires a receptor potential of at least 1 mV to detect the presence of a stimulus, the maximal stimulus that can be detected is 50 times as great as the minimally described stimulus. The power law relationship between stimulus intensity and receptor potential magnitude enables this relatively small range of receptor potential amplitudes (50 to 1) to encode a range of stimulus intensities that is greater that 1 million to 1.

Fig. 34–5 also illustrates that sensory perception and the firing frequency of a sensory neuron are related to stimulus intensity by the Steven's power law. This substantiates the concept that stimulus intensity is encoded by the rate of sensory nerve discharge.

With no stimulus the membrane is at rest (Fig. 34–6, A). A subthreshold stimulus produces a generator potential too small to cause membrane excitation (Fig. 34–6, B). A brief, but intense, stimulus can cause a single action potential (Fig. 34–6, C). Maintaining this stimulus leads to a train of action

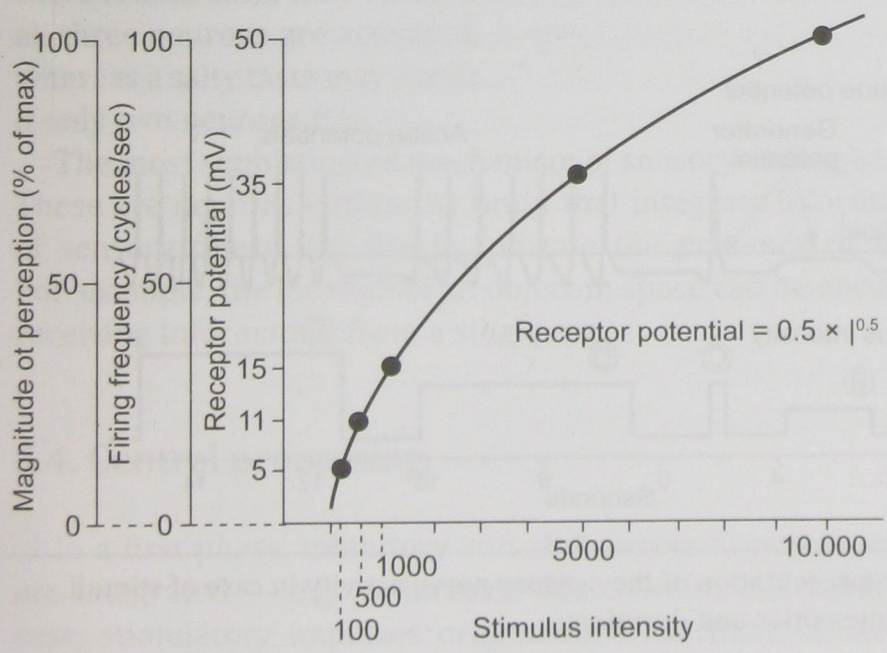


Fig. 34–5. The scheme of the dependency between generator potential and stimulus.

potentials (Fig. 34–6, D). Increasing the stimulus intensity leads to an increase in the action potential firing rate (Fig. 34–6, E).

B. Stimulus location is encoded primarily by the location of the sensory projection to the cerebral cortex. This mechanism of encoding, called *topographic representation*, is used by the visual and somatosensory systems to localize the point of stimulus application.

Each sensory neuron receives information from a particular sensory area called its *receptive field*. The smaller the receptive field, the more precise the encoding of stimulus localization. For example, the receptive fields of the fovea are the smallest within the eye. Similarly, the receptive fields within the fingertips are much smaller than those on the hands and back, where it is difficult to precisely locate the point of a mechanical stimulus.

Stimulus localization can be made more precise by *lateral inhibition*. Fig. 34–7 demonstrates three overlapping receptive fields in the skin. For two different stimuli to be perceived they have to stimulate the receptive fields separated by an unstimulated receptive field. Without lateral inhibition the two stimuli produced equal amounts of discharge in all three neurons (Fig. 34–7, A). In the presence of *lateral inhibition*, however, the neuron with the receptive field in the center is presynaptically inhibited by collaterals from the neuron with receptive fields located laterally (Fig. 34–7, B). As a result the receptive field in the centre does not fire, and two stimuli are perceived.

C. Stimulus quality is encoded by a variety of mechanisms. The simplest one uses a *labeled line* in which the stimulus is encoded by the particular neural

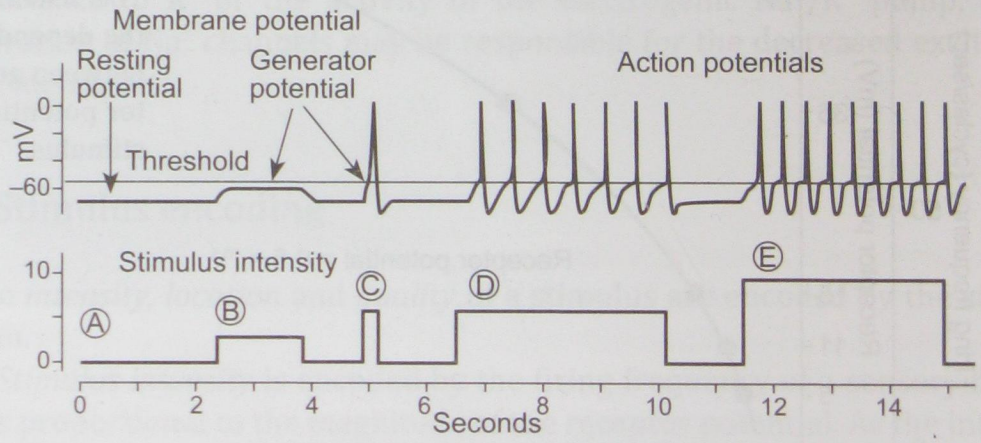


Fig. 34–6. Graphical representation of the sensory nerve activity in case of stimuli applying of different intensities and durations.

pathway that is stimulated. The basic sensory modalities are encoded in this way. Thus, the sensation of touch is elicited whether the receptors on the skin are excitated by mechanical deformation or by electrical stimulation. A more complex mechanism of coding uses the *pattern of activity* within the neural pathway that is carrying information to the brain. In *temporal pattern coding,* the same neuron can carry two different types of sensory information depending on its pattern of activity. 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.

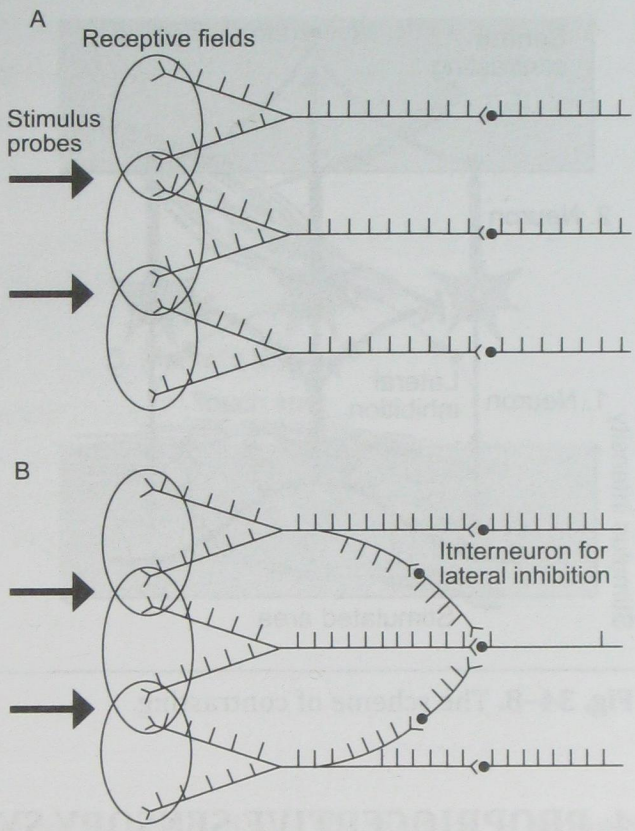


Fig. 34–7. The scheme of the lateral inhibition in the receptive fields for stimuli perceiving.

The most sophisticated mechanism of sensory coding uses *feature detectors.* These are neurons within the brain that integrate information from a variety of sensory fibers 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.

3.4. Central processing

In a first phase, inhibitory and stimulatory impulses conducted to the CNS are integrated — e.g., to increase the contrast of stimuli (Fig. 34–8). In this case, stimulatory impulses originating from adjacent sensor are attenuated

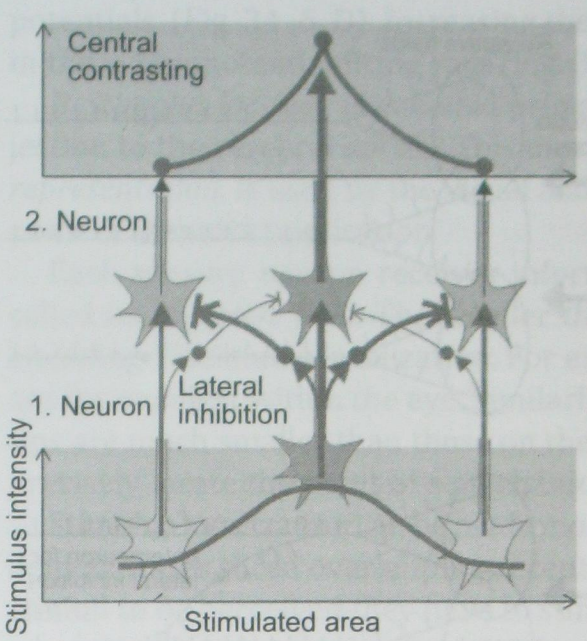


Fig. 34–8. The scheme of contrasting.

in the process (lateral inhibition). In a second step, a sensory impression of the stimuli (e.g. "green" or "sweet") takes form in low-level areas of the sensory cortex. This is the first step of subjective sensory physiology. Consciousness is a prerequisite for this process. Sensory impressions are followed by their interpretation. The result of it is called perception, which is based on experience and reason, and is subject to individual interpretation. The impression "green," for example, can evoke the perception "There is a tree" or "This is a meadow."

4. PROPRIOCEPTIVE SENSORY SYSTEM

The skin contains receptors that are adapted to encode information about touch, pain and temperature. Approximately 1 million sensory nerve fibers innervate the skin. Most of these are unmyelinated nerve fibers that are responsible for crude somatosensory mechanical sensation. Although far fewer in number, the large myelinated (A fibers) sensory fibers encode the important sensory qualities of touch, vibration and pressure. The sensations of temperature and pain are encoded by small myelinated (A-delta) fibers and unmyelinated (C) fibers.

Proprioception involves the following types of the sensitivity: (i) mechanical sensation, (ii) temperature sensation and (iii) pain sensation (nociception).

4.1. Mechanical sensation

Precise information about mechanical stimulation from the *hairless region* of the skin is provided by four receptors (Fig. 34–9).

A. The pacinian corpuscle is a very rapidly adapting receptor with a large receptive field that is used to encode high-frequency (100–400 Hz) *vibratory*

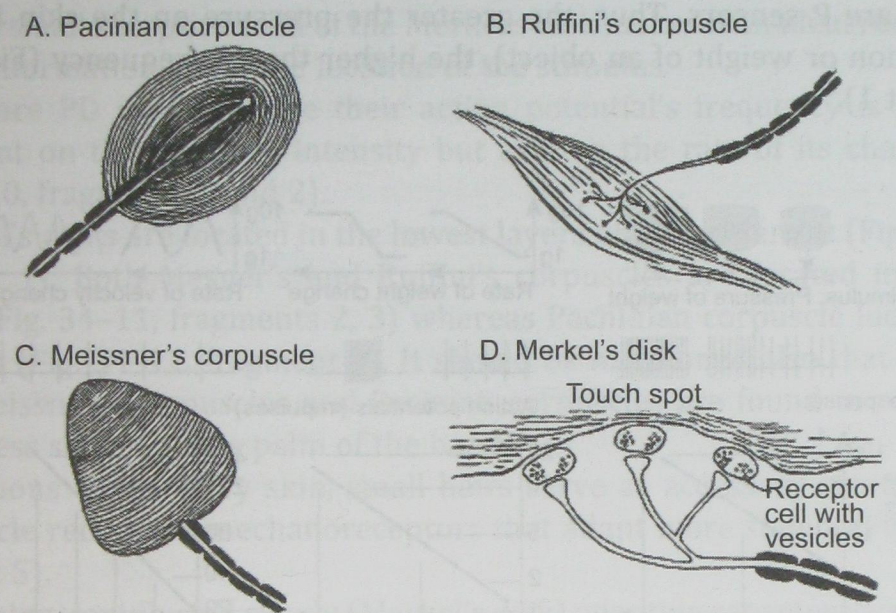


Fig. 34–9. Mechanoreceptors.

sensation (Fig. 34–9, A). The receptor is located on the end of a group B myelinated fiber, which is inserted into an onion-like lamellar capsule that is approximately 1 mm in diameter.

When a mechanical stimulus deforms the outer lamellae of the capsule, the deformation is transmitted through the capsule to the nerve terminal. The deformation of the nerve terminal increases the membrane permeability to Na^+ and K^+, producing a depolarizing receptor potential. The receptor potential amplitude increases in proportion to the magnitude of the deformation. However, because the skin is moved over an object is especially important to individuals using braille.

The pacinian corpuscle action potentials' frequency is proportional to the vibration frequency (Fig. 34–10, fragment 3).

B. The spindle-shaped *Ruffini's corpuscle* is a slowly adapting receptor that encodes pressure (Fig. 34–9, B). It has a large receptive field that is used to encode the magnitude of a stimulus. The receptor is located on the terminal of a group B axon that is covered by a liquid-filled collagen capsule. Collagen strands within the capsule make contact with the nerve fiber and the overlying skin. Any deformation or stretch of the skin causes the nerve terminal to depolarize and generate action.

They are P sensors. Thus, the greater the pressure on the skin (depth of indentation or weight of an object), the higher the AP frequency (Fig. 34–10, fragment 1).

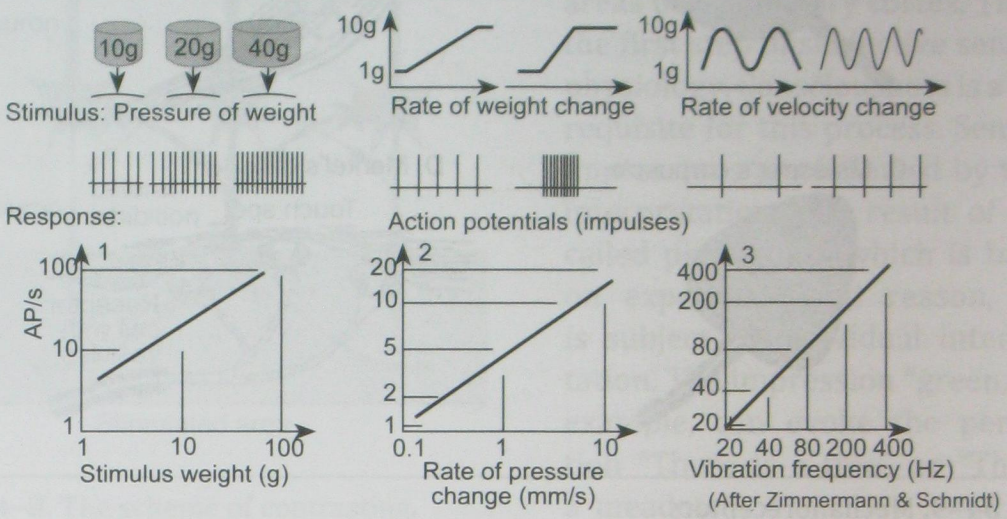


Fig. 34–10. Response of skin sensors for pressure, touch and vibration (fragments 1–3, respectively).

C. Meissner's corpuscle is a rapidly adapting receptor that participates in the *touch sensation* and low-frequency (10–100 Hz) vibration (Fig. 34–9, C). [Hair follicle receptors (Fig. 34–11, 5), which respond to bending of the hairs, assume these functions in the hairy areas of the skin]. The receptor is located at the end of a single group B afferent fiber that is inserted into a small capsule.

They are P sensors that respond only to pressure changes. Meissner's corpuscle can be stimulated only by deformation of the small region of the skin lying just above the receptor; therefore, it has a small receptive field that is used to encode the rate of stimulus application.

D. Merkel's disk is a slowly adapting receptor with a small receptive field that is also used to encode the *touch sensation* (Fig. 34–9, D). The epithelial sensory cells form synaptic connections with branches of a single group B afferent fiber.

Merkel's disk is unique because the transducer is not on the nerve terminal but on the epithelial cells that make up the disk. The disk is about 0.25 mm in diameter and can be stimulated only if the stimulus is applied directly to the

disk. The small receptive field of the Merkers disk makes it an ideal receptor to encode information about the location of the stimulus.

They are PD sensors since their action potential's frequency is not only dependent on the pressure intensity but also on the rate of its change (see Fig. 34–10, fragments 1 and 2).

Merkel's disks are located in the lowest layers of the epidermis (Fig. 34– 11, fragment 1). Both Mesner's and Ruffini's corpuscles are located inside the dermis (Fig. 34–11, fragments 2, 3) whereas Pachinian corpuscle localization is deeper (Fig. 34–11, fragment 4). It should be mentioned also that Merkel's disks, Meissner's corpuscles and Pacinian corpuscles are found in regions of the hairless skin (e.g., the palm of the hand).

In regions of the hairy skin, small hairs serve as accessory structures for hair-follicle receptors, mechanoreceptors that adapt more slowly (Fig. 34–11, fragment 5).

Resuming, rapidly and slowly (Merkel's disk) adapting receptors are densely distributed in the mouth, lips and fingertips, especially in the index and middle finger (about 100/cm^2). They can distinguish closely adjacent stimuli as separate, i.e., each afferent axon has a narrow receptive field. Since the signals do not converge as they travel to the CNS the ability of these sensors in the mouth, lips and fingertips to distinguish between two closely adjacent tactile stimuli. Correspondently, their resolution, is very high.

Slowly (Ruffini's corpuscle) adapting receptors receptors and Pacinian corpuscles have a broad receptive field. Pacinian corpuscles are therefore well adapted to detect vibrations, e.g., Earth tremors.

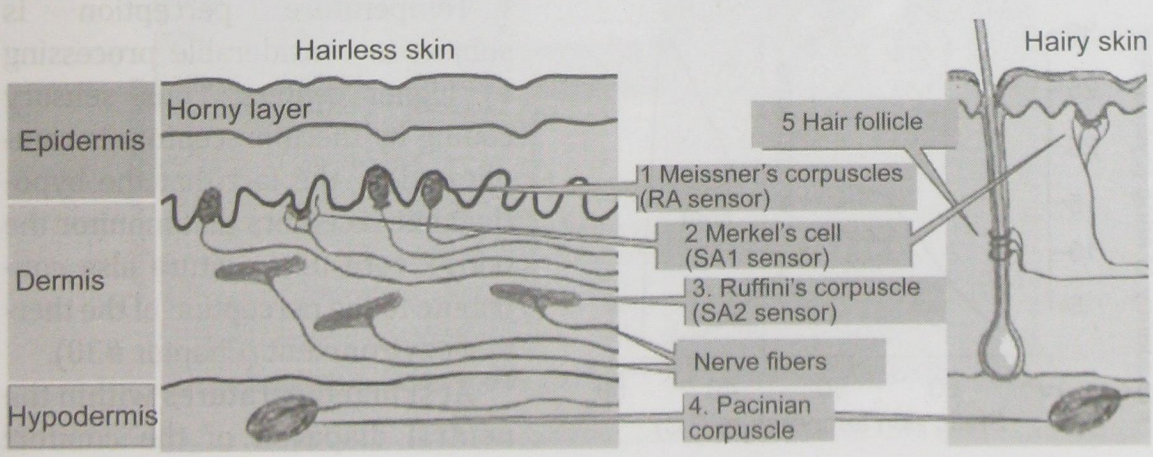


Fig. 34–11. Skin receptors localization.

4.2. Temperature sensation

Temperature sensation is encoded by thermoreceptors located on the free endings of small myelinated (A-delta group, cold receptors) and non-myelinated (C group, warm receptors) fibers. Separate receptors with discrete receptive fields exist for encoding warm and cold sensations. Warm fibers respond only when a warm object is applied to them, and cold fibers respond only to cool objects. Warm fibers are not as numerous as cold fibers.

Cold receptors form a population with a broad response peak at about 30°C, the warm receptor population has its peak at about 43°C. The lower the temperature (in the 20–36°C range), the higher the AP frequency of the cold receptors. The reverse applies to warm receptors in the 36–43°C range (Fig. 34–12).

Both sets of receptors share some common features (i) they are sensitive only to thermal stimulation, and (ii) they both have a phasic response that is rapidly adapting and responds only to temperature changes and a tonic response that depends on the local temperature.

Warm fibers transiently increase their firing rate when skin temperatures increase, and decrease their firing rate when skin temperatures decrease. Because the thermoreceptors respond transiently to the direction of a temperature change, the sensation produced by a small change in temperature depends on the current skin temperature. For example, a stimulus of 35°C will feel warm if the skin is at 30°C, and cool if the skin is at 40°C.

Cold fibers transiently increase their firing rate when skin temperatures decrease, and transiently decrease their firing rate when skin temperatures increase.

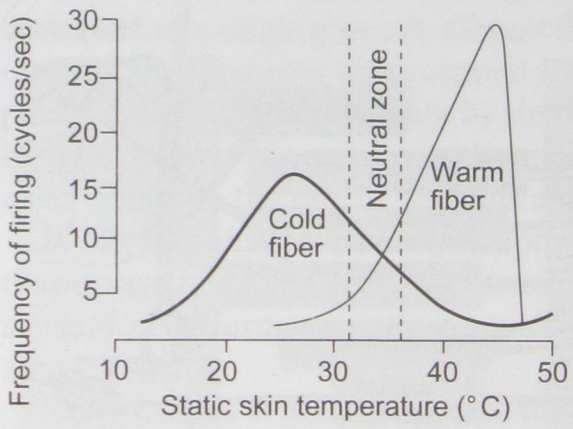


Fig. 34–12. Response of thermoreceptors.

Temperature perception is subject to considerable processing by higher centers. Thus, sensory coding by thermoreceptors is complicated by the fact that the hypothalamic receptors that monitor the body's core temperature also contribute to the perception of the thermal environment (Chapter #30).

At skin temperatures within the neutral diapason of the comfort zone (31-36°C) complete perceptual adaptation occurs (Fig. 34–

12). It should be mentioned also that in case of extreme cold and extreme hot temperature thermoreceptors change their thermal way of sensation and started to generate pain sensation.

5. OLFACTORY SENSATION

The receptor organ for olfaction is the olfactory mucosa, an area of approximately 5 cm^2 located in the roof of the nasal cavity (superior nasal concha) adjacent to the nasal septum (Fig. 34–13, fragment 1). Because of its superior position in the nasal cavity, the olfactory mucosa is not directly exposed to the flow of inspired air entering the nose.

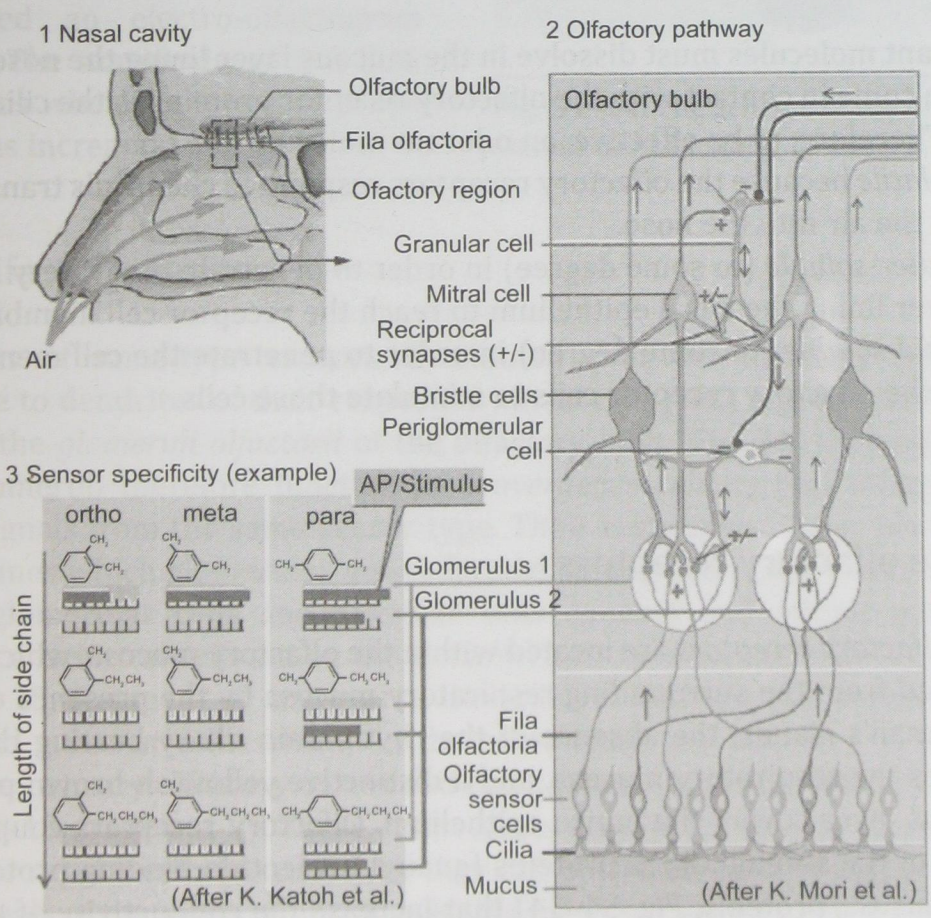


Fig. 34–13. Olfactory region, olfactory pathway and olfactory sensor specificity (1–3, respectively).

Olfactory stimuli are detected by 10^7 specialized receptors located on the free nerve endings of the olfactory nerve (cranial nerve I and some branches of the cranial nerve V) fibers. These olfactory sensor cells (Fig. 34–13, fragment 1) are bipolar neurons. Their dendrites branch to form 5–20 mucus-covered cilia, whereas the axons extend centrally in bundles called fila olfactoria (Fig. 34–13, fragments 1, 2).

5.1. Odorant molecules

Odorant molecules come in contact with the olfactory mucosa by sniffing (i.e., by short, forceful inspirations). Sniffing produces turbulence in the airflow and thereby transports molecules to the receptor cells.

Odorant molecules must dissolve in the mucous layer lining the nose before they can come in contact with the olfactory receptor proteins in the cilial membrane. Therefore, to be effective, an odorant molecule must be:

1. *Volatile* because the olfactory receptors respond to chemicals transported by the air into the nose.
2. *Water-soluble* (to some degree) in order to penetrate the watery mucous layer lining the nasal epithelium to reach the receptor cell membrane
3. *Lipid-soluble* (to some degree) in order to penetrate the cell membranes of the olfactory receptor ceils to stimulate those cells.

5.2. The olfactory receptors

The olfactory receptors are located within the olfactory mucosa, which is distinguished from the surrounding respiratory mucosa by the presence of tubular Bowman's glands, the absence of the rhythmmic ciliary beating that characterizes the respiratory mucosa, and a distinctive yellowish-brown pigment. A mucous layer covers the entire epithelium. Olfactory receptors couple with subtype of the stimulatory G proteins (guanyl nucleotide-binding proteins) — G_s-proteins G_{olf} proteins; Fig. 34–14) that increase the conductivity of the sensor cell membrane to cations, thereby increasing the influx of Na^+ and Ca^{2+} and thus depolarizing the cell. Therefore, odorant molecules adsorption on the cilia membranes of the receptor cells generates a receptor potential in these receptor

cells. Important that a specific olfactory receptor does not respond to a particular compound or category of compounds; instead, an individual receptor responds to many odors. 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.

The electrical response recorded from the olfactory mucosa in response to an olfactory stimulus is called an electro-olfactogram (EOG). The amplitude of the EOG increases when the intensity of the stimulus increases.

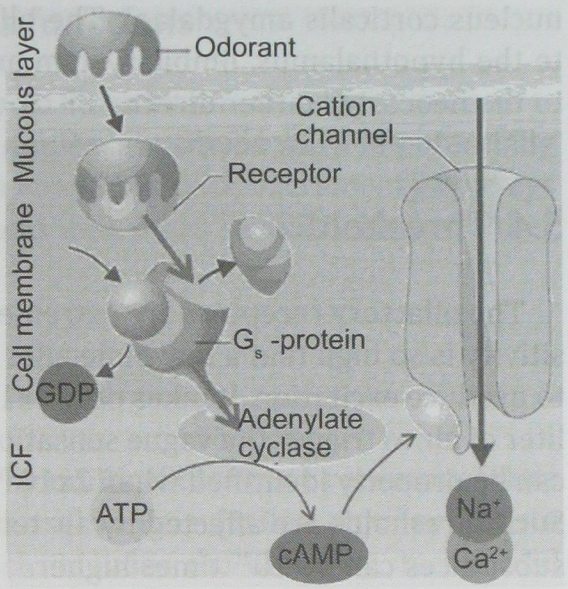


Fig. 34–14. The scheme of the olfactory stimuli transduction.

5.3. Olfactory pathway

Axons of same-type sensors distributed over the olfactory epithelium synapse to dendrites of their respective *mitral cells* (MC) and *bristle cells* (BC) within the *glomeruli olfactorii* of the olfactory bulb (Fig. 34–13, fragment 2). The glomeruli therefore function as *convergence centers* that integrate and relay signals from the same sensor type. Their respective sensor protein also determines which glomerulus newly formed sensor axons will connect to.

Periglomerular cells and granular cells connect and inhibit mitral and bristle cells. Mitral cells act on the same reciprocal synapses (see "+/–" on the Fig. 34–13, fragment 2) in reverse direction to activate the periglomerular cells and granular cells which, on the other hand, are inhibited by efferents from the primary olfactory cortex and contralateral anterior olfactory nucleus (Fig. 34–13, fragment 2). These connections enable the cells to inhibit themselves or nearby cells (contrast), or they can be disinhibited by higher centers. The signals of the axons of mitral cells (Fig. 34–13, fragment 1) reach the anterior olfactory nucleus. Its neurons cross over (in the anterior commissure) to the mitral cells of the contralateral bulb and form the olfactory tract projecting to the primary olfactory cortex (prepiriform cortex, tuberculum olfactorium,

nucleus corticalis amygdalae). The olfactory input processed there is relayed to the hypothalamus, limbic system and reticular formation; it is also relayed to the neocortex either directly or by way of the thalamus.

5.4. Threshold

The olfactory receptors are extremely sensitive. For many odorants, the sensitivity is so high that a few molecules interacting with receptors are sufficient to produce excitation. It takes only $4x10^{-15}$ g of methylmercaptan (in garlic) per liter of air to trigger the vague sensation of smell (*absolute threshold*). The odor can be properly identified when $2x10^{-13}$ g/l is present (*identification threshold*). Such thresholds are affected by air temperature and humidity; those for other substances can be 10^{10} times higher.

5.5. Odor discrimination

The olfactory system can discriminate among a vast number of odorants. In some instances, the olfactory system can discriminate dextro- and levorotatory forms as well as cis- and trans- conformations of a molecule. The different molecular moieties of an odorant molecule therefore activate different types of sensors (Fig. 34–13, fragment 3, top right). Jasmine leaves and wine contain several dozens and hundreds of odorants, respectively, so their overall scent is a more complex perception (integrated in the rhinencephalon).

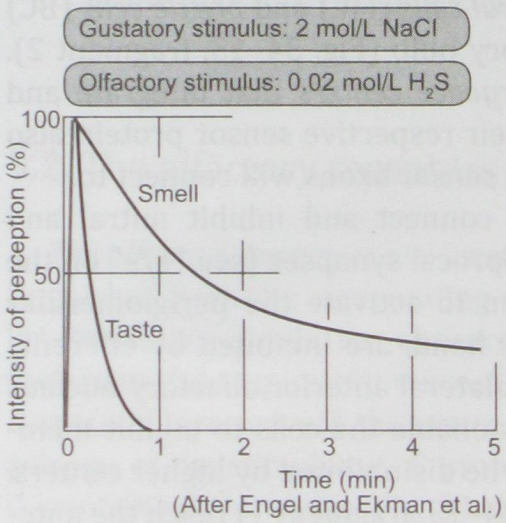


Fig. 34–15. Adaptation of smell and taste.

5.6. Adaptation

Olfactory sensation adapts very rapidly with continued exposure to an odorant. Adaptation to smell is sensor-dependent and neuronal (Fig. 34–15).

5.7. The sense of smell has various functions

The sense of smell has various *functions*. Pleasant smells trigger the secretion of saliva and gastric juices, whereas unpleasant smells warn of potentially spoiled food. Body odor permits hygiene control (sweat, excrement), conveys social information (e.g., family, enemy) and influences sexual behavior. Other aromas influence the emotional state.

6. GUSTATORY SENSATION

Gustatory (taste) stimuli are detected by multicellular taste receptors (*taste buds*) within the mouth, pharynx and dorsal tongue (called *papillae*). Taste must be distinguished from flavor, which includes the olfactory, tactile and thermal attributes of food.

6.1. The taste buds

(Fig. 34–16) consist of clusters of 50–100 secondary sensory cells on the tongue (renewed in 2-week cycles); humans have around 5 000 taste buds.

6.2. Sensory stimuli

Sensory stimuli from the taste buds *are conducted* to endings of the VII, IX and X cranial nerves, relayed by the nucleus tractus solitarii, and converge at a high frequency on (i) the postcentral gyrus via the thalamus and (ii) the hypothalamus and limbic system via the pons (Fig. 34–17).

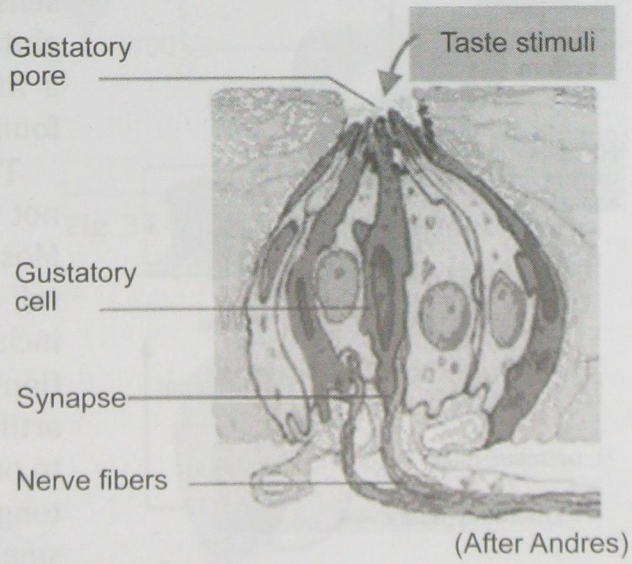


Fig. 34–16. Taste bud.

6.3. Taste modalities

The qualities of taste distinguishable in humans are conven-tionnally defined as sweet, sour, salty and bitter. The specific taste sensor cells for these quailties are distributed over the whole tongue but differ with respect to their densities.

1. *A sweet sensation* is produced by various classes of organic molecules including sugars, glycols, and aldehydes. The tip of the tongue is the area most sensitive to sweet stimuli.
2. *A sour sensation* is produced by acids; this sensation relates, to some degree, to the pH of applied stimulus solutions. The posterior half of each side of the tongue is the area most sensitive to sour stimuli.
3. *A salty sensation* is produced by the anions of ionizable salts. The front half of each side of the tongue is the area most sensitive to salty stimuli.
4. *A bitter sensation* is produced by alkaloids, such as quinine and caffeine. Many alkaloids are harmful when swallowed. The back of the tongue is the area most sensitive to bitter stimuli.

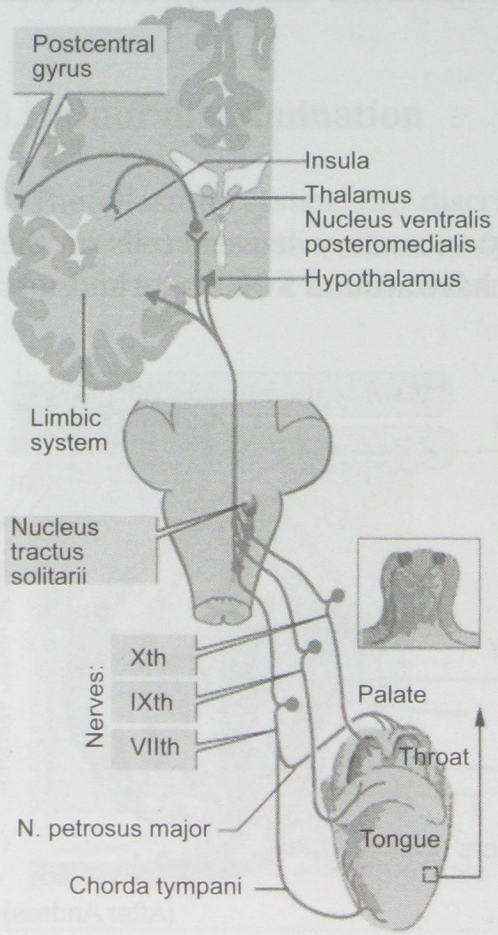


Fig. 34–17. Gustatory pathways.

Umami, or savoriness (a Japanese word meaning "delicious"), — the sensation caused by monosodium-L-glutamate (MSG), is now classified as a fifth quality of taste. MSG is chiefly found in protein-rich foods.

The broad surface of the tongue is not as well supplied with taste buds. Most taste experiences involve several different sensory modalities, including taste, smell, mechanoreception (for texture), and temperature; artificially confining the taste sensation to only the four modalities found on the tongue (e.g., by blocking the sense of smell) greatly diminishes the range of taste perceptions.

6.4. Taste thresholds

The threshold for recognition of taste stimuli applied to the tongue is roughly 10^{-5} for quinine sulfate and saccharin, 10^{-3} for HCl and 10^{-2} for sucrose and NaCl. The concentration of the gustatory stimulus determines whether its taste will be perceived as pleasant or unpleasant (Fig. 34–18).

6.5. Adaptation

Taste sensations adapt rapidly, often almost completely within a minute or so of continuous stimulation. The intensity of a sensation resulting from continuous application of a taste stimulus declines with time until the solution becomes tasteless. The firing rate of afferent nerve from the tongue similarly declines. Gustatory sensation adapts quicker comparing with the olfactory adaptation (Fig. 34–15).

6.6. The sense of taste

The sense of taste has a protective *function* as spoiled or bitter-tasting food (low taste threshold) is often poisonous. Tasting substances also stimulate the secretion of saliva and gastric juices.

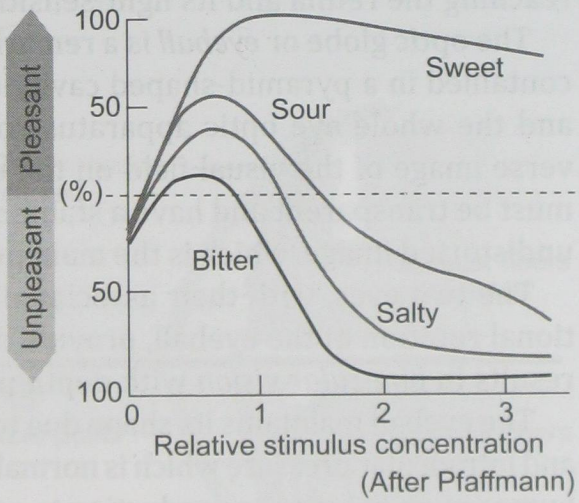


Fig. 34–18. Evaluation of taste stimuli.

Chapter 35. Physiology of vision

The eye is a unique complex sensory organ involving both sensory elements and elaborate accessory structures that process information both before and after it is detected by the light-sensitive cells.

1. THE STRUCTURE OF THE EYE

Light entering the eye must pass through the cornea, aqueous humor, lens and vitreous body, which are collectively called the optical apparatus, before reaching the retina and its light-sensitive photosensors.

The optic globe or *eyeball* is a remarkable, mobile, nearly spherical structure contained in a pyramid-shaped cavity of the skull called the orbit. The eyeball and the whole eye optic apparatus composition produces a reduced and inverse image of the visual field on the retina. Hence, all parts of the apparatus must be transparent and have a stable shape and smooth surface to produce an undistorted image, which is the main purpose of tear fluid in case of the cornea.

The two eyes, with their associated extraocular muscles that permit directional rotation of the eyeball, provide different images of the same object. This results in *binocular vision* with depth perception.

The eyeball maintains its shape due to its tough outer coat or sclera (Fig. 35–1) and intraocular pressure which is normally 10–21 mmHg. The drainage of aqueous humor must balance its production to maintain a constant ocular pressure.

The lens is a remarkable structure that functions to bring images into focus on the retina. The lens is an avascular, transparent, biconvex body, the posterior side of which is more convex than the anterior side. A thin highly elastic lens capsule, which is attached to the surrounding ciliary body by delicate suspensory radial ligaments called *zonules*, holds the lens in place (Fig. 35–1).

The suspensory ligaments and lens capsule normally are under tension, causing the lens to have a flattened shape for distant vision (*far vision*, Fig. 35–2, top). Contraction of the muscle fibers of the ciliary body narrows the diameter of the ciliary body, relaxes the fibers of the suspensory ligaments, and allows the lens to relax to a more spherical or convex shape for *near vision* (Fig. 35–2, bottom). A clear image of the object can be formed if the eye accommodates

for near vision (objects closer than 6 cm).

Accommodation — the eye adjustment for clear vision of objects at different distances — occurs by a reflex that has three components:

1. ***Bulging the lens.*** Contraction of the ciliary muscles releases the tension of the zonular fibers. The elastic capsule surrounding the lens retracts, increasing the convexity (and thus the power) of the lens;
2. ***Pupillary constriction.*** By reducing the area through which light can enter the eye spherical aberration is reduced that follows by depth of the focus increasing (the degree to which the image can form in front of or behind the retina and still appear to be in focus). Nearsighted persons typically squint to increase their depth of focus.
3. ***Eyes convergence*** in order to both eyes were focused on the object.

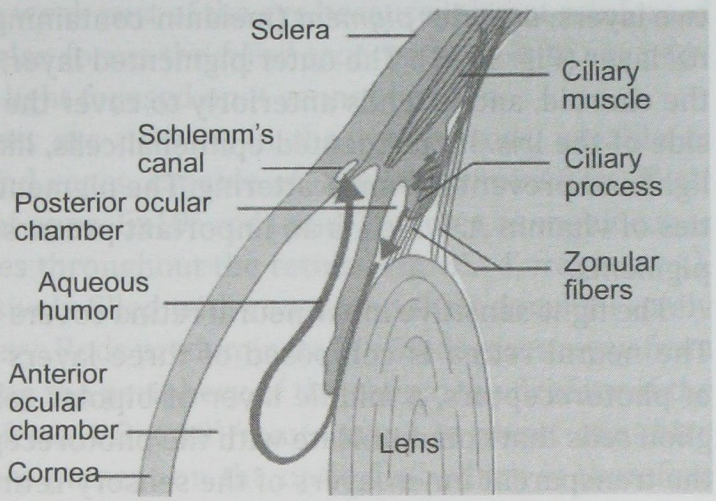


Fig. 35–1. The inflow and drainage of an aqueous liquid.

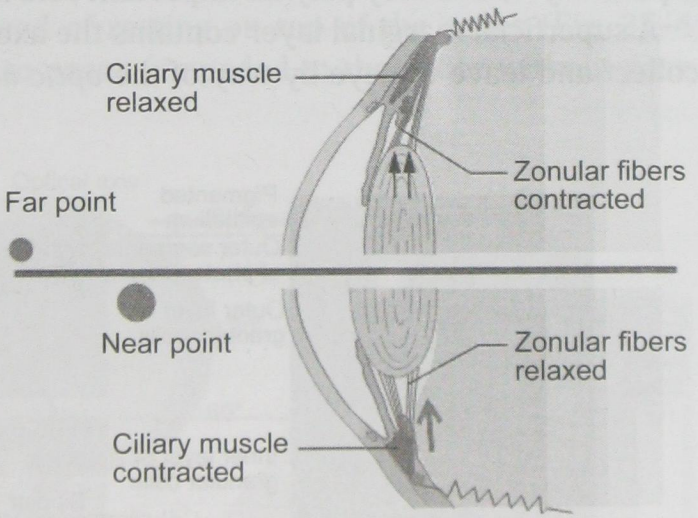


Fig. 35–2. The scheme of accommodation.

The function of the *retina* is to receive visual images, partially analyze them, and transmit this modified information to the brain. The retina is composed of

two layers: *an outer pigment* (melanin-containing) epithelium and an *inner neural* layer (Fig. 35–3). The outer pigmented layer, a single-cell-thick lining abuts the choroid, and extends anteriorly to cover the ciliary body and the posterior side of the iris. Its pigmented epithelial cells, like those of the choroid, absorb light and prevent it from scattering. The pigment layer also stores large quantities of vitamin A, which is an important precursor of the photosensitive visual pigments.

The light-sensitive inner neural retina covers the inner aspect of the eyeball. The neural retina is composed of three layers of neurons: a posterior layer of photoreceptors, a middle layer of bipolar cells, and an inner layer of ganglion cells that communicate with the photoreceptors. Light must pass through the transparent inner layers of the sensory retina before it reaches the photoreceptors. Local currents produced in response to light spread from the photoreceptors to the bipolar neurons and other interneurons and then to the innermost ganglionic cells, where action potentials are generated. The interneurons, which are composed of horizontal and amacrine cells, have cell bodies in the bipolar layer, and they play an important role in modulating retinal function.

A superficial marginal layer contains the axons of the ganglion cells as they collect and leave the eye by way of the *optic nerve*. The optic disk, where the

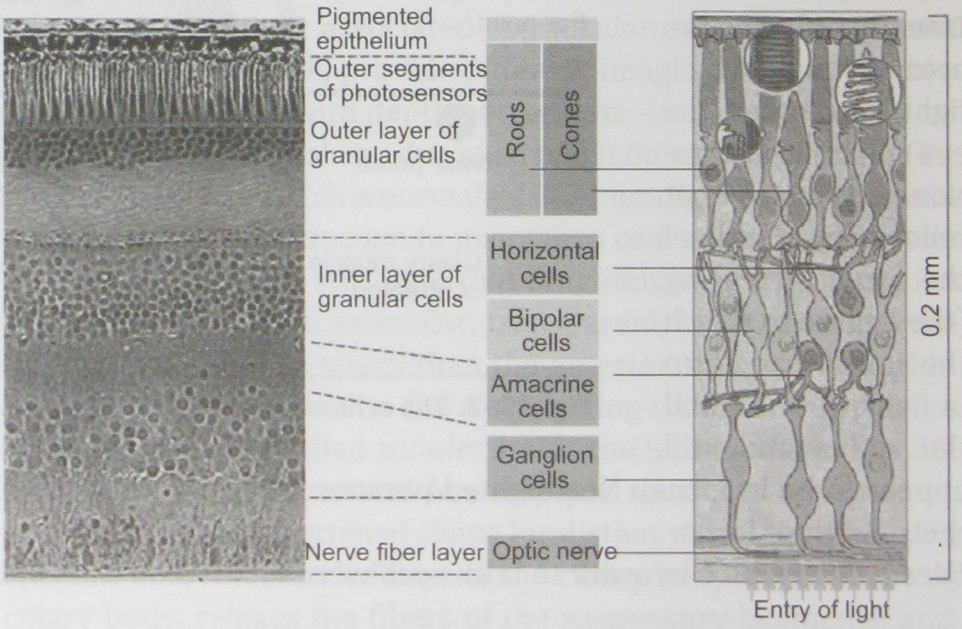


Fig. 35–3. Retina.

optic nerve exits the eye, is the weak part of the eye because it is not reinforced by the sclera. The optic disk also forms the *blind spot*, because it is not reinforced by photoreceptors and light focused on it cannot be seen.

Two types of *photoreceptors* are present in the retina: *rods*, capable of black–white discrimination, and *cones*, capable of color discrimination. Their comparative ratio composes of appr. $6x10^6$ rods and 20 times as many cones distributed at variable densities throughout the retina (Fig. 35–4, fragment 1).

The *fovea centralis* is exclusively filled with cones, and their density rapidly decreases towards the periphery. Rods predominate 20–30 degrees away from the fovea centralis. Approaching the periphery of the retina, the density of the rods decreases continuously from 1.5 x $10^5/mm^2$ (max.) to about one-third this value. No photosensors are present on the optic disk which is therefore referred to as the *blind spot* in the visual field.

An area approximately 1.5 mm in diameter near the center of the retina, called the macula lutea (i.e., "yellow spot"), is especially adapted for acute and detailed vision. This area is composed entirely of cones. In the central portion of the macula, the fovea centralis, the blood vessels and innermost layers are displaced to one side instead of resting on top of the cones (Fig. 35–4, fragment 2). This allows light to pass unimpeded to the cones without pass-

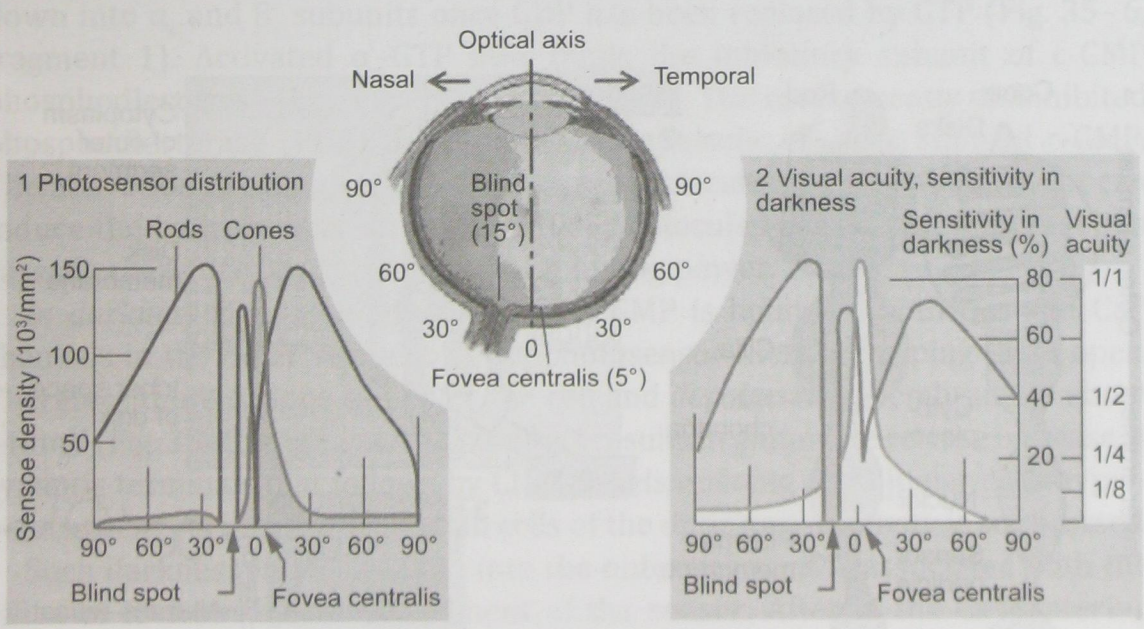


Fig. 35–4. Photosensor distribution, sensitivity in darkness and visual acuity of the retina.

ing through several layers of the retina. Many of these cones are connected one- to- one with ganglion cells, an arrangement that favors high acuity.

Rod-based vision (*scotopic vision*) is particularly sensitive to detecting light, especially moving light stimuli, at the expense of clear pattern discrimination. Rod vision is particularly adapted for night and low-level illumination. Dark adaptation is the process by which rod sensitivity increases to the optimum level. This requires approximately 4 hrs in total or near-total darkness and involves only rods.

Cone receptors, which are selectively sensitive to different wavelengths of light, provide the basis for color vision *(photopic vision)*. Three types of cones, or cone-color systems, respond to the blue, green and red portions of the visible electromagnetic spectrum. Cones do not have the dark adaptation of rods. Consequently, the dark-adapted eye is a rod receptor eye with only black-gray-white experience. The light-adapted eye adds the capacity for color discrimination.

Both rods and cones contain light-absorbing visual pigments and a variety of enzymes and transmitters that decompose on exposure to light and, in the process, generate the currents that lead to the action potentials generated by the ganglionic cells. This process is called *photoelectric transduction*. Both types of photoreceptors are thin, elongated, mitochondria-filled cells with a single, highly modified cilium (Fig. 35–5, fragment 1). The cilium has a short base, or

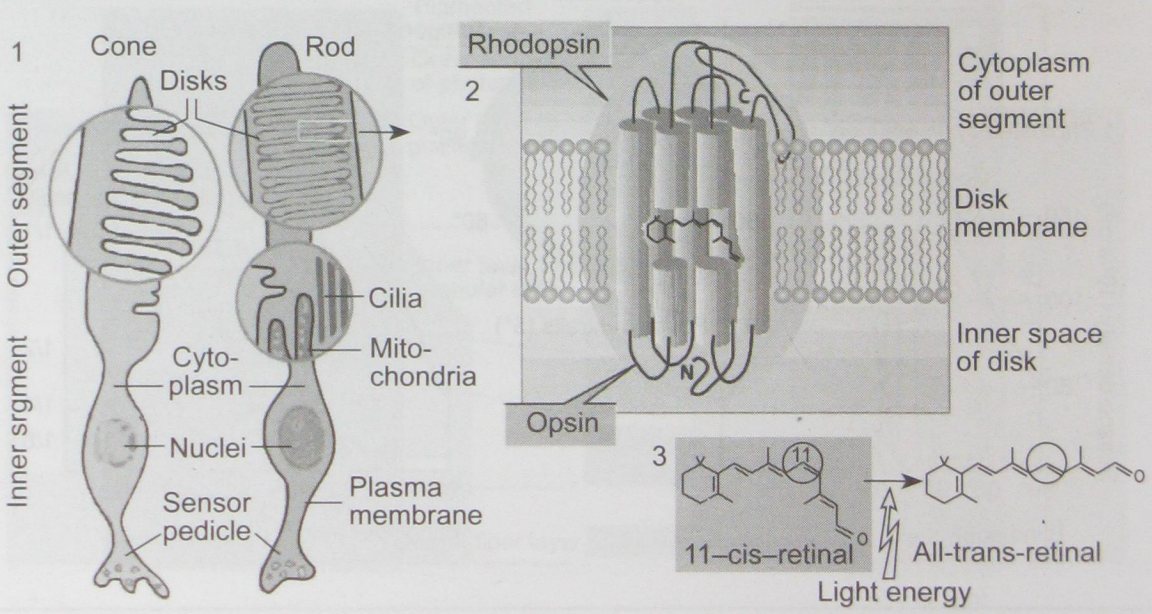


Fig. 35–5. Photosensors.

inner segment, and a highly modified outer segment. The plasma membrane of the outer segment is tightly folded to form membranous disks (rods) or conical shapes (cones) containing visual pigment.

The light-sensitive chemical in the rods is called rhodopsin (Fig. 35–5, fragment 2) and the light-sensitive chemicals in the cones are called cone or color pigments. Both rhodopsin and color pigment are incorporated into membranes of these disks in the form of transmembrane proteins. These disks are continuously synthesized at the base of the outer segment and shed at the distal end. The discarded membranes are phagocytized by the retinal pigment cells. If this phagocytosis is disrupted, as in retinitis pigmentosa, the sensory retina degenerates.

Very important is to discuss the question concerning the chains of *biochemical reactions* in response to light rays penetration through the eye optic system to retina.

Photic stimuli (light) trigger a primary photochemical reaction in rhodopsin in which 11-cis-retinal is converted to all-trans-retinal (Fig. 35–5, fragment 3). Even without continued light energy application the reaction yields bathorhodopsin, the intermediates lumirhodopsin and metarhodopsin I, and finally metarhodopsin II within roughly 10^{-3} s (Fig. 35–6, fragment 1).

Metarhodopsin II reacts with a G_s- protein called *transducin* which breaks down into α_s and β_γ subunits once GDP has been replaced by GTP (Fig. 35– 6, fragment 1). Activated α_s-GTP now binds the inhibitory subunit of c-GMP phosphodiesterase (I_{PDE}, Fig. 35–6, fragment 2). The consequently disinhibited phosphodiesterase (PDE) then lowers the cytosolic concentration of c-GMP. The activation of a single retinal rhodopsin molecule by a quantum of light can induce the hydrolysis of up to 10^6 c-GMP molecules per second. The reaction cascade therefore has tremendous amplifying power.

In darkness (Fig. 35–6, fragment 3), c-GMP is bound to both Na^+ and Ca^{2+} channels in the outer segment of the photosensor thereby keeping them open. Therefore these cations can enter the cell and *depolarize* it membrane to about –40mV (Fig. 35–6, fragment 4). This effect results in glutamate release increase at synaptic terminals that follows by Cl^- channels opening and bipolar cells hyperpolarization. As the result ganglion cells of the optic nerve activity is suppressed.

Such darkness-induced influx into the outer segment is associated with the efflux of K^+ from the inner segment of the sensor. After it, the Ca^{2+} entering the outer segment is immediately transported out of the cell by a $3Na^+/Ca^{2+}$ exchanger, so the cytosolic Ca^{2+} concentration $[Ca^{2+}]_i$ in darkness remains constant at appr. 350–500 nmol/l (Fig. 35–6, fragment 6).

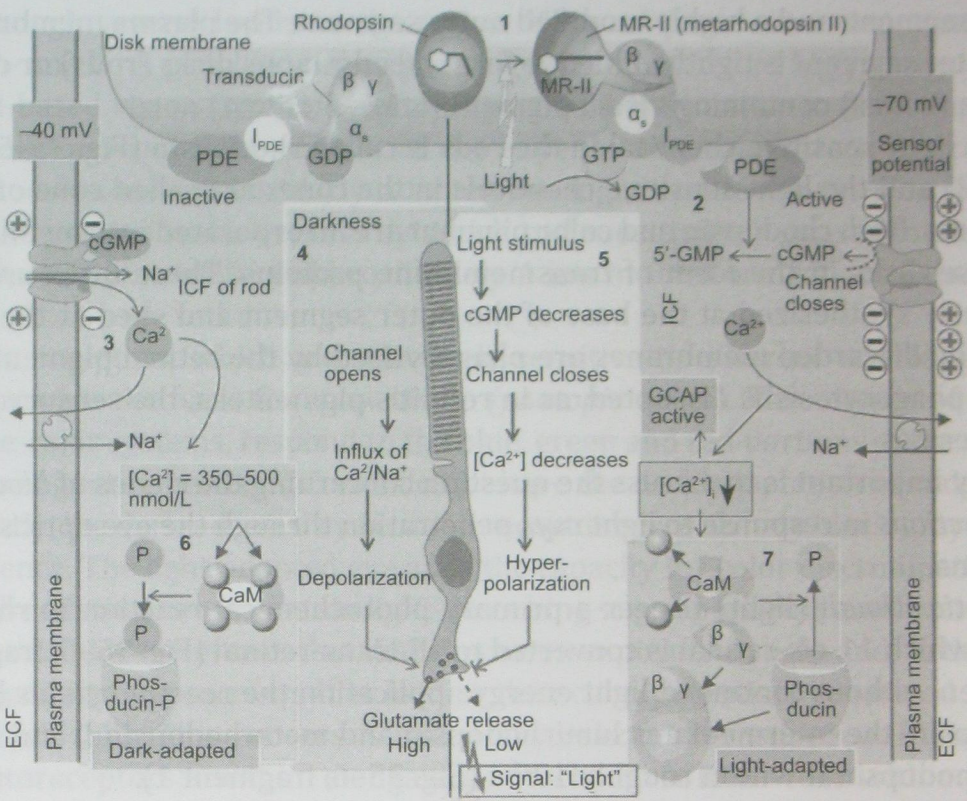


Fig. 35–6. The scheme of chain reaction of photosensors activation in light and in the darkness.

Reverse electro-chemical process occurs in case of the *light stimuli* influence in response to which the cytosolic c-GMP concentration decreases (Fig. 35–6, fragment 5). This results in c-GMP molecule dissociation from the cation channels allowing them to close. The photosensor then *hyperpolarizes* to appr. –70mV (*sensor potential*: Fig. 35–6, fragment, right part). This inhibits the release of glutamate at the sensor pedicle (Fig. 35–6, fragment 5) which subsequently causes bipolar cells membrane depolarization. As the result glutamate release from the bipolar cells increased that follows by the action potentials production in ganglion cells of optic nerve fibers. It should be mentioned that receptor potentials of the rods and bipolar cells are local, graded potentials. It is only in the ganglion cells that propagated action potentials are generated.

All the chains of the electrical reactions in the rods are in the dark and in the light are summarized in the Fig. 35–7.

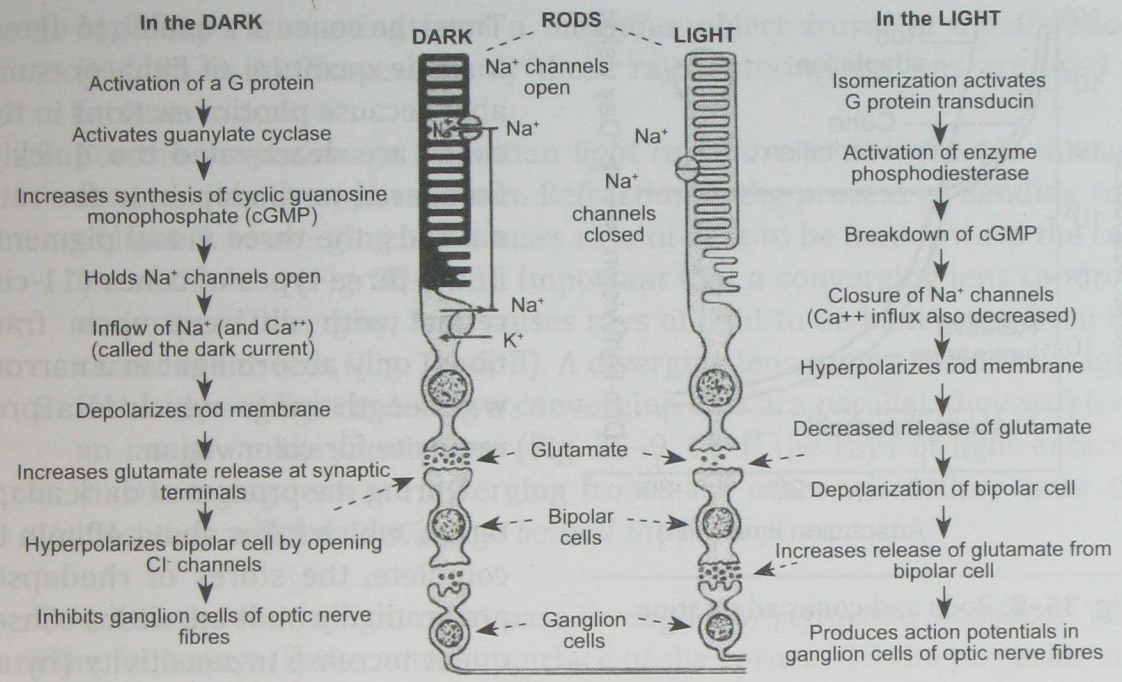


Fig. 35–7. The summarized scheme of the electrical reactions in the rods are in the dark and in the light.

Photic reactions deactivation results in G_s-protein deactivation. The α_s-GTP molecule and β_γ subunit then reunite to *transducin* which regeneration accelerates GAP (GTPase-activating protein). *Phosducin*, another protein, is phosphorylated in the dark (Fig. 35–6, fragment 6) and dephosphorylated in light (Fig. 35–6, fragment 7). The latter form binds to the β_γ subunit thereby blocking the transducin regeneration. This plays an important role in light adaptation.

In the dark, the $[Ca^{2+}]_i$ is high, and calmodulin-bound Ca^{2+} stimulates phosducin phosphorylation with the aid of cAMP and phosphokinase A (Fig. 35–6, fragment 6). In light, the $[Ca^{2+}]_i$ is low; phosducin is dephosphorylated and rapid regeneration of transducin is not possible (Fig. 35–6, fragment 7). Moreover, Ca^{2+} accelerates the phosphorylation of MR-II in light with the aid of another Ca^{2+} binding protein, *recoverin*. Ca^{2+} is therefore essential for the adaptation of photosensors.

The processes in cone cells are similar, although there are three different opsins (with different spectral sensitivities) and the specific transduction mechanism is also different. The overall sensitivity of the transduction process is about 100 times less than that of the rods. In the light, much rhodopsin is in its unconjugated form, and the sensitivity of the rod cell is relatively low.

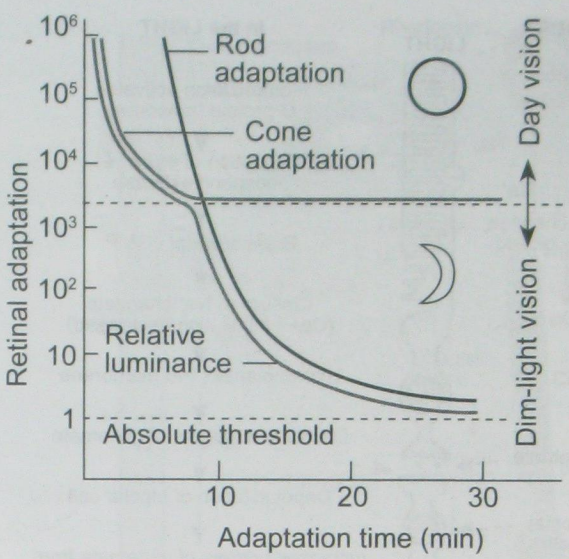


Fig. 35–8. Rods and cones adaptation.

Thus, the cones are unable to detect a single quantum of light, presumably because photic reactions in the cones are deactivated too quickly. Compared to rhodopsin in the retinal rods, the three visual pigments in the three types of cones (11-cis-retinal with different opsin fractions) only absorb light in a narrow wavelength range which is a prerequisite for color vision.

During the process of dark adaptation, which takes about 40 min to complete, the stores of rhodopsin are gradually built up, with a consequent increase in sensitivity (by as much as 25,000 times).

The maximum level of adaptation is reached in about 30 min (Fig. 35–8). The minimum light intensity that can just be detected after maximum dark adaptation is the *absolute visual threshold*.

Cone cells adapt more quickly than rods but their final sensitivity is much lower. The reverse process, light adaptation, takes about 5 min.

2. THE OPTIC SYSTEM OF THE EYE

The adequate stimulus for human visual receptors is light, which may be defined as electromagnetic radiation between the wave lengths of 770 nm (red) and 380 nm (violet). The physics of the vision is based on the simple experiment during which white light passing through a prism is split into a color spectrum ranging from red to violet (colors of the rainbow). The familiar colors of the spectrum all lay between these limits. A wide range of intensities, from a single photon to the direct light of the sun, exists in nature.

In order for an object to be perceived by the visual system, a clear image of the object must be formed on the retina.

Several physical principles govern the process of image formation by a lens system:

1. *Light rays* emanating from a luminous object *travel in all directions* (Fig. 35–9, A). However, only those rays entering the lens are used to form an image.
2. For the proper image formation *light rays have to be refracted* through the complex eye optic system. Refraction is the process of bending rays of light. A converging lens causes rays of light to be bent toward the center of the lens (Fig. 35–6, B). Important that a converging lens can form an image. A *diverging lens* causes rays of light to be bent away from the center of the lens (Fig. 35–6, B). A diverging lens cannot form an image.
3. If the rays of light entering a converging lens are parallel, they will form an image *on the focal plane* (Fig. 35–9, C). If the rays of light entering a converging lens are diverging from each other when they enter the lens, they will form an image *behind the focal plane*.

To reach the retina, light must pass through *four refractive interfaces*. They are the following: (i) the anterior surface of the cornea, (ii) the posterior surface of the cornea, (iii) the anterior surface of the lens, and (iv) the posterior surface of the lens. The greatest amount of refraction of the eye occurs at the anterior surface of the cornea, because the difference between the indices of refraction of air and of the cornea is larger than the difference between the indices of refraction of the other ocular media. The cornea contributes about two-thirds of the eye's refractive power. The amount of refraction contributed by the cornea is fixed. The rest one-third of the eye's refractive power nearly equally delivers through the rest three refractive interfaces.

Light absorption. To absorb light photosensors must be photosensitive. With this aim rods (Fig. 35–5) contain rhodopsin which is responsible for (scotopic, black-and-white) night vision. Rhodopsin absorbs light at wavelengths of appr. 400–600 nm, the max. absorption value equals to 500 nm. Relatively speaking, greenish blue

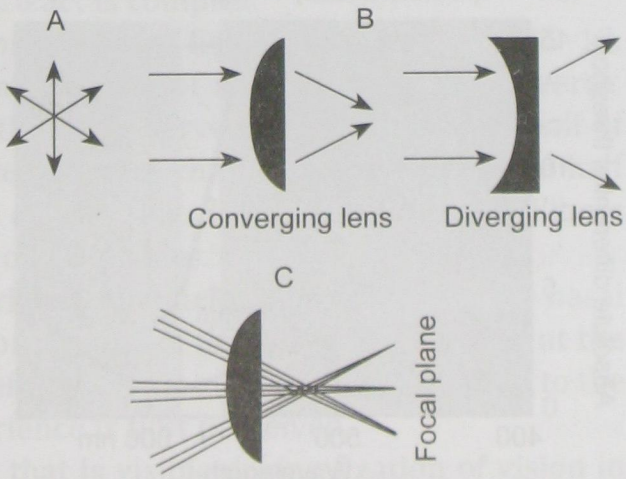


Fig. 35–9. The scheme illustrating some fundamental optic principles.

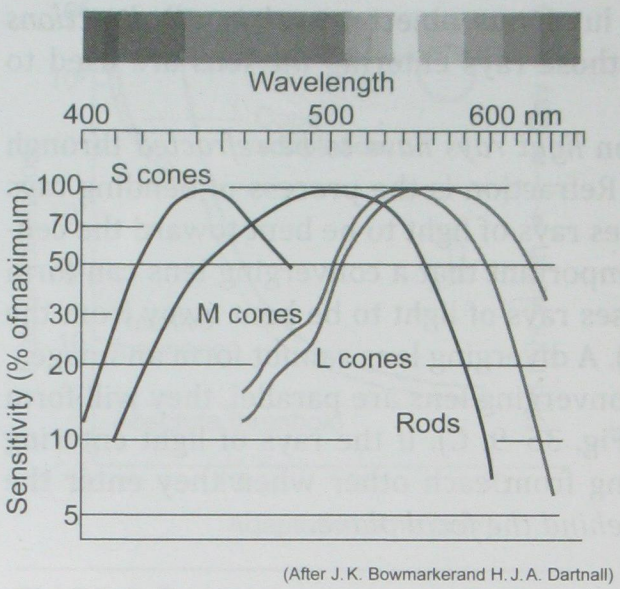


Fig. 35–10. The scheme illustrating rods and cones physiological sensitivity.

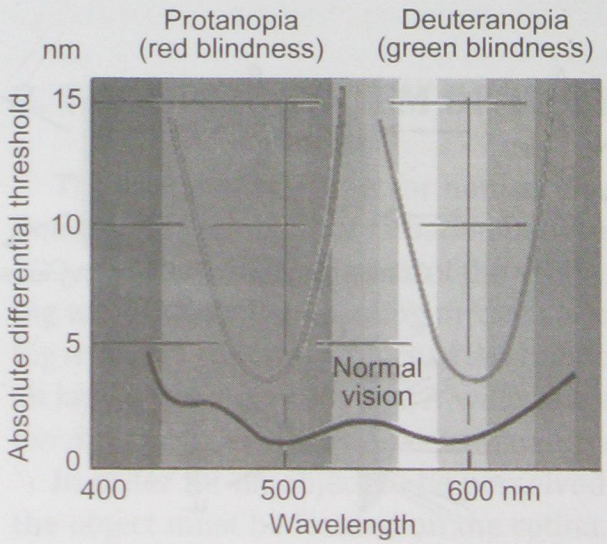


Fig. 35–11. Differential threshold for wavelength.

light therefore appears brightest and red appears darkest at night. Wearing red glasses in daylight therefore leaves the rods adapted for darkness.

Three types of color-sensitive cones are responsible for day vision (photopic):

- S cones, which absorb short-wave (S) blue-violet light (λ_{max}=420 nm);
- M cones, which absorb medium-wave (M) blue-green to yellow light (λ_{max}=535 nm);
- L cones, which absorb long- wave (L) yellow to red light (λ_{max}=565 nm; Fig. 35–10).

This is the basis for the *theory of opponent colours* (after Hering) explanation of colour vision.

Sensory information relayed by the three types of cones (peripheral application of the *trichromatic theory of color vision* after Lomonosov, Young and Helmholtz) and transduction of these visual impulses to brightness and opponent color channels in the retina and lateral geniculate body enables the visual cortex to recognize different types of colors. The human eye can distinguish 200 shades of color and different degrees of color saturation. The *absolute differential threshold* for color vision is 1–2 nm (Fig. 35–11).

Color perception is more complex. White paper, for example, will look white in white light (sunlight), yellow light (light bulb) and red light. We also do not perceive the different shades of color in a house that is partially illuminated by sunlight and partially in the shade. This *color constancy* is the result of retinal and central processing of the retinal signal.

3. NEURAL PATHWAYS

Full visual function requires the normally developed brain-related functions of photoreception and the pupillary reflex. These functions depend on the integrity of all visual pathways, including retinal circuitry and the pathway from the optic nerve to the visual cortex and other visual regions of the brain and brain stem.

Visual information is carried to the brain by axons of the retinal ganglion cells, which form the optic nerve. The two optic nerves meet and fuse in the optic chiasm, beyond which they are continued as the optic tracts (Fig. 35–12). In the optic chiasm, axons from the nasal retina of each eye cross to the opposite side and join with the axons of the temporal retina of the contralateral eye to form the optic tracts. One optic tract contains fibers from both eyes that transmit information from the same visual field.

Fibers of the optic tracts synapse in the lateral geniculate nucleus (LGN) of the thalamus. Axons from these neurons in the LGN form the optic radiations to the primary visual cortex in the occipital lobe. The pattern of information transmission established in the optic tract is complex.

Objects located in the nasal half of the visual field of each eye (Fig. 35–12, lateral parts) are imaged in the temporal half of each retina and vice versa. Along the visual pathway, fibers of the optic nerve from the temporal half of each retina remain on the same side, whereas the fibers from the nasal half of each retina decussate at the optic chiasm (Fig. 35–12, medial parts). Fibers from the fovea centralis are present on both sides.

For example, the axons from the right visual field, represented by the nasal retina of the right eye and the temporal retina of the left eye, are united at the chiasm. They continue through the left optic tract and left optic radiation to the left visual cortex, where visual experience is first perceived.

The *visual field* refers to the area that is visible during fixation of vision in one direction. As with a camera, the simple lens system of the eye inverts the image of the external world on each retina. In addition, the right and left sides of the visual field also are reversed. The right binocular visual field (the na-

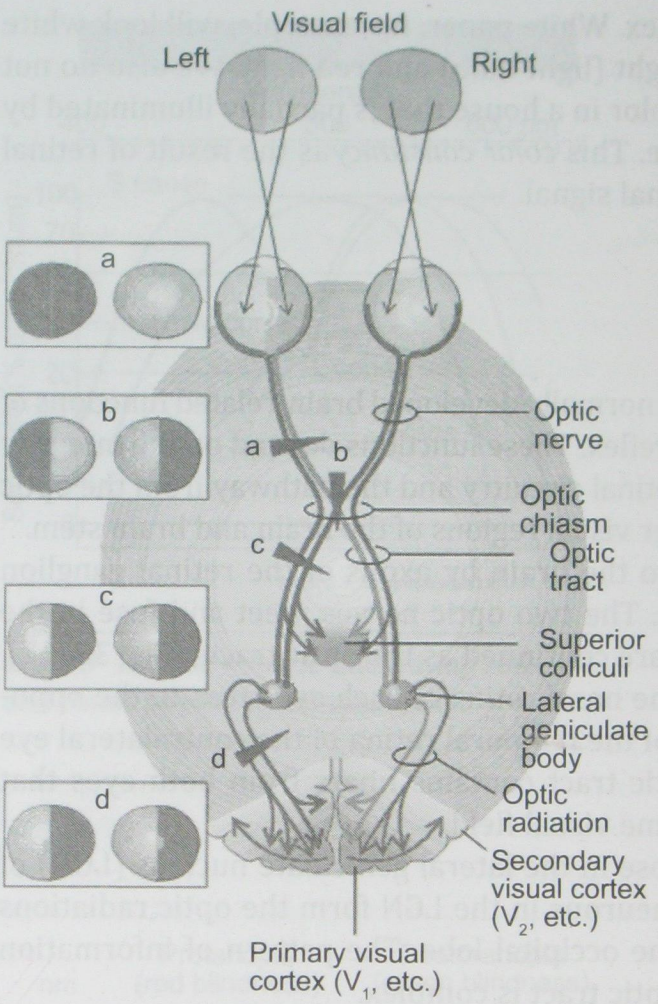


Fig. 35–12. The scheme of the visual pathway and visual field deficits.

sal half of the right eye and the temporal half of the left eye) is seen by the left retinal halves of each eye. Likewise, the left binocular field is seen by the right retinal halves of each eye.

Most of the visual field is binocular, or seen by both eyes. This binocular field is subdivided into central and peripheral portions. Central portions of the retina provide high visual acuity and correspond to the field focused on the central fovea. The peripheral and surrounding portion provides the capacity to detect objects, particularly moving objects. Beyond the visual field shared by both eyes, the left lateral periphery of the visual field is seen exclusively by the left nasal retina, and the right peripheral field by the right nasal retina.

Visual field defects result from damage to the visual pathways or the visual cortex. The testing of visual fields of each eye and of the two eyes together are useful in localizing lesions affecting the system. Perimetry or visual field testing, in which the visual field of each eye is measured and plotted in an arc, is used to identify defects and determine the location of lesions.

Lesions of the left optic nerve for instance cause deficits in the entire left visual field (blindness in one eye — *anopia*, Fig. 35–12, a), whereas lesions of the left optic tract produce deficits in the right halves of both visual fields (*hemianopia*, Fig. 35–12, c). Damage of the median optic chiasm

results in bilateral temporal deficits, i.e., bitemporal hemianopia (Fig. 35–12, b). Occipital lesions may spare the fibers from the macula (Fig. 35–12, d) because of the separation in the brain of these fibers from the others subservient vision.

Axons of the optic tract also project to subcortical regions of the brain such as the pretectal region which regulates the diameter of the pupils; the superior colliculi (Fig. 35–12), which are involved in oculomotor function; the hypothalamus, which is responsible for circadian rhythms.

The left primary visual cortex receives two representations of the right visual field. Physical separation of information from the left and right visual fields is maintained in the visual cortex. Interaction between these disparate representations occurs and provides the basis for the sensation of depth in the near visual field.

The primary visual cortex (area 17) surrounds the calcarine fissure, which lies in the occipital lobe. It is at this level that visual sensation is first experienced.

Immediately surrounding area 17 are the visual association cortices (areas 18 and 19) and several other association cortices. These association cortices with their thalamic nuclei must be functional for added meaningfulness of visual perception. This higher-order aspect of the visual experience depends on previous learning.

Chapter 36. Physiology of auditory and vestibular sensory systems

Like the visual system, the auditory system permits analysis of energy (in this case, sound energy) emitted or reflected from objects even when they are located at a considerable distance from the body.

The ears are paired organs consisting of an external and middle ear, which function in capturing, transmitting, and amplifying sound, and an inner ear that contains the receptive organs that are stimulated by sound waves (e. g., hearing) or head position and movement (e. g., vestibular function).

The auditory system also provides a mechanism for object recognition. Although a number of animal species have much more acute hearing, humans exhibit a more highly developed sense of auditory recognition involving extensive cortical circuits to interpret the complex vocalizations required for spoken communication. Unfortunately, much more is understood about the mechanical transduction of sound waves and auditory acuity than is currently known about higher cortical functions responsible for auditory recognition.

1. SOUND TRANSMISSION THROUGH THE MIDDLE AND INNER EARS

Sound waves are transmitted to the organ of hearing via the external ear and the auditory canal, which terminates at the tympanic membrane or eardrum. The sound waves are conducted through the air (air conduction) and set the eardrum in vibration. These are transmitted via the auditory ossicles of the tympanic cavity (middle ear) to the membrane of the oval window (Fig. 36-1), where the internal or inner ear (labyrinth) begins.

The external ear is a funnel-shaped structure that conducts sound waves to the tympanic membrane. It consists of the auricle, the external acoustic meatus, and the lateral surface of the tympanic membrane (Fig. 36–1). A thin layer of skin containing fine hairs, sebaceous glands, and ceruminous glands lines the ear canal. Ceruminous glands secrete cerumen, or earwax, which has certain antimicrobial properties and is thought to serve a protective function.

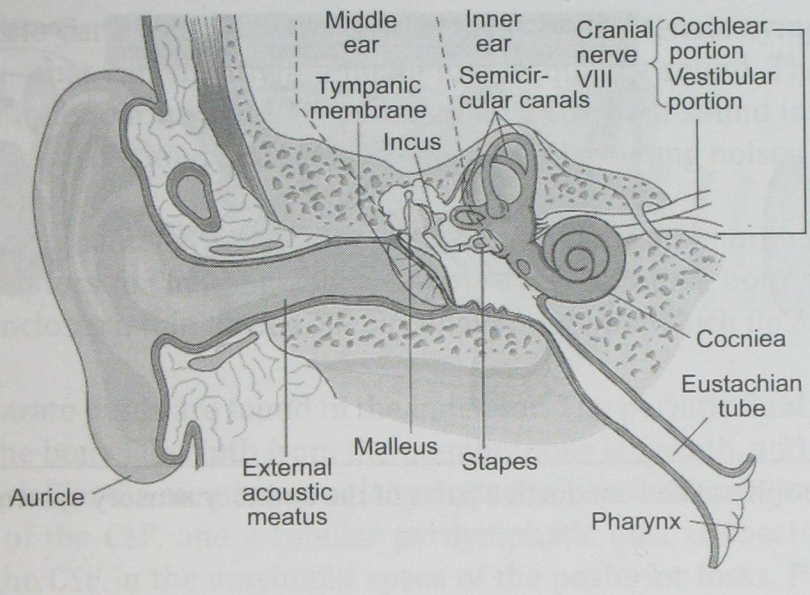


Fig. 36–1. Outer, middle and internal subdivisions of the ear.

The middle ear is a small air-filled cavity located with the petrous (stony) portion of the middle ear. Its lateral wall is formed by the tympanic membrane and its medial wall by the bone dividing the middle and inner ear (Fig. 36–1). Posteriorly, the middle ear is connected with small air pockets in the temporal bone called mastoid air spaces or cells.

Three tiny bones, the auditory ossicles (malleus, incus and stapes) are suspended from the roof of the middle ear cavity and connect the tympanic membrane with the oval window (Fig. 36–1). They are connected by synovial joints and are covered with the epithelial lining of the cavity.

Sound waves are transmitted to the organ of hearing via the external ear and the auditory canal, which terminates at the tympanic membrane or eardrum. The sound waves are conducted through the air (air conduction) and set the eardrum in vibration. These are transmitted via the auditory ossicles of the tympanic cavity (middle ear) to the membrane of the oval window (Fig. 36–2), where the internal or inner ear (labyrinth) begins.

The malleus ("hammer") has its handle firmly fixed to the upper portion of the tympanic membrane. The head of the malleus articulates with the incus ("anvil"), which articulates with the stapes ("stirrup"), which is inserted and sealed into the oval window by an annular ligament. Arrangement of the ear ossicles is such that their lever movements transmit vibrations from the tym-

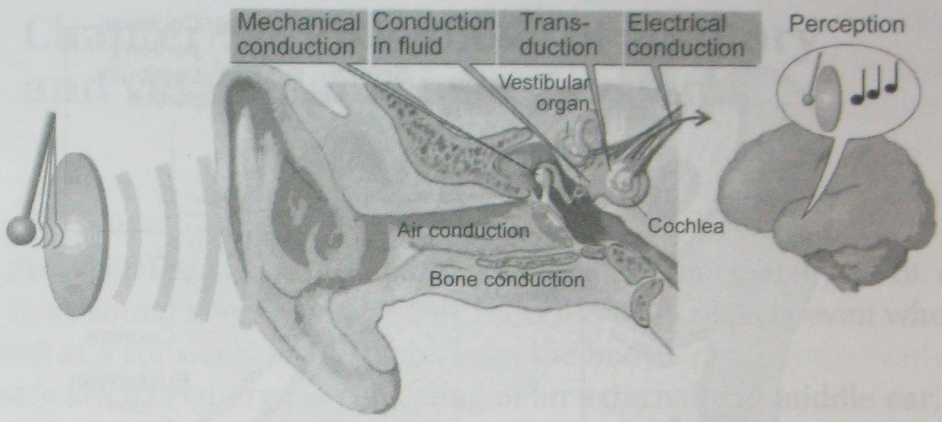


Fig. 36–2. Peripheral and conductive parts of the auditory sensory system.

panic membrane to the oval window and from there to the fluid in the inner ear.

Two tissue-covered openings in the medial wall, the oval and the round windows, provide for the transmission of sound waves between the air-filled middle ear and the fluid-filled inner ear. It is the piston-like action of the stapes footplate that sets up compression waves in the inner ear fluid.

The middle ear is connected to the nasopharynx by the eustachian or auditory tube, which is located in a gap in the bone between the anterior and medial walls of the middle ear (Fig. 36–1). Opening of the eustachian tube, which normally occurs with swallowing and yawning reflexes, provides the mechanism for equalizing the pressure of the middle ear with that of the atmosphere. This equalization ensures that the pressures on both sides of the tympanic membrane are the same, so that sound transmission is not reduced and rupture does not result from sudden changes in external pressure, as occurs during plane travel.

In the middle ear, the malleus, incus and stapes conduct the vibrations of the tympanic membrane to the oval window. Their job is to conduct the sound from the low wave resistance/impedance in air to the high resistance in fluid with as little loss of energy as possible. This *impedance transformation* is based on a 22-fold pressure amplification [(i) tympanic membrane area/oval window area is 17:1, and (ii) leverage arm action of the auditory ossicles amplifies force by a factor of 1.3]. Impairment of impedance transforming capacity due, e.g., to destruction of the ossicles, causes roughly 20 dB of hearing loss (*conduction deafness*).

The middle ear contains two small muscles — the tensor tympani and the stapedius — that can slightly attenuate low-frequency sound. The main functions of the inner ear muscles are to maintain a constant sound intensity level, protect the ear from loud sounds and to reduce distracting noises produced by the listener.

The inner ear contains a labyrinth, or system of intercommunicating channels, and the receptors for hearing and position sense. An outer bony wall, or *bony labyrinth*, encloses a thin walled, *membranous labyrinth* which floats in the bony labyrinth.

Two separate fluids are found in the inner ear. The periotic fluid or *perilymph* separates the bony labyrinth from the membranous labyrinth, and the otic fluid or *endolymph* fills the membranous labyrinth. Periotic fluid composition is similar to that of the CSF, and a tubular perilymphatic duct connects the periotic fluid with the CSF in the arachnoid space of the posterior fossa. Perilymph circulates in Corti's tunnel and Nuel's spaces (Fig. 36–3). Otic fluid has a K^+ content that is similar to that of ICF.

The inner ear consists of the equilibrium organ and the *cochlea*, a spiraling bony tube that is 3–4 cm in length like a snail shell that winds around a central bone column called the modiolus (Fig. 36–4). Inside the cochlea is an endolymph-filled duct called the *scala media* (cochlear duct); the ductus reuniens connects the base of the cochlear duct to the endolymph-filled part of

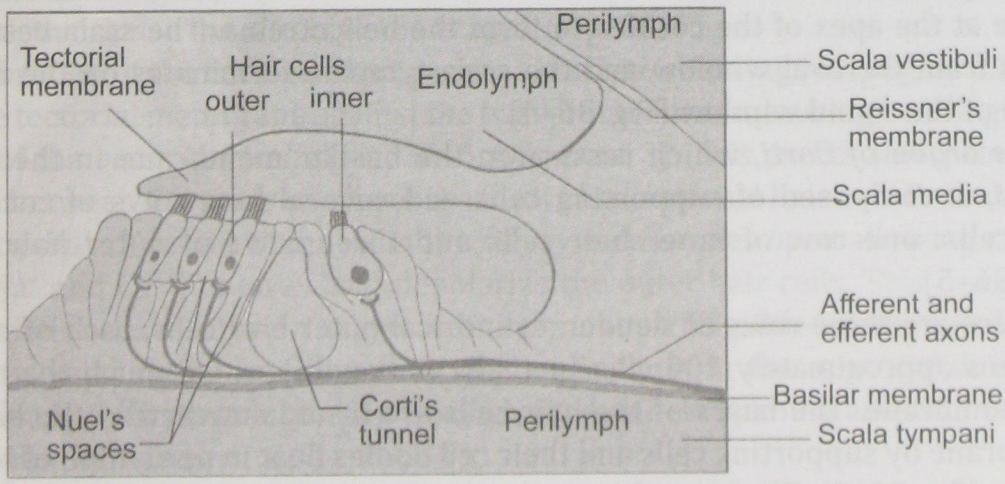


Fig. 36–3. Organ of Corti has been removed from the cochlear duct and greatly enlarged to show the inner and outer hair cells, the basilar membrane and cochlear nerve fibers.

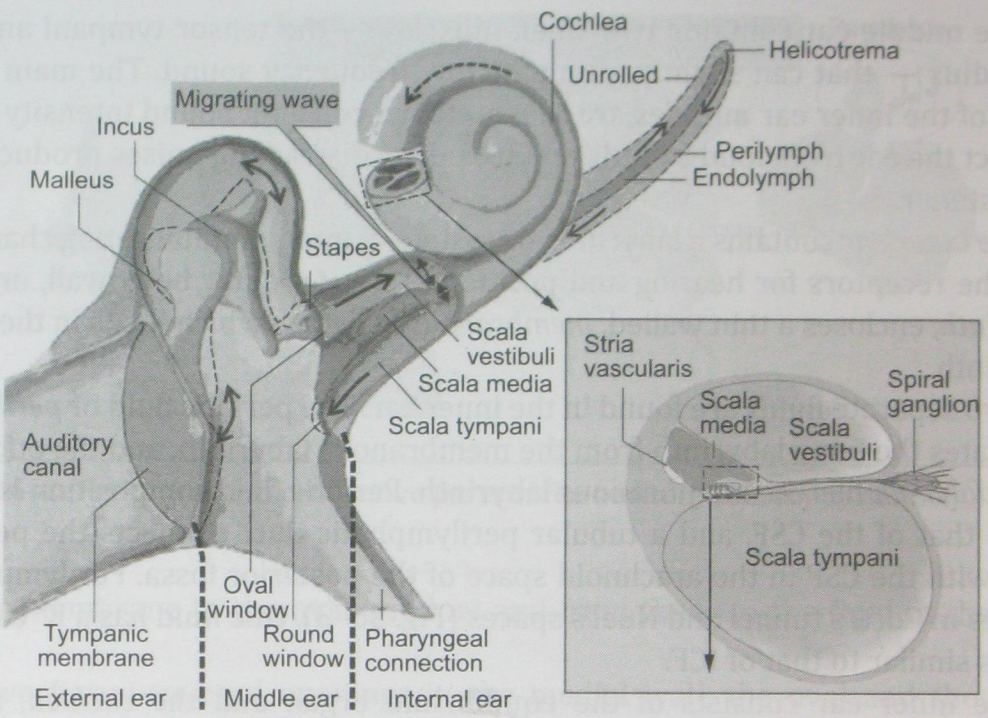


Fig. 36–4. Path taken by sound waves reaching the inner ear.

the equilibrium organ. The scala media is accompanied on either side by two perilymph-filled cavities: *the scala vestibuli* and *scala tympani*. These cavities merge at the apex of the cochlea to form the helicotrema. The scala vestibuli arises from the oval window and the scala tympani terminates on the membrane of the round window (Fig. 36–4).

The organ of Corti, which rests atop the basilar membrane in the scala media, is composed of supporting cells and several long rows of cochlear hair cells: one row of inner hair cells and three rows of outer hair cells (Fig. 36–5).

There are three rows of slender, cylindrical outer hair cells, each of which contains approximately 100 cilia (actually microvilli) which touch the tectorial membrane. The bases of the hair cells are firmly attached to the basilar membrane by supporting cells and their cell bodies float in perilymph of Nuel's spaces (Fig. 36–3). The outer hair cells are principally innervated by efferent, mostly cholinergic neurons from the spiral ganglion. The inner hair cells are pear-shaped and completely surrounded by supporting cells. Their cilia project freely into the endolymph. The inner hair cells are arranged in a single row and

synapse with over 90 % of the afferent fibers of the spiral ganglion. Efferent axons from the n. olivaris superior lateralis synapse with the afferent endings.

Sound waves, delivered to the oval window by the stapes footplate, are transmitted via the periotic fluid (perilymph) the round window (Fig. 36–4).

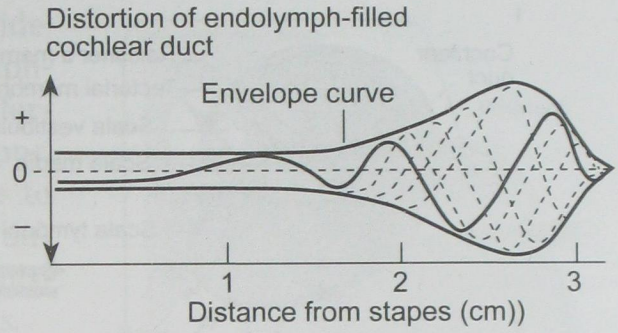


Fig. 36–5. The scheme of the migrating wave in cochlea.

The walls of the endolymph-filled cochlear duct, i.e. Reissner's membrane and the basilar membrane give against the pressure wave *(migrating wave*, Fig. 36–5).

It can therefore take a "short cut" to reach the round window without crossing the helicotrema. Since the cochlear duct is deformed in waves, Reissner's membrane and the basilar membrane vibrate alternately towards the scala vestibuli and scala tympani (Fig. 36–6, fragments 1, 2).

The velocity and wavelength of the migrating wave that started at the oval window decrease continuously while their amplitude increases to a maximum and then quickly subsides (Fig. 36–5, *an envelope curve*).

The site of the maximum excursion of the cochlear duct is characteristic of the wavelength of the stimulating sound. The higher the frequency of the sound, the closer the site is to the stapes (Fig. 36–7).

Vibration of the cochlear duct causes a discrete shearing (of roughly 0.3 nm) of the tectorial membrane against the basilar membrane, causing bending of the cilia of the *outer hair cells* (Fig. 36–6, fragment 3). This exerts also a shearing force between the rows of cilia of the individual external hair cell. Probably via the "tip links", cation channels in the ciliary membranes open allowing cations (K^+, Na^+ and Ca^{2+}) to enter and depolarize the outer hair cells. This causes the outer hair cells to shorten in sync with stimulation (Fig. 36–6, fragment 3). The successive shearing force on the cilia bends them in the opposite direction. This leads to hyperpolarization through K^+ channels opening and extension of the outer hair cells.

These outer hair cell electromotility contributes to the cochlear amplification (appr. 100-fold or 40 dB amplification) which occurs before sound waves reach the actual sound sensors, i.e. inner hair cells. The electromotility causes endolymph waves in the subtectorial space which exert *shearing forces* on the

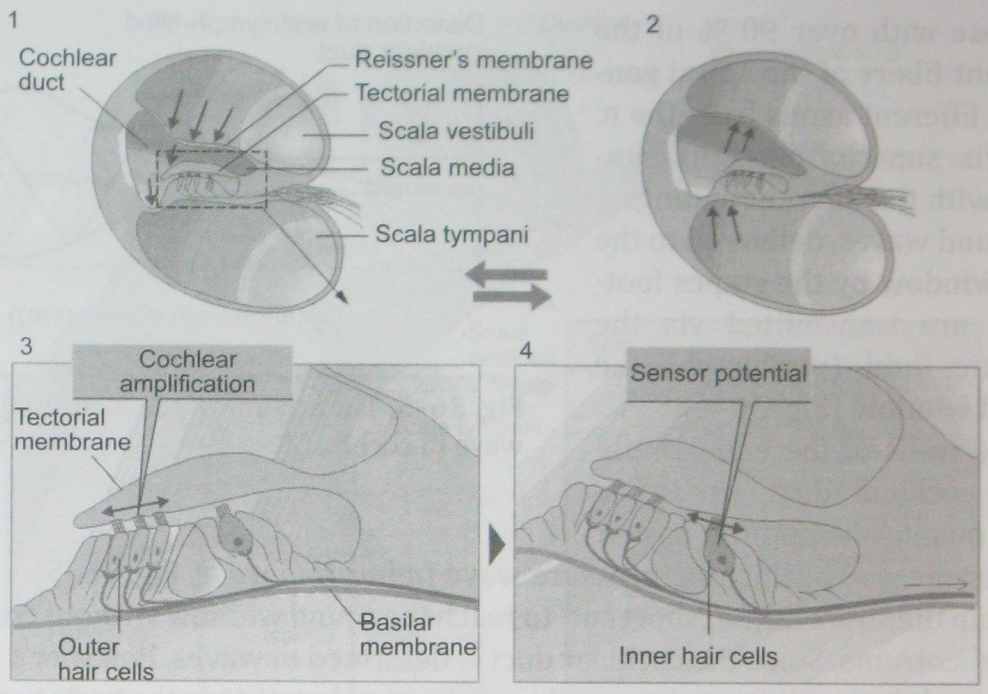


Fig. 36–6. The scheme showing how the up-and-down movement of the basilar and tectorial membrane causes the stereocilia extending from the hair cells to bend back and forth.

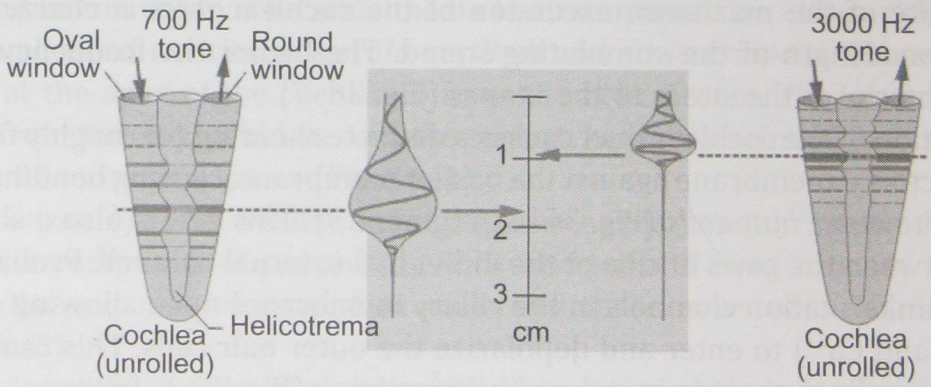


Fig. 36–7. Frequency-dependent mechanical events in cochlea.

inner hair cell cilia at the site of maximum reaction to the sound frequency (Fig. 36–6, fragment 4), resulting in opening of transduction channels and depolarization of the cells. This, in turn, leads to transmitter release by internal hair cells and the subsequent conduction of impulses to the CNS.

Inner ear potentials. On the cilia side, the hair cells border with the endolymph-filled space, which has a potential difference (*endocochlear potential*) of appr. from +80 mV till +110 mV relative to perilymph (Fig. 36–8). This potential difference is maintained by an active transport mechanism in the stria vascularis.

Since the cell potential of outer (inner) hair cells is –70 mV (–40 mV), a potential difference of roughly 150–180 mV (120–150 mV) prevails across the cell membrane occupied by cilia. Since the K^+ levels both in the endolymph and hair cells are roughly equal (≈140 mmol/l), the prevailing *K^+ equilibrium potential* equals to 0 mV.

These high potentials provide the driving forces for the influx not only of Ca^{2+} and Na^+ but also of K^+ prerequisites for provoking the sensor potential.

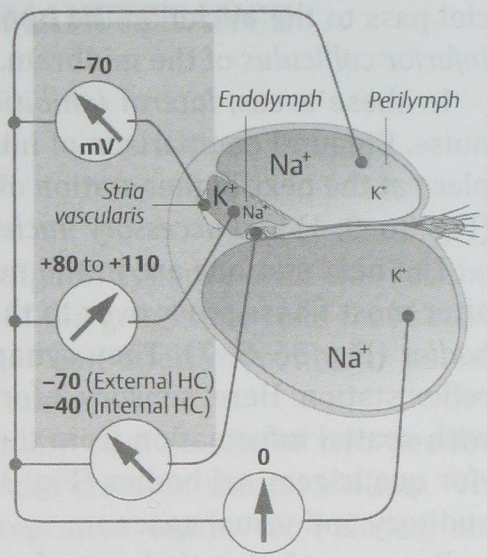


Fig. 36–8. Electrical potentials in the cochlea together with electrolyte distribution.

2. AUDITORY PATHWAY

The auditory nerve fibers with somata positioned in the spiral ganglion of the cochlea (organ of Corti) in the *vestibulocochlear or auditory nerve* project from the *cochlea* (Fig. 36–9, 1) to the *anterolateral* (Fig. 36–9, 2), *posteroventral* and *dorsal cochlear nuclei* in the caudal pons (Fig. 36–9, 3). Afferents in these three nuclei are arranged according to tone frequency at different levels of complexity. Many secondary nerve fibers from the cochlear nu-

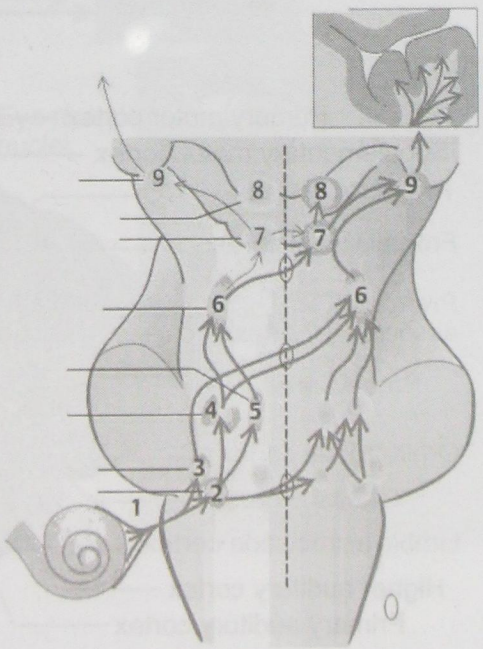


Fig. 36–9. Afferent auditory pathway.

clei pass to the nuclei on the opposite side of the pons or rostrally toward the *inferior colliculus* of the midbrain.

In these areas, *lateral inhibition* (see Chapter # 34, Fig. 34–3) suppresses noise. Binaural comparison of intensity and transit time of sound waves takes place at the next-higher station of the auditory pathway — in the *superior olive* (Fig. 36–9, 4) and *accessory nucleus* (Fig. 36–9, 5).

The next stations are in the nucleus of *lateral lemniscus* (Fig. 36–9, 6) and, after most fibers cross over to the opposite side — the *inferior quadrigeminal bodies* (Fig. 36–9, 7). They synapse with numerous afferents and serve as a reflex station. Here, sensory information from the cochlear nuclei is compared with spatial information from the superior olive. Via connections to the superior quadrigeminal bodies (Fig. 36–9, 8), they also ensure coordination of the auditory and visual space.

By way of the *thalamus* [*medial geniculate body* (MGB), Fig. 36–9, 9] the afferents ultimately reach the *primary auditory cortex* in the temporal lobe (Fig. 36–9, 10; Brodmann's area 41) and the surrounding *secondary auditory areas* (areas 42 and 22, Fig. 36–10), which is necessary for the meaningfulness of sound, borders the primary cortex. Because some of the fibers from each ear cross, each auditory cortex receives impulses from both ears.

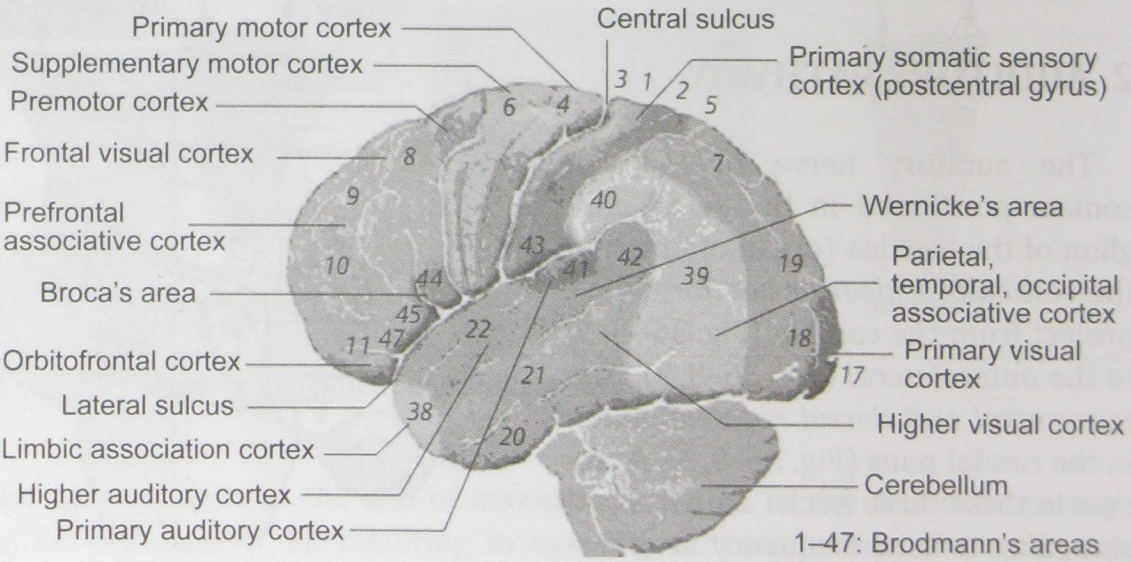


Fig. 36–10. Areas of the cortex according to Brodmann's division.

3. THE VESTIBULAR SYSTEM AND MAINTENANCE OF EQUILIBRIUM

3.1. Vestibular pathways

The bipolar neurons of the vestibular ganglion synapse with the vestibular nuclei (Fig. 36–11). The nerve fibers from the ganglionic cells that supply the vestibular apparatus become the superior and inferior vestibular nerves, which become part of the vestibulocochlear nerve.

Impulses from the vestibular nerves initially pass to one of two destinations — the vestibular nuclear complex in the stem or the cerebellum. The vestibular nuclei, which form the main integrative center for balance, also receives input from visual and somatic receptors, particularly from proprioceptors in the neck muscles that report the angle or inclination of the head. The vestibular nuclei integrate this information and then send impulses to the brain stem centers that control the extrinsic eye movements (III, IV and VI cranial nerves) and reflex movements of the neck, limb, and trunk muscles

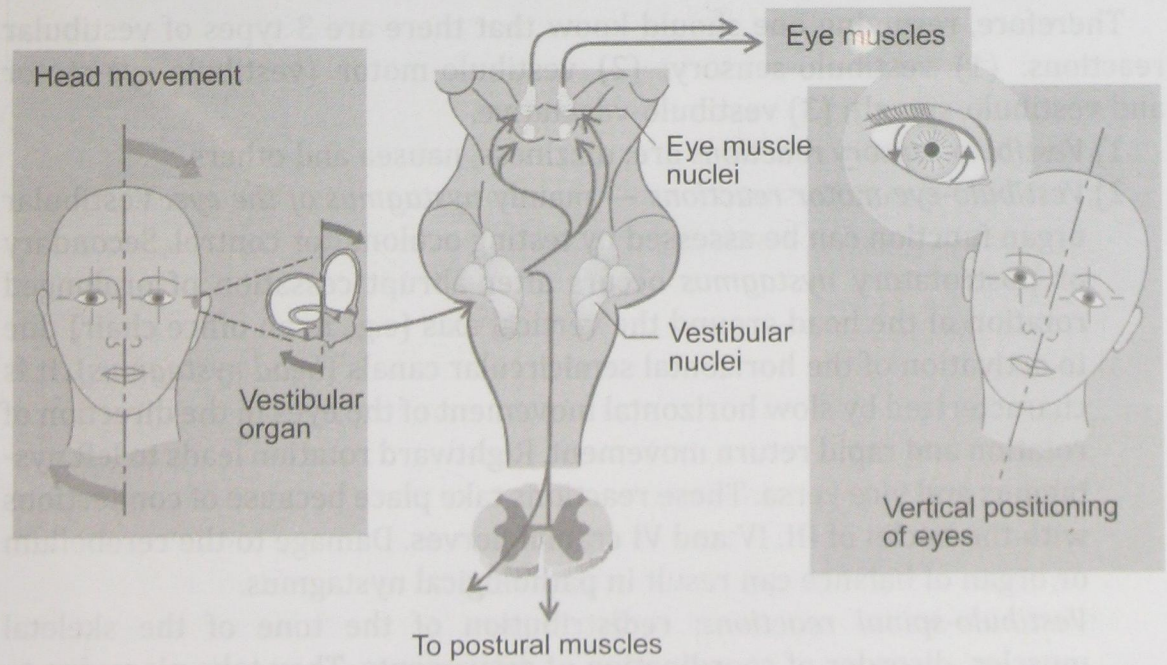


Fig. 36–11. The vestibular pathways. The scheme of oculomotor control.

(via the vestibulospinal tracts) and to the postcentral gyrus (conscious spatial orientation).

Vestibular reflexes have to (i) maintain the balance of the body (*postural motor*, Fig. 36–11, left side) and (ii) keep the visual field in focus despite changes in head and body position (*oculomotor control*, Fig. 36–11, right side).

Since the vestibular organ cannot determine whether the head alone or the entire body moves (sense of movement) or changed position (postural sense), the vestibular nuclei must also receive and process visual information and that from propriosensors in the neck muscles. Efferent fibers project bilaterally to the eye muscle nuclei, and any change in head position is immediately corrected by opposing eye movement (Fig. 36–11, right side). This vestibulo-ocular reflex maintains spatial orientation.

Neurons of the vestibular nuclei also project to the thalamus, the temporal cortex, the somesthetic area of the parietal cortex, and the chemoreceptor trigger zone. The thalamic and cortical projections provide the basis for the subjective experiences of position in space and of rotation. Connections with the chemoreceptor trigger zone stimulate the vomiting center in the brain. This account for the nausea and vomiting that often are associated with vestibular disorders.

Therefore, resuming one should know that there are 3 types of vestibular reactions: (1) vestibulo-sensory; (2) vestibulo-motor (vestibulo-eye-motor and vestibulo-spinal); (3) vestibulo-vegetative.

1) *Vestibulo-sensory reactions* are: dizziness, nausea and others.
2) *Vestibulo-eye-motor reactions* — mainly *nystagmus of the eye*. Vestibular organ function can be assessed by testing oculomotor control. Secondary or postrotatory *nystagmus* occurs after abrupt cessation of prolonged rotation of the head around the vertical axis (e.g., in an office chair) due to activation of the horizontal semicircular canals (*head nystagmus)*. It is characterized by slow horizontal movement of the eyes in the direction of rotation and rapid return movement. Rightward rotation leads to left nystagmus and vice versa. These reactions take place because of connections with the nuclei of III, IV and VI cranial nerves. Damage to the cerebellum or organ of balance can result in pathological nystagmus.

   Vestibulo-spinal reactions: redistribution of the tone of the skeletal muscles, disorder of coordination of movements. They take place due to connections of the vestibular nuclei with the spinal cord and with the cerebellum.

3) *Vestibulo-vegetative reactions*: change of the cardiac rhythm, constriction or dilatation of the vessels, decline of arterial pressure, intensified movements of the stomach, intestine, vomiting, increased salivation. They can be realized because of connections of the vestibular nuclei with nuclei of n. vagus and reticular formation.

Unit XIII PHYSIOLOGY OF HUMAN BEHAVIOUR

Chapter 37. Unconditioned and conditioned reflectory activity

1. UNCONDITIONED AND CONDITIONED REFLEXES

Higher nervous activity is expressed in complex reflex reactions accomplished with the participation of the brain cortex and the adjacent subcortical formations. The first concepts of the reflex character of the cerebral activity were developed by I. M. Sechenov in his monograph "Reflexes of the Brain".

A famous Russian physiologist Ivan Petrovich Pavlov of Russia (who was awarded a Nobel Prize in 1904 for his works in gastric physiology, involving conditioned reflex in the beginning of that century) is regarded as the father of the subject of conditioned reflexes which were the background for the theory of the higher nervous activity studying. An outline about the idea of conditioned reflex has been given in Chapter #25 concerning the mechanisms of salivary secretion.

When a piece of meat is put within the mouth cavity of a dog or any other animal or man, it salivates. This is an example of *unconditioned reflex*, that is, this reflex is present since birth of the animal and requires no training/ practice/ experience. Therefore, the *unconditioned reflexes* (offered by I.P.Pavlov) — are individual and specific stereotype reactions of organism on the inside irritants, which are fastened genetically and realized with the help of the Central Nervous System.

Closely to the unconditioned reflexes are *instincts* defined as *complicated unconditioned reactions.*

On the other hand, as the dog in its childhood, learns to recognize a piece of meat by sighting or smelling, (in its grown up stage) the *mere* sighting or smelling

of meat (without any actual contact of meat with the tongue) results in salivation. This is *conditioned reflex*, because it is dependent on experience. Hence, the *conditioned reflex* (offered by I.P.Pavlov) — is a systemic adaptative reaction, individually acquired during life or special study and based on the temporary connection between conditioned irritant and unconditioned action. The differences between unconnditioned and conditioned reflexes given below (Table 38–1).

Table 38–1. Difference between the conditioned and unconditioned reflexes

N	Unconditioned reflexes	Conditioned reflexes
1	Inborn, inherited reactions	Reactions acquired by organism in the course of its individual development on the basis of "life experience".
2	Specific, i.e., characteristic of all representatives of a given species.	Individual and can be present in some members of a species and absent in others
3	Relatively stably	Unstable and may be developed, reinforced and extinguished under definite conditions.
4	They are evoked in response to adequate stimuli applied to definite receptive fields.	Can be developed under the action of the most varied stimuli applied to different receptive fields.
5	Performance of the unconditioned reflexes can be on the basis of the lower divisions of the CNS (brain stem and spinal cord. F. i., knee reflex).	They are needed in higher parts of the brain (brain cortex).
6		They are built on the basis of unconditioned ones.
7	Majority of them are constant for all the life.	They are temporary and need connections between the centers. Conditioned reflexes can extinguish (disappear).

From experience, the dog knows what is the appearance or smell of meat. It *is obvious* that this is also a form of *learning* (read later Chapter # 40).

The typical procedure for *classical conditioned reflex* induction involves presentations of a *neutral* stimulus along with a *stimulus of some significance*. The neutral stimulus could be any event that does not result in an overt behavioral response from the organism under investigation. Pavlov referred to this as

a *conditioned stimulus* (CS). Conversely, presentation of the significant stimulus necessarily evokes an innate, often reflexive, response. Pavlov called these the *unconditioned stimulus* (US) and *unconditioned response* (UR), respectively. If the CS and the US are repeatedly paired, eventually the two stimuli become associated and the organism begins to produce a behavioral response to the CS. Pavlov called this the *conditioned response* (CR).

A further improvement of the above picture can be made. A bell, as originally shown by I. P. Pavlov, can be ringed and a piece of meat may be supplied following the ringing of the bell. After only a few such sessions it will be seen that the dog now begins to salivate as soon as it hears the sound of the bell, even if there is no following up by supplying the meat. Or in other words, the dog has learnt to associate (conditioned to) the ringing of the bell with the meat (Fig. 38–1).

This form of conditioning is called *classical or Pavlovion conditioning*. In this type of conditioning, the subject (the dog) has no control over the stimulus. Moreover, the conditioning stimulus (bell in the above example) is followed by the unconditioned stimulus, viz, the meat.

For a conditioned reflex to appear it is necessary that certain conditions are fulfilled:

1) coexistence in time, several times repeated, of the indifferent agent and the unconditioned stimulus (in the previous example, the sound of the bell presentation of the meat);

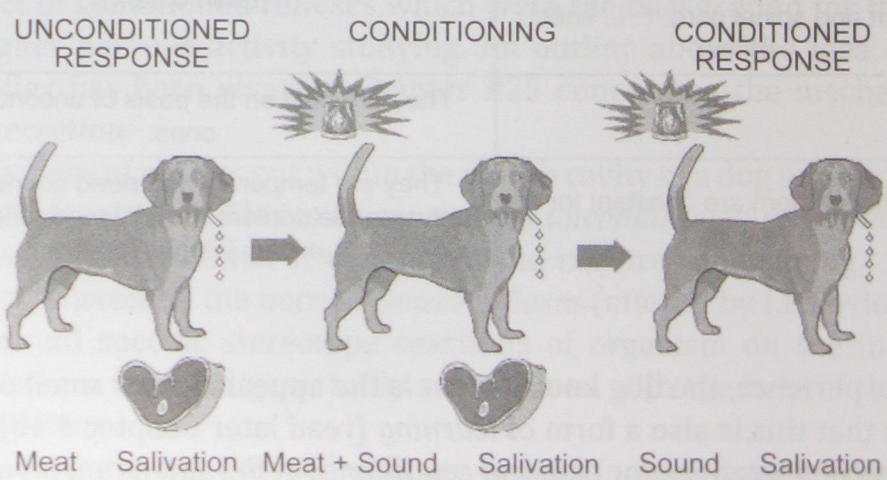


Fig. 38–1. The classic experience of Pavlov is that of the dog, the bell and the salivation to the view of a piece of meat.

2) the indifferent agent should somewhat precede (1-5 sec) the unconditioned stimulus. If we first give the meat and then ring the bell, the conditioned response is not established;
3) unconditioned stimulus should be stronger by its biological importance (food is a stronger stimulus than bell or light);
4) conditioned stimulus should be indifferent. The orientation reaction to it should be extinguished;
5) absence of other stimuli that could induce externally caused inhibition. If we simultaneously whip the animal or splash it with cold water, we provoke inhibition, setting loose defensive reactions;
6) the animal should be healthy and active. The centers, responsible for the conditioned and unconditioned stimuli perception, should have high excitability;
7) stimuli from the visceral organs (fullness of the urinary bladder, rectum) should be excluded;
8) the experiments should take place at the same time every day;
9) several dozens of combinations (10-20) of conditioned and unconditioned stimuli should be done during the experiment;
10) in order to preserve the conditioned reflex, it is necessary to reinforce it periodically. Once the reflex is formed, the mere sound of the bell substitutes the presentation of the meat. But, if we repeatedly ring the bell without showing the meat, after a certain time the animal stops reacting with salivation and digestive secretion.

Another form of conditioning is known as the *'Instrumental Conditioning'*, (also called 'Skinners conditioned') where the subject has control over the stimulus. The special chamber used in this technique for making instrumental conditioned reflexes and studying their mechanisms, was worked out by Skinner.

For example, if a chronically placed electrode will be placed in a 'reward area' (e.g., the lateral hypothalamus), the animal will repeatedly press the bar; the animal reflectory presses the bar and it is now conditioned to press the bar whenever it sees it. Here, the animal has control over the stimulus.

2. IMPORTANCE OF CONDITIONED REFLEX

Conditioned reflex is a very important tool for studying the process of learning and behavior in animals (even in men). Some psychiatric disorders are

related to conditional reflexes. Finally, the response of a society to a situation, depends, as a whole, on the conditioning of the society; and that is why, the response, to some extent, may be predicted.

We will discuss some conditioned reflex characteristic features below:

1. ***Reinforcement.*** Suppose in a Pavlovian conditioning experiment, a bell is rung (conditioning stimulus) and then meat is supplied to the dog; as a result, the reflex is established, so that when the bell rings the dog salivates (the response or behavior). By reinforcement it is meant that the whole procedure is repeated over and over again.
2. ***Conditioned (internal) inhibition.*** If the conditioning stimulus be now (in an animal where the sequence, ringing ® salivation has been established) made without the follow through by the unconditioned reflex (e.g. supply of the meat in the above example), and the whole procedure is repeated several times, no response (salivation) will occur any more; that is ringing the bell will no more elicit salivation. This phenomenon is called '*internal inhibition*'.

External inhibition can be understood with the help of a typical example: normally dogs are annoyed by cats. Now, a dog, which has been conditioned to the sequence, ringing of the bell ® salivation, suddenly sees a cat after the bell has been rung, there will be no salivation (despite the ringing of the bell); this (i. e, suppression of the expected effect due to distraction of mind) is called external inhibition.

3. HIGHER AND AUTONOMIC CONDITIONING

These also can be made and by this procedure *heart rate, glomerular filtration rate, pupillary aperture* and many other functions of the body, normally which are regarded as if dependent of the will can be influenced and the mind (including the emotion) controlled. A dramatic example is given below.

Normally in any man, if light is shown to the eye, there is constriction of the pupil. Now, a device can be so made, that when a man presses a key and whispers to himself, "start", there is completion of an electric circuit resulting in automatic shining of light to his eyes. After a few such sessions, when he merely says "start" (*without* actually pressing, so that no light shines to his eyes) there will be pupillary constriction.

The classical Indian Vaishnav, whenever he feels enraged (an emotion), utters, '*Hare Krishna*' and thus suppresses his emotion. The words '*Hare Krishna*' are associated with mental calmness, bliss and purity and so the emotion (the rage) disappears. The words Hare Krishna are the conditioning stimulus and control of rage (Which normally is not easy to control) is the effect.

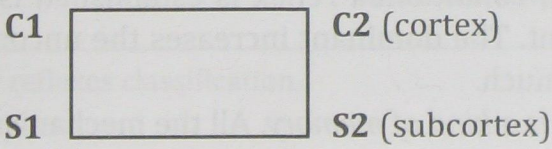
4. MECHANISM OF CONDITIONING

It may be that the phenomenon of conditioned reflex is due to development of new neural connection.

Pavlov's concept of the conditioned reflex building is based on the formation of the temporary connections between the groups of cortical cells, receiving conditioned stimulation and those receiving unconditioned.

It is the idea of a horizontal connection.

(C1 — C2)



Thus when the conditioned reflex causing salivation as a result of ringing of bell develops, it is to be presumed that between the auditory area of temporal lobe and salivary nucleus of the brain stem, new neural connections have developed. Pavlov further thought that this type of new neural connections can develop only in presence of cortex, that is, to produce conditioned reflex, presence of cerebral cortex is essential. It is now known that such conditioning can develop at lower levels too.

But later it was established, that the subcortical structures play a great role. And the connection "bend" deeper on the subcortical level (cortico-subcortico-cortical pathways). It was approved that if after the conditioned reflex formation the connection C1-C2 is transected, the realization of the reflex is possible (Fig. 38–2). After this author came to the conclusion about multilevel organization of the temporary connection (Fig. 38–3).

However, alternate theories to explain the mechanism of conditioning, also exist. One such (alternative) theory is that, no new neural circuits develop. What happens is that in an existing circuit (required to produce the chain of

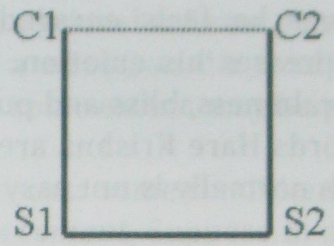


Fig. 38-2

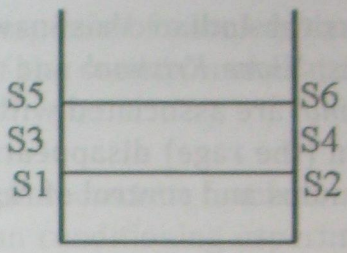


Fig. 38-3

The schematical representation of the subcortical structures participation in the conditioned reflexes formation.

events, e.g, *hearing* the gong of the bell, *salivation*, the effect), the neurotransmitters in the synapses of the circuit are secreted more quickly and more vigorously resulting in the effect.

The majority of investigators now believe that an important role in the mechanism by which conditioned reflex is established is played by the phenomenon of dominant. The dominant increases the unconditioned reflex centers excitability too much.

Conditioned reflex is a kind of memory. All the mechanisms of the short-term and then long-term memory are important for the temporary connection formation (see Chapter #40). The most probable theory today, therefore, is that functional changes occurring during combination of the stimuli are important only for brief impression of traces of stimuli on the cortex ("short memory"), however, is based on some molecular, plastic changes taking place in the synapses or even in the cerebrocortical cells themselves. Various theories are offered to explain the essence of the plastic changes.

According to one of them, frequent transmission of impulses to the cells along the same pathways results in proliferation of presynaptic terminals and an increase in the number of synapses. According to another theory, repeated passage of impulses through a synapse causes persistent changes in the enzyme systems in the nerve endings, as a result of which the transmission of impulses is made easier.

In recent years great importance has been attached to the fact, that stimulation of a nerve cell increases its content of nucleic acid. The impression of traces of preceding stimulation of the cell is supposed to be due to changes in the structure of the ribonucleic acid, which leads to a change of protein

synthesis in the cell protoplasm, processes, and synapses, and this may lead to proliferation of synapses on certain nerve cells.

5. CONDITIONED INHIBITION

According to both unconditioned and conditioned reflexes one could identify the same types of inhibition. Their classification is given below (Fig. 38–4).

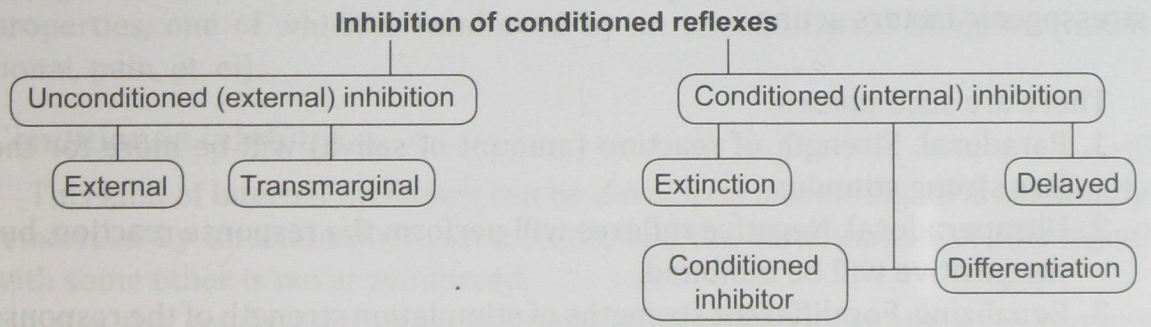


Fig. 38–4. Conditioned reflexes classification.

Unconditioned (external and transmarginal) kinds of inhibition are caused by inborn properties of the nervous system. They have not to be developed.

External as a kind of unconditioned inhibition can be caused by the action of various external stimuli.

For example: An unknown person comes during the experiment. It will cause the orientation reflex "What is this?" The amount of saliva in the dog will decrease.

In human it is like distraction of attention.

If the same situation will be repeated several times, the orientation reaction will extinct. But, there are some factors, which always cause inhibition of conditioned reflexes. They are: pain, strong stimulation of the visceral organs (urinary bladder, signals for defecation, sexual excitement).

As impulses from these receptors are addressed to the centers, outside of the conditioned connection, this kind of inhibition is known as external.

Transmarginal inhibition. If the strength of the conditioned stimulus is too excessive, instead of increasing the strength of response reaction, its inhibi-

tion takes place. It has defense importance and prevents the exhaustion of the CNS.

Transmarginal inhibition can be caused also by weak, but monotonous and long-acting stimulus. For example: sound of rain causes sleep.

This kind of inhibition is manifestation of "Pavlov's law of strength" break. It should be noticed that the result of the increased strength of the conditioned stimulus may be as transmarginal inhibition, but it also may be as "the phase states".

The *phase states* are such disorders of the conditioned-reflex activity, when the Pavlov's law of strength is broken. It may be the result of a very strong or stressogenic factors acting.

There are such phases:

1. Paradoxal. Strength of reaction (amount of saliva) will be more for the less strong stimulus.
2. Ultraparadoxal. Negative reflexes will perform the response reaction, but the positive will be inhibited.
3. Equalizing. For different strengths of stimulation strength of the response will be the equal.
4. Inhibitory. Pavlov's law of strength is correct, but manifestation of reaction (amount of saliva) will be less for all the stimuli.

6. CONDITIONED (INTERNAL) KINDS OF INHIBITION OF CONDITIONED REFLEXES

In contrast to the *unconditioned inhibition*, which is produced by the first application of the stimulus, *conditioned inhibition* has to be developed. It is also known as internal because the conditioned stimuli are addressed to the centers, inside of the conditioned connection.

Extinction of the conditioned reflexes can be formed if conditioned stimulus is not reinforced by an unconditioned stimulus many times.

In the man extinction is the same as forgetting. It is a good feature of CNS because we forget only useless information.

In natural condition usually extinct only useless conditioned reflexes which lost their importance.

Differentiation of conditioned stimuli

If a nutritional conditioned reflex to some stimulus is established to a tone of 1000 Hz, the dog performs the same response to 900 and 1100 Hz. The phenomenon is known as generalization of the conditioned reflex.

But, if the tone of 1000 Hz will be constantly reinforced by an unconditioned stimulus, and signals close to it will be without reinforcement, the dog begins to distinguish the tone of 1000 Hz from the others. The reactions to the others frequencies will extinguish and secretion of saliva will stop.

Differentiation is discrimination between two signals similar in physical properties, one of which is reinforced by an unconditioned stimulus (nutritional, pain, et. el).

Conditioned inhibitor

This kind of internal inhibition can be developed if some signal is constantly reinforced by unconditioned stimulation, but the combination of this signal with some other is never reinforced.

Later the combination loses its positive signal significance and becomes *"conditioned inhibitor"*. For instance, turning on the lamp is constantly reinforced by food, but combination of turning on the lamp and the bell is never used with unconditioned stimulation. The dog begins to distinguish the separate signal (light of the lamp) and its combination with the other.

Delay

Delayed conditioned reflexes are performed by the response delay as a result of delay of the unconditioned stimulus (reinforcment).

If an unconditioned stimulus (e. g., food) is used in short time (1–5 seconds) after the beginning of the conditioned agent, the conditioned reflex reaction (salivation) occurs immediately after the beginning of signal stimulation.

But if the reinforcing stimulus is gradually delayed (maximum to 2–3 min in the dogs) after the beginning of the conditioned stimulation, the conditioned reflex reaction will delay too.

For example. After a sound of the bell salivation will be 2 min. later. In the man the similar processes are patience, self-control.

Chapter 38. Physiology of memory

Many definitions of *memory* associate it with a function of the brain. We sorted many of them but one of the best is seemed to be the following one. **Memory is the ability of an organism to store, retain, and subsequently recall information.** In the recent decades, memory and learning have become one of the principal parts of a new branch of science that represents a marriage between cognitive psychology and neuroscience, called cognitive neuroscience.

1. CLASSIFICATIONS

There are several ways of classifying memories, based on duration, nature and retrieval of information. Therefore, it sholuld be mentioned that from an information processing perspective there are three main stages in the formation and retrieval of memory:

- *Encoding* (processing and combining of received information);
- *Storage* (creation of a permanent record of the encoded information);
- *Retrieval/Recall* (calling back the stored information in response to some cue for use in some process or activity).

A. Classification by duration

A basic and generally accepted classification of memory is based on the duration of memory retention, and identifies three distinct types of memory:

- *sensory memory*,
- *short-term memory*, and
- *long-term memory*.

The *sensory memory* corresponds approximately to the initial moment that an item is perceived. Some of this information in the sensory area proceeds to the sensory store, which is referred to as *short-term memory*. Sensory memory is characterized by the duration of memory retention from milliseconds to seconds and short-term memory from seconds to minutes.

These stores are generally characterised as of strictly limited capacity and duration, whereas in general stored information can be retrieved in a period of

time which ranges from days to years; this type of memory is called *long-term memory*.

It may be that short-term memory is supported by transient changes in neuronal communication, whereas long-term memories are maintained by more stable and permanent changes in neural structure that are dependent on protein synthesis. Some psychologists, however, argue that the distinction between long- and short-term memories is arbitrary, and is merely a reflection of differing levels of activation within a single store.

If we are given a random seven-digit number, we may remember it only for a few seconds and then forget (short-term memory). On the other hand, we can remember telephone numbers for many years (assuming we use them often enough). Those long-lasting memories are said to be stored in long-term memory.

Additionally, the term *working memory* is used to refer to the short-term store needed for certain mental tasks — it is not a synonym for *short-term memory*, since it is defined not in terms of duration, but rather in terms of purpose. Some theories consider working memory to be the combination of short-term memory and some attentional control. For instance, when we are asked to mentally multiply 45 by 4, we have to perform a series of simple calculations (additions and multiplications) to arrive at the final answer. The ability to store the information regarding the instructions and intermediate results is what is referred to as working memory.

B. Classification by nature of memory

According to the nature one could identify the following types of the nature:

- *neural memory*
- *immunological memory*, and
- *genetical memory*.

C. Classification by information type

Long-term memory can be divided into *declarative (explicit)* and *procedural (implicit)* memories.

Declarative memory requires conscious recall, in that some conscious process must call back the information. It is sometimes called *explicit memory*, since it consists of information that is explicitly stored and retrieved.

Declarative memory can be further sub-divided into semantic memory, which concerns facts taken independent of context; and episodic memory, which concerns information specific to a particular context, such as a time

and place. Semantic memory allows the encoding of abstract knowledge about the world, such as "Paris is the capital of France". Episodic memory, on the other hand, is used for more personal memories, such as the sensations, emotions, and personal associations of a particular place or time. Autobiographical memory — memory for particular events within one's own life — is generally viewed as either equivalent to, or a subset of, episodic memory. Visual memory is part of memory preserving some characteristics of our senses pertaining to visual experience. We are able to place in memory information that resembles objects, places, animals or people in sort of a mental image. Visual memory can result in priming and it is assumed some kind of perceptual representational system or PRS underlies this phenomenon.

In contrast, procedural memory (or *implicit memory*) is not based on the conscious recall of information, but on implicit learning. Procedural memory is primarily employed in learning motor skills and should be considered a subset of implicit memory. It is revealed when we do better in a given task due only to repetition — no new explicit memories have been formed, but we are unconsciously accessing aspects of those previous experiences. Procedural memory involved in motor learning depends on the cerebellum and basal ganglia.

D. Classification by temporal direction

A further major way to distinguish different memory functions is whether the content to be remembered is in the past, retrospective memory, or whether the content is to be remembered in the future, prospective memory.

Thus, retrospective memory as a category includes semantic memory and episodic/autobiographical memory. In contrast, prospective memory is memory for future intentions, or *remembering to remember*. Prospective memory can be further broken down into event- and time-based prospective remembering. Time-based prospective memories are triggered by a time-cue, such as going to the doctor (action) at 4pm (cue). Event-based prospective memories are intentions triggered by cues, such as remembering to post a letter (action) after seeing a mailbox (cue). Cues do not need to be related to the action (as the mailbox example is), and lists, sticky-notes, knotted hankerchiefs, or string around the finger (see box) are all examples of cues that are produced by people as a strategy to enhance prospective memory.

E. Classification of the steps of memory

Memory functions through three steps:

- Acquisition

- Consolidation
- Retrieval

Acquisition. Before you can remember something, you first must learn the information. This is called acquisition. This acquired information is then put into temporary nerve-cell pathways in the brain. These pathways are where you store short-term memory.

Consolidation. In order for something to be placed in long-term memory, the nerve pathways have to be strengthened and reinforced. This process, called consolidation, can take weeks or even months. There are several factors that affect whether or not information will be put into long-term memory. For example, you are more likely to retain information if it relates to pre-existing memories or somehow stimulates you emotionally. Also, it doesn't hurt to have a good night's sleep, as this too helps you retain information.

Retrieval. When people retrieve information, they are literally "recalling" it from the nerve pathways. The brain reactivates a particular pathway, and information is remembered. This process can be fast or slow, depending on how familiar you are with the information and how well you learned it in the first place.

2. PHYSIOLOGICAL MECHANISMS OF MEMORY

Overall, the mechanisms of memory are not well understood. Brain areas such as the hippocampus, the amygdala, or the mamillary bodies are thought to be involved in certain kinds of memory.

For example, the hippocampus is believed to be involved in spatial learning and declarative learning. Damage to certain areas in patients and animal models and subsequent memory deficits is a primary source of information. However, rather than implicating a specific area, it could be that damage to adjacent areas, or to a pathway traveling through the area is actually responsible for the observed deficit. Further, it is not sufficient to describe memory, and its counterpart, learning, as solely dependent on specific brain regions.

Learning and memory are attributed to changes in neuronal synapses, thought to be mediated by long-term potentiation and long-term depression. Other scientists who have investigated the nature of memory such as neurologists John Eccles and Wilder Penfield and biologist Rupert Sheldrake have suggested that memories are a field phenomenon and are not stored in the brain at all, but rather accessed through neurological structures.

Basically, memory and learning are inextricably linked because part of the learning process involves the assimilation of new information and its commitment to memory. The most likely sites of learning in the human brain are the large association areas of the cerebral cortex, in coordination with subcortical structures deep in the temporal lobe, including the hippocampus and amygdala. The association areas draw on sensory information received from the primary visual, auditory, somatic sensory, and olfactory cortices and on emotional feelings transmitted via the limbic system. This information is integrated with previously learned skills and stored memory, which presumably also reside in the association areas.

The learning process itself is poorly understood, but it can be studied experimentally at the synaptic level in isolated slices of mammalian brain or in more simple invertebrate nervous systems. Synapses subjected to repeated presynaptic neuronal stimulation show changes in the excitability of postsynaptic neurons. These changes include the facilitation of neuronal firing, altered patterns of neurotransmitter release, second messenger formation, and, in intact organisms, evidence that learning occurred.

The phenomenon of increased excitability and altered chemical state on repeated synaptic stimulation is known as *long-term potentiation*, a persistence beyond the cessation of electrical stimulation, as is expected of learning and memory. An early event in long-term potentiation is a series of protein phosphorylations induced by receptor-activated second messengers and leading to activation of a host of intracellular proteins and altered excitability. In addition to biochemical changes in synaptic efficacy associated with learning at the cellular level, structural alterations occur. The number of connections between sets of neurons increases as a result of experience.

Much of our knowledge about human memory formation and retrieval is based on studies of patients in whom stroke, brain injury, or surgery resulted in memory disorders.

Such knowledge is then examined in more rigorous experiments in nonhuman primates capable of cognitive functions. From these combined approaches, we know that the prefrontal cortex is essential for coordinating the formation of memory, starting from a learning experience in the cerebral cortex, then processing the information and communicating it to the subcortical limbic structures. The prefrontal cortex receives sensory input from the parietal, occipital, and temporal lobes and emotional input from the limbic system. Drawing on skills such as language and mathematical ability, the prefrontal cortex integrates these inputs in light of previously acquired learning. The prefrontal cortex can thus be considered the site of *working memory*, where new

experiences are processed, as opposed to sites that consolidate the memory and store it. The processed information is then transmitted to the hippocampus, where it is consolidated over several hours into a more permanent form that is stored in, and can be retrieved from, the association cortices.

3. DECLARATIVE AND PROCEDURAL MEMORY

A remarkable finding from studies of surgical patients who had bilateral resections of the medial temporal lobe is that there are two fundamentally different memory systems in the brain.

Declarative memory refers to memory of events and facts and the ability to consciously access them. Patients with bilateral medial temporal lobectomies lose their ability to form any new declarative memories. However, they retain their ability to learn and remember new skills and procedures. This type of memory is called *procedural memory* and involves several different regions of the brain, depending on the type of procedure.

In contrast to declarative memory, structures in the medial temporal lobe are not involved in procedural memory. Learning and remembering new motor skills and habits requires the striatum, motor areas of the cortex, and the cerebellum. Emotional associations require the amygdala. Conditioned reflexes require the cerebellum. An early demonstration of the dichotomy between declarative and procedural memory came from studies by Dr. Brenda Milner on a patient of Dr. Wilder Penfield in the mid-1950s. This patient (H.M.) had received a bilateral medial temporal lobectomy to treat severe epilepsy and, since that time, has been unable to form any new declarative memories. This deficit is called *anterograde amnesia*. Dr. Milner was quite surprised to learn that H.M. could learn a relatively difficult mirror-drawing task, in which (like anyone else) he got better with repeated trials and retained the skill over time. However, he could not remember ever having done the task before.

4. SHORT-TERM MEMORY

Short-term memory, also known as working memory, stores information that you need to remember in the following seconds, minutes or hours. An example would be a telephone message that you are given and must remember until you pass it on.

Declarative memory can be divided into that which can be recalled for only a brief period (seconds to minutes), and that which can be recalled for weeks to years. Newly acquired learning experiences can be readily recalled for only a few minutes or more using *short-term memory*. An example of short-term memory is looking up a telephone number, repeating it mentally until you finish dialling the number, then promptly forgetting it as you focus your attention on starting the conversation. Short-term memory is a product of working memory; the decision to process information further for permanent storage is based on judgment as to its importance or on whether it is associated with a significant event or emotional state.

An active process involving the hippocampus must be employed to make a memory more permanent.

5. LONG-TERM MEMORY

Long-term memory stores information that your brain retains because it is important to you. Basic information remembered includes names of family and friends, your address, as well as information on how to do certain activities and tasks.

The conversion of short-term to *long-term memory* is facilitated by repetition, by adding more than one sensory modality to learn the new experience (e.g., writing down a newly acquired fact at the same time one hears it spoken) and, even more effective, by tying the experience (through the limbic system) to a strong, meaningful emotional context. The role of the hippocampus in consolidating the memory is reinforced by its participation in generating the emotional state with which the new experience is associated. As determined by studying patients such as H.M., the most important regions of the medial temporal lobe for long-term declarative memory formation are the hippocampus and parahippocampal cortex.

Once a long-term memory is formed, the hippocampus is not required for subsequent retrieval of the memory. Thus, H.M. showed no evidence of a loss of memories laid down prior to surgery; this type of memory loss is known as *retrograde amnesia*. Nor was there loss of intellectual capacity, mathematical skills, or other cognitive functions.

An extreme example of H.M.'s memory loss is that Dr. Milner, who worked with him for years, had to introduce herself to her patient every time they met, even though he could readily remember people and events that had occurred before his surgery.

6. LONG-TERM POTENTIATION

Long-term potentiation (LTP) considered to be one of the most important neurophysiological mechanisms involved both in the short-term and long-term memory elucidation. LTP was discovered in the mammalian hippocampus by Terje Lømo in 1966.

LTP is the long-lasting enhancement in efficacy of the synapse between two neurons (Fig. 40–1). Though its biological mechanisms have not yet been fully determined, LTP is believed to contribute to synaptic plasticity in living animals, providing the foundation for a highly adaptable nervous system. Most neuroscientific learning theories regard long-term potentiation and its opposing process, long-term depression, as the cellular basis of learning and memory.

Above the diagram are examples of the raw EPSP field potentials before and after tetanus.

Experimentally, a series of short, high-frequency electric stimulations to a nerve cell synapse can strengthen, or *potentiate*, that synapse for minutes to hours. In living cells, LTP occurs naturally and can last from hours to days, months, and years.

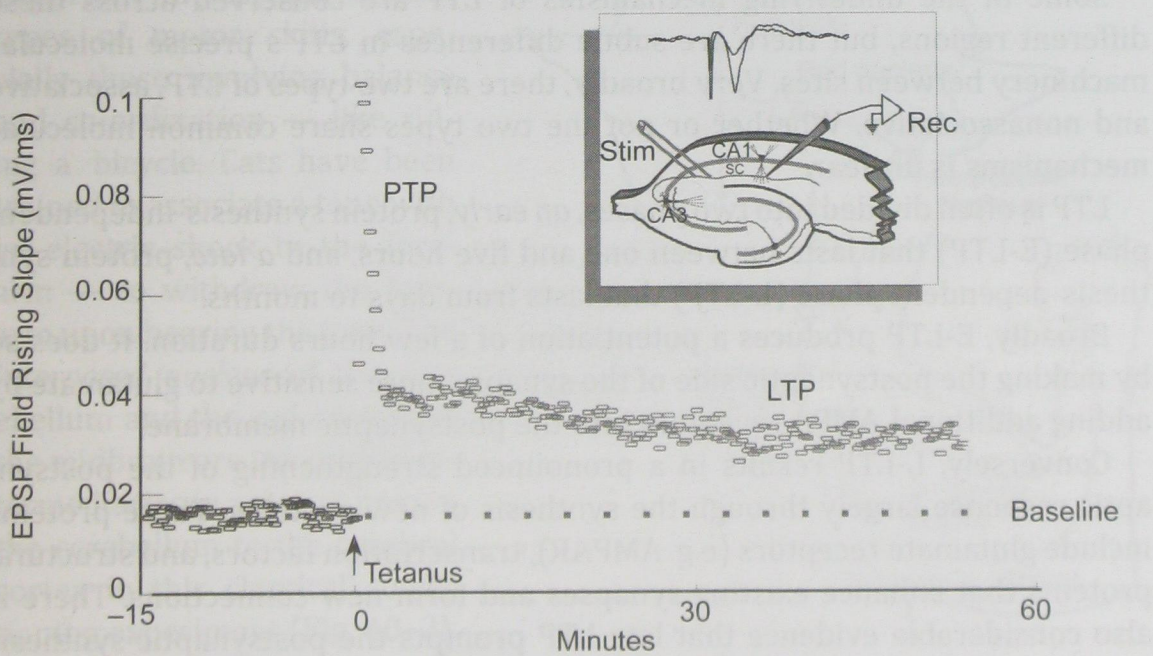


Fig. 40–1. An example of long-term potentiation recorded at Schaffer collateral/CA_1 synapses in the hippocampus of rat brain.

LTP was first observed by Terje Lømo in 1966 in the Oslo, Norway, laboratory of Per Andersen. There, Lømo conducted a series of neurophysiological experiments on anesthetized rabbits to explore the role of the hippocampus in short-term memory.

Isolating the connections between two parts of the hippocampus, the perforant pathway and dentate gyrus, Lømo observed the electrical changes in the dentate gyrus elicited by stimulation of the perforant pathway. As expected, a single pulse of electrical stimulation to the perforant pathway elicited an excitatory postsynaptic potential (EPSP) in the dentate gyrus. What Lømo did not expect was that the postsynaptic responses to these single-pulse stimuli could be enhanced by first delivering a high-frequency train of stimuli to the synapse.

When such a train of stimuli was applied, subsequent single-pulse stimuli elicited stronger, prolonged EPSPs. This phenomenon was soon dubbed "long-term potentiation".

Since its original discovery in the rabbit hippocampus, LTP has been observed in a variety of other neural structures, including the cerebral cortex, cerebellum, amygdala, and many others. It was suggested that LTP may even occur at all excitatory synapses in the mammalian brain.

Some of the underlying mechanisms of LTP are conserved across these different regions, but there are subtle differences in LTP's precise molecular machinery between sites. Very broadly, there are two types of LTP, associative and nonassociative. Whether or not the two types share common molecular mechanisms is unclear.

LTP is often divided into two phases, *an early*, protein synthesis-independent phase (E-LTP) that lasts between one and five hours, and *a late*, protein synthesis-dependent phase (L-LTP) that lasts from days to months.

Broadly, E-LTP produces a potentiation of a few hours duration. It does so by making the postsynaptic side of the synapse more sensitive to glutamate by adding additional AMPA receptors into the postsynaptic membrane.

Conversely, L-LTP results in a pronounced strengthening of the postsynaptic response largely through the synthesis of new proteins. These proteins include glutamate receptors (e.g. AMPAR), transcription factors, and structural proteins that enhance existing synapses and form new connections. There is also considerable evidence that late LTP prompts the postsynaptic synthesis of a retrograde messenger that diffuses to the presynaptic cell increasing the probability of neurotransmitter vesicle release on subsequent stimuli. All of this is largely hypothetical. The proposed mechanism of L-LTP are only weakly

supported by existing data. Many investigators in the field doubt the very existence of L-LTP.

7. SOME CENTRAL NERVOUS SYSTEM LEARNING MECHANISMS

Most of what is known about the basis of learning and memory in the vertebrate brain involves correlating activity of certain types of neuron groups with certain types of learning. Details of cell physiology are rarely obtainable by current methods.

There are evidently many different areas in the human nervous system where learning occurs — and many means by which learning occurs. Preserving many or even any of these mechanisms may not be essential to the preservation of identity — or even to the preservation of memory. Preserving long-term habituation of spinal reflexes does not seem crucial for maintaining identity.

The cerebellum seems to be important in learning certain types of motor skills, especially those involving balance and co-ordination — like riding a bicycle. Cats have been trained to associate a tone with an electric shock to the forearm — to withdraw the forearm upon hearing the tone. The *interposed nucleus* of the cerebellum and the *red nucleus* of the midbrain are the critical intermediaries in a circuit linking the cerebellum to the cerebral cortex in this classical conditioning experiment (Fig. 40–2). New synapses are apparently formed in the red nucleus in association with this learning.

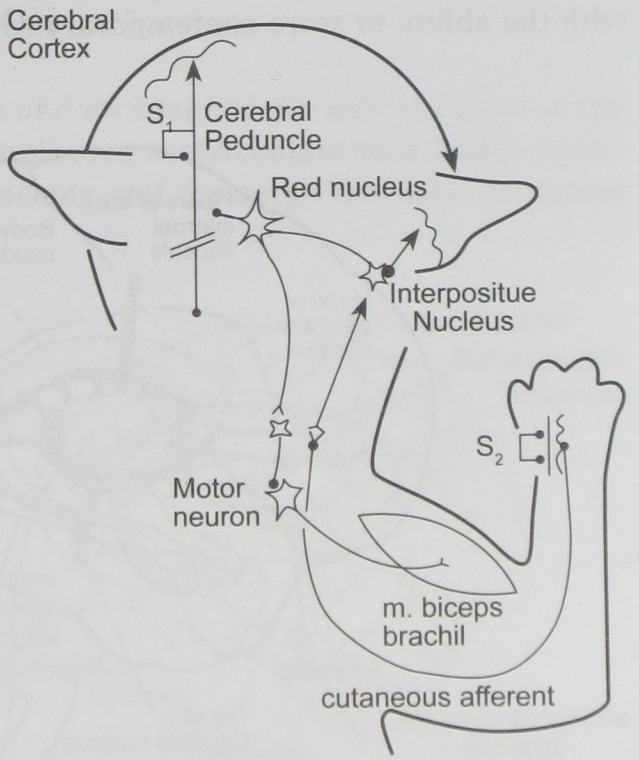


Fig. 40–2. Pathways of Interposed and Red Nucleus.

Rats raised in an environment where they could exercise and be active showed 23 % more spines on the dendrites of Purkinje cells of the cerebellum than rats in cages allowing only enough space for access to food and water. Similarly, rats raised in an "enriched" environment allowing for exploration showed increased branching of dendrites in the primary visual area of the cerebral cortex.

Operant conditioning is similar to classical conditioning, except that learning occurs through reward or punishment of spontaneous behaviors, rather than through the induction of conditioned reflexes to novel stimuli. A laboratory animal that learns to press a lever to get a food pellet is demonstrating operant conditioning. The basal ganglia are evidently very important in learning that links rewards with motor activity. The caudate nucleus and the putamen are basal ganglia that receive inputs primarily from the cerebral cortex — the caudate nucleus from the association areas and the putamen from the somatosensory and somatomotor gyri. These two nuclei are collectively called the striatum (Fig. 40–3).

A form of short-term memory referred to as *working memory* is associated with the ability to store contemporary representations of the outside world.

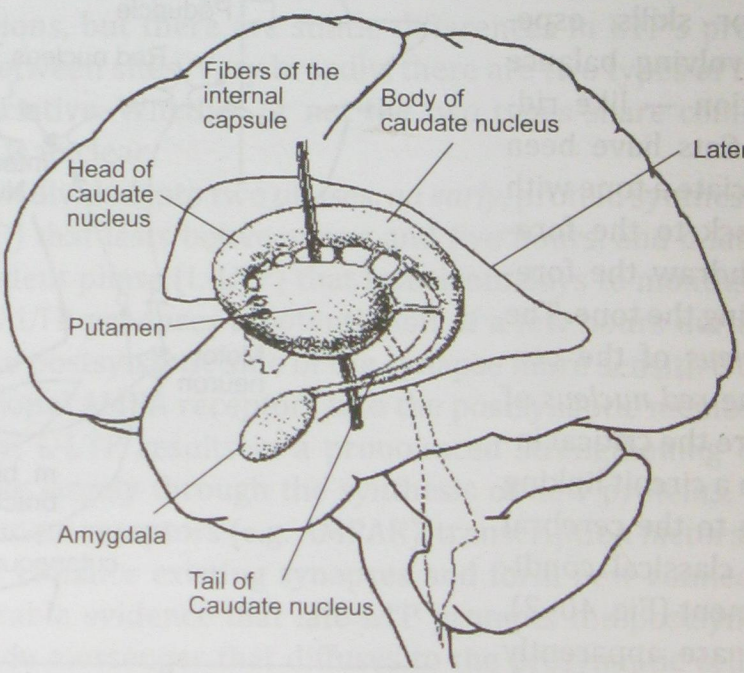


Fig. 40–3. Basal ganglia in learning.

The apparent locus for working memory is in the prefrontal area of the cerebral cortex. Monkeys shown that food is hidden behind an obscuring object, but who are restrained from immediately taking the food, maintain knowledge of the food behind the obscuring object. The monkeys will take the food after a delay, when restraints have been removed. Monkeys with areas of their prefrontal cortex surgically removed will forget about the food immediately after they can no longer see it ("out of sight, out of mind"). Dopamine seems to be the most important neurotransmitter for neurons involved in working memory.

Chronic alcoholics show learning-deficit symptoms of "Korsakoff's syndrome" as a consequence of damage to the mammillary bodies of the hypothalamus and the dorsomedial nucleus of the thalamus. Such patients have difficulty discriminating objects and words. Acquisition of new knowledge based on discrimination is very slow and short-term memory is poor, but long-term memory is not so impaired.

8. CHOLINERGIC INNERVATION

The primacy of the hippocampus and its connections with the base of the forebrain for memory formation implicates acetylcholine as a major transmitter in cognitive function and learning and memory. The basal forebrain region contains prominent populations of cholinergic neurons that project to the hippocampus and to all regions of the cerebral cortex (Fig. 40–4). These cholinergic neurons are known generically as basal forebrain nuclei and include the septal nuclei, the nucleus basalis, and the nucleus accumbens. Another major cholinergic projection derives from a region of the brainstem reticular formation known as

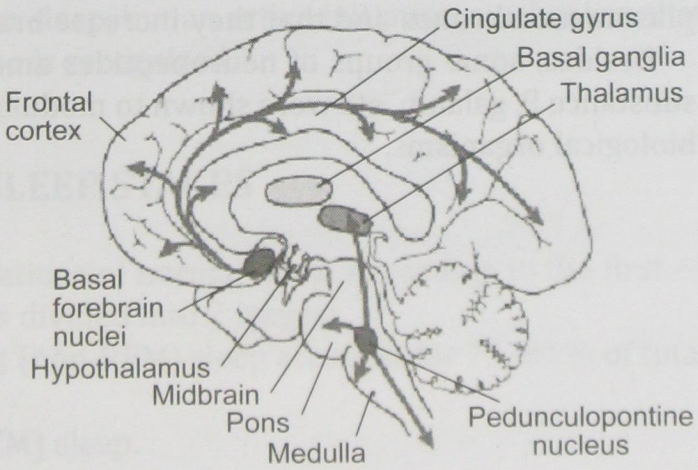


Fig. 40–4. The origins and projections of major cholinergic neurons.

the pedunculopontine nucleus, which projects to the thalamus, spinal cord, and other regions of the brainstem.

Approximately 90 % of brainstem inputs to all nuclei of the thalamus are cholinergic. Cortical cholinergic connections are thought to control selective attention, a function congruent with the cholinergic brainstem projections through the ascending reticular activating system. Loss of cholinergic function is associated with dementia, an impairment of memory, abstract thinking, and judgment.

Other cholinergic neurons include motor neurons and autonomic preganglionic neurons, as well as a major interneuronal pool in the striatum.

9. NEUROPEPTIDES AND MEMORY

If in rats the pituitary glands is removed by surgery, these subjects may under certain experimental conditions exhibit a severe learning deficit, which can readily be amended by treatment with pituitary peptide hormones (ACTH, vasopressin) or their fragments. These small peptide fragments have lost the endocrine activity of the parental molecule, but still possess the behavioural activity. From animal and human studies it can be inferred that neuropeptides derived from ACTH affect motivational aspect of learning. Neurophysiological experiments revealed that these peptides facilitate neurotransmission in some brain circuits and neurochemical studies showed that they alter the turnover of neurotransmitters and the degree of phosphorylation of synaptic plasma membranes and that they increase brain protein synthesis.

Besides, some groups of neuropeptides among which are opioid peptides, substance P, galanin, etc were shown to modulate the learning processes of the biological organisms.

Chapter 39. Physiology of sleep

Sleep is the regular state of natural rest observed in all mammals, birds and fish. Sleep is a state of unconsciousness in which the brain is relatively more responsive to internal than to external stimuli. Sleep is not actually "unconsciousness," but rather, it is a natural state of rest characterized by a reduction in voluntary body movement, decreased reaction to external stimuli, an increased rate of anabolism (the synthesis of cell structures) and a decreased rate of catabolism (the breakdown of cell structures). Therefore, since consciousness is literally the awareness of the surroundings, being asleep is just an altered state of consciousness, as opposed to being unconscious. It is heavily influenced by circadian rhythms and by hormonal and environmental factors. Sleep performs a restorative function for the brain and body as evidenced by the myriad symptoms of metabolic dysfunction that result when animals are deprived of sleep.

The predictable cycling of sleep and the reversal of relative external unresponsiveness are features that assist in distinguishing sleep from other states of unconsciousness. The brain gradually becomes less responsive to visual, auditory, and other environmental stimuli during the transition from wake to sleep, which is considered by some to be stage I of sleep.

Historically, sleep was thought to be a passive state that was initiated through withdrawal of sensory input. Currently, withdrawal of sensory awareness is believed to be a factor in sleep, but an active initiation mechanism that facilitates brain withdrawal also is recognized.

1. CLASSIFICATION OF SLEEP STAGES

There two possible classifications of normal sleep. According to the first — the easiest one, normal sleep is divided into 2 stages:

1. Non–rapid eye movement (non-REM) sleep accounts for 75–80 % of total sleep time and
2. Rapid eye movement (REM) sleep.

The second classification includes the following 5 stages that involve previously done variant of classification:

1. *Stage # 1* (light sleep) with near-disappearance of the alpha waves seen in awake states, and appearance for the first time of theta waves. The stage is sometimes referred to as *somnolence*, or "drowsy sleep". It appears at sleep onset (as it is mostly a transition state into Stage 2), and can be associated with so-called hypnagogic hallucinations. In this period, the subject loses some muscle tone, and conscious awareness of the external environment: Stage 1 can be thought of as a gateway state between wake and sleep;
2. *Stage # 2* (so-called stage of 'sleep-spindles' [12–16 Hz] and "K- complexes"). The EMG lowers, and conscious awareness of the external environment disappears. This occupies 45–55 % of total sleep;
3. *Stage # 3* with delta waves, also called delta rhythms (1–2 Hz), is considered part of slow-wave sleep and functions primarily as a transition into stage four. Overall it occupies 3–8 % of total sleep time;
4. *Stage # 4* is true delta sleep. It predominates the first third of the night and accounts for 10–15 % of total sleep time. This is often described as the deepest stage of sleep; it is exceedingly difficult to wake a subject in this state. This is the stage in which night terrors and sleepwalking occur;
5. *Stage # 5* — stage of Rapid eye movement (REM) sleep, associated with dreaming, especially bizarre, visual, and seemingly random dreams. REM sleep is predominant in the final third of a sleep period, its timing linked to circadian rhythm and body temperature. The EEG in this period is aroused and looks similar to stage 1, and sometimes includes beta waves.

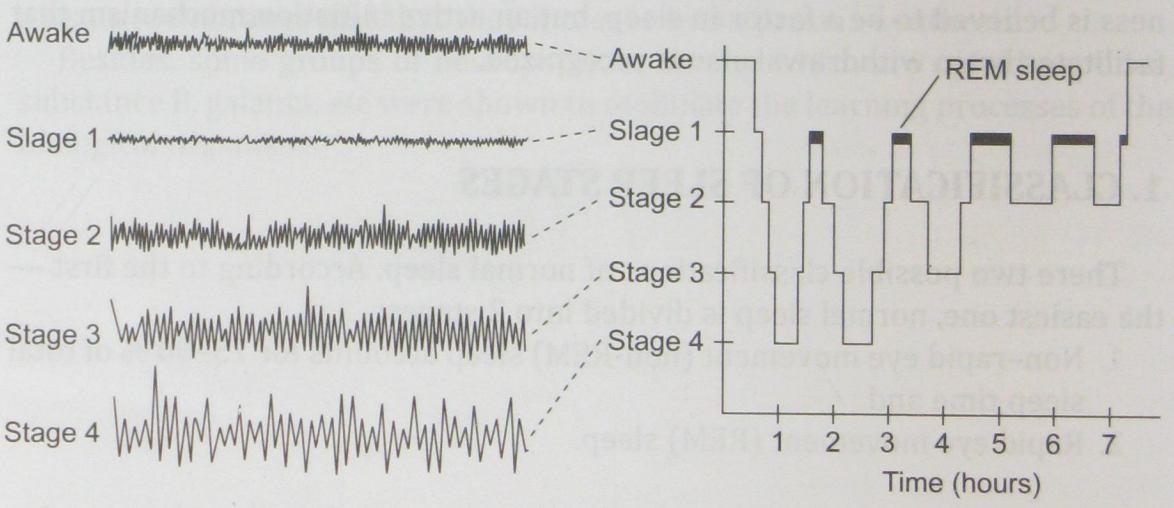


Fig. 41–1. EEG changes during the different stages of normal sleep.

The EEG pattern recorded during sleep varies in a cyclic fashion that repeats approximately every 90 minutes, starting from the time of falling asleep to awakening 7 to 8 hrs later (Fig. 41–1). For more details — see part 9 in this chapter. These cycles are associated with two different forms of sleep, which follow each other sequentially.

Slow-wave (non-REM) sleep comprises stages 1–4. Sleep is an active process that cycles at an ultradian rhythm of about 90 minutes. waking usually transitions into non-REM sleep. REM follows non-REM sleep and occurs 4–5 times during a normal 8- to 9-hour sleep period. The first REM period of the night may be less than 10 minutes in duration, while the last may exceed 60 minutes (Fig. 41–2).

For the purpose of analysis, overnight sleep has been divided into 3 equal time periods: sleep in the first third of the night, which comprises the highest percentage of non-REM; sleep in the middle third of the night; and sleep in the last third of the night, the majority of which is REM. Awakening after a full night's sleep is usually from REM sleep.

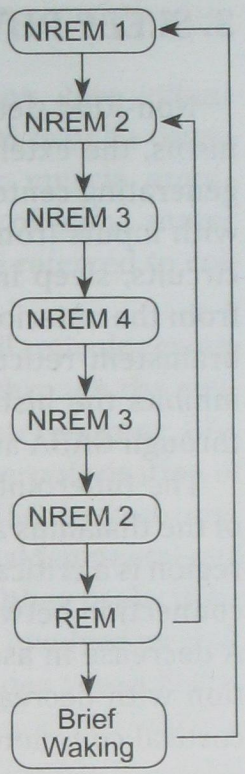


Fig. 41–2. The stages of sleep.

3. SLEEP PHYSIOLOGY

Non-REM sleep is controlled by complex initiating and maintenance mechanisms, the extent of which is not fully elaborated. Probably no single sleep-generating center exists. A more likely mechanism is sleep-generating circuits with inputs from brainstem and hypothalamic neuronal groups. Within these circuits, sleep initiation may begin with the emergence of inhibitory signals from the anterior hypothalamic preoptic nucleus directed caudally toward the brainstem reticular core and posterior hypothalamus. The preoptic nucleus inhibits the histaminergic posterior hypothalamic tuberoinfundibular region through GABA and probably acetylcholine.

The tuberoinfundibular region projects rostrally to the intralaminar nuclei of the thalamus and to the cerebral cortex. Inhibition of the tuberoinfundibular region is a critical step toward falling asleep because it results in functional disconnection between the brain stem and the more rostral thalamus and cortex. A decrease in ascending thalamic cholinergic transmissions occurs in association with decreasing cortical responsiveness. In addition to inhibiting higher cortical consciousness, the tuberoinfundibular tract projects caudally into the pontine reticular system and inhibits afferent transmissions from ascending cholinergic tracts.

Non-REM is an active state that is maintained partly through oscillations between the thalamus and the cortex. The 3 major oscillation systems are sleep spindles, delta oscillations, and slow cortical oscillations. Sleep spindles, a hallmark of stage II sleep, are generated by bursts of hyperpolarizing GABAnergic thalamic reticular neurons. These bursts inhibit thalamocortical projection neurons. As deafferentation spreads, corticothalamic projections back to the thalamus synchronize. As hyperpolarization of the thalamic reticular neurons progresses, delta waves are produced by interactions from both thalamic reticular and cortical pyramidal sources. Slow cortical oscillations are produced in neocortical networks by cyclic hyperpolarizations and depolarizations.

Although the functions of non-REM sleep remain speculative, several theories have been put forth. One theory proposes that decreased metabolic demand facilitates replenishment of glycogen stores. Another theory, which utilizes neuronal plasticity, suggests that the oscillating depolarizations and hyperpolarizations consolidate memory and remove redundant or excess synapses.

REM sleep is generated by mesencephalic and pontine cholinergic neurons, hence these are referred to as REM-on neurons. As REM sleep initiates,

monoadrenergic locus ceruleus and serotonergic raphe neurons become inactive and are called REM-off neurons.

REM is characterized by muscle atonia, cortical activation, low-voltage desynchronization of the EEG, and rapid eye movements. REM may be considered to have both tonic and phasic characteristics. Tonic muscle atonia is present throughout REM sleep. It results from inhibition of alpha motor neurons by clusters of peri–locus ceruleus neurons, which are referred to collectively as the dorsolateral small cell reticular group.

Projection of the presumed cholinergic dorsolateral small cell reticular group is through the medullary reticular formation, which projects through the ventrolateral reticulospinal tract to inhibitory spinal and bulbar interneurons. Glycinergic interneurons produce postsynaptic inhibition and hyperpolarization of the spinal alpha motor neurons. Tonic cortical activation with EEG desynchronization is promoted by projections from cholinergic lateral dorsal tegmental and pedunculopontine tegmental neurons to the thalamic nuclei. Other projections through brainstem reticular formation neurons are likely to be involved as well.

Phasic rapid eye movements are composed of lateral saccades generated in the paramedian pontine reticular formation and vertical saccades thought to be generated in the mesencephalic reticular formation. REM density is a term used to describe the frequency per minute of the eye movement bursts.

Phasic pontine-geniculate-occipital (PGO) spikes are another neurophysiological feature of REM sleep. These spikes appear to be generated by lateral dorsal tegmental and pedunculopontine tegmental neuronal bursts. They are projected to the lateral geniculate and other thalamic nuclei, and then to the occipital cortex. PGO bursts precede rapid eye movements by several seconds. Increases in PGO bursts are seen after REM sleep deprivation.

Other phasic features of REM sleep include periodic skeletal muscle twitches, increased heart rate variability, pupil dilation, and increased respiratory rate.

During non-REM sleep, the metabolic demand of the brain decreases. This is supported by oxygen positron emission tomography (PET) studies, which show that, during non-REM sleep, the blood flow throughout the entire brain progressively decreases. PET studies also show that, during REM sleep, blood flow increases in the thalamus and the primary visual, motor, and sensory cortices, while remaining comparatively decreased in the prefrontal and parietal associational regions. The increase in blood flow to the primary cortical regions may explain the vivid nature of REM dreaming, while the continued decrease in blood flow to the prefrontal cortex may explain the unquestioning acceptance of even the most bizarre dream content.

4. PHYSIOLOGIC CHANGES DURING SLEEP

A variety of behavioral and physiologic changes occur during normal wakefulness, non-REM sleep, and REM sleep.

These changes are most commonly noted in the somatic and autonomic nervous systems (ANS); in the respiratory, cardiovascular, and gastrointestinal systems; in endocrine, renal, and sexual function; and in thermoregulation. The fundamental changes in the ANS consist of an increase in the parasympathetic tone and a decrease in sympathetic activity during non-REM sleep, with further increase of parasympathetic tone and decrease in sympathetic activity during REM sleep.

During the REM sleep phase, however, sympathetic activity increases intermittently. Sympathetic nerve activity in muscle and the vascular beds of the skin, as measured by microneurographic technique, is reduced during non-REM sleep but is increased during REM sleep.

During both non-REM and REM sleep, the respiratory neurons in the pontomedullary region show a decreased firing rate. Muscle tone in the upper airway decreases in non-REM sleep and disappears in REM sleep, resulting in an increase in upper airway resistance. Two separate systems — the metabolic (or automatic) and voluntary (or behavioral) systems — control respiration during sleep and wakefulness. Both the metabolic and the voluntary system operate during wakefulness, whereas only the metabolic system operates during non-REM sleep.

The wakefulness stimuli that act through the ascending reticular activating system also act as tonic stimuli to ventilation. Hypercapnic and hypoxic ventilatory responses are moderately reduced in non-REM sleep but are more markedly decreased during REM sleep. Tidal volume and alveolar ventilation decrease during sleep; arterial oxygen tension is mildly decreased and arterial carbon dioxide tension is increased during both non-REM and REM sleep. Thus, respiration is vulnerable during normal sleep, and a few periods of apnea may occur, especially at the onset of sleep and during REM sleep.

Heart rate, blood pressure, cardiac output, and peripheral vascular resistance decrease during non-REM sleep and decrease still further during REM sleep. Cerebral blood flow and cerebral metabolic rates for glucose and oxygen decrease during non-REM sleep but increase to above waking values during REM sleep.

Profound changes in endocrine secretions are registered during sleep. Growth hormone secretion exhibits a pulsatile increase during non-REM sleep

in the first third of the normal sleep period. Prolactin secretion also rises 30 to 90 minutes after the onset of sleep. Sleep inhibits cortisol secretion. Secretion of thyroid-stimulating hormone reaches a peak in the evening and then decreases throughout the night. Testosterone levels in men increase during sleep, rising from trough levels at 8 p.m. to peak levels at 8 a.m., but no clear relation has been demonstrated between levels of gonadotropic hormones and the sleep-wake cycle in children or adults. Melatonin, which is released by the pineal gland, attains its highest secretion level between 3 a.m. and 5 a.m., then decreases to low levels during the day.

Body temperature begins to fall at the onset of sleep and reaches its lowest point during the third sleep cycle. Thermoregulation is maintained during non-REM sleep but is nonexistent in REM sleep. Penile erection and clitoral tumescence occur during REM sleep.

5. FUNCTIONAL MECHANISMS OF SLEEP

The neuroanatomic substrates for REM sleep and non-REM sleep are located in separate parts of the central nervous system.

REM sleep is generated by neurons in the pontomesencephalic region. The cells that activate REM sleep are cholinergic neurons located in the pedunculopontine tegmental nucleus and the laterodorsal tegmental nucleus in the pontomesencephalic region; these cells fire maximally during REM sleep. The cells that deactivate REM sleep are aminergic neurons located in the locus coeruleus and raphe nuclei; these cells are inactive during REM sleep. A reciprocal interaction in the brain stem between REM-activating and REM-deactivating neurons is responsible for REM generation and maintenance.

The active hypnogenic neurons for non-REM sleep are located primarily in the preoptic area of the hypothalamus, the basal forebrain area, and the solitary nucleus region of the medulla. The reticular nucleus of the thalamus is thought to be responsible for sleep spindle generation.

Sleep most likely results from both passive and active mechanisms. Muscle hypotonia during REM sleep is thought to depend on the inhibitory postsynaptic potentials generated by dorsal pontine descending axons. The roles of neurotransmitters and neuropeptides are not clearly delineated; they are believed to modulate various sleep stages and cycles. It has been shown that adenosine may fulfil the major criteria for neural sleep factor, which mediates the somnogenic effects of prolonged wakefulness.

6. THE "BIOLOGICAL CLOCK" OF HYPOTHALAMUS

Many physiological functions, including body temperature and sleep/wake cycles, vary throughout the day in a pattern that repeats itself daily. Others, such as the female menstrual cycle, repeat themselves approximately every 28 days. Still others, such as reproductive function in seasonal breeders, repeat annually.

The hypothalamus is thought to play a major role in regulating all of these biological rhythms. Furthermore, these rhythms appear to be endogenous (within the body) because they persist even in the absence of time cues, such as day/night cycles for light and dark periods, lunar cycles for monthly rhythms, or changes in temperature and day length for seasonal change. Accordingly, most organisms, including humans, are said to possess an endogenous timekeeper, a so-called *'biological clock'* that times the body's regulated functions.

Most homeostatically regulated functions exhibit peaks and valleys of activity that recur approximately daily. These are called *circadian rhythms* or diurnal rhythms. The circadian rhythms of the body are driven by the suprachiasmatic nucleus (SCN), a center in the hypothalamus that serves as the brain's biological clock. The SCN, which influences many hypothalamic nuclei via its efferent connections, has the properties of an oscillator whose spontaneous firing patterns change dramatically during a day/night cycle.

This diurnal cycle of activity is maintained *in vitro* and is an internal property of SCN cells. The molecular basis of the cellular rhythm is a series of transcriptional/translational feedback loops. The genes involved in these loops are apparently conserved from prokaryotes to humans. An important pathway influencing the SCN is the afferent retinohypothalamic tract of the optic nerve, which originates in the retina and enters the brain through the optic chiasm and terminates in the SCN. This pathway is the principal means by which light signals from the outside world transmit the day/night rhythm to the brain's internal clock, thereby entraining the endogenous oscillator to the external clock.

Fig. 41–3 illustrates some of the circadian rhythms of the body. One of the most vivid is alertness, which peaks in the afternoon and is lowest in the hours preceding and following sleep. Another, body temperature, ranges approximately 1°C throughout the day, with the low point occurring during sleep. The mechanism by which the SCN can regulate diverse functions is related to its control of the production of *melatonin* by the pineal gland. Melatonin levels increase with decreasing light as night ensues. Other homeostatically regulated functions exhibit diurnal patterns as well; when they are all in synchrony, they function harmoniously and impart a feeling of well-being.

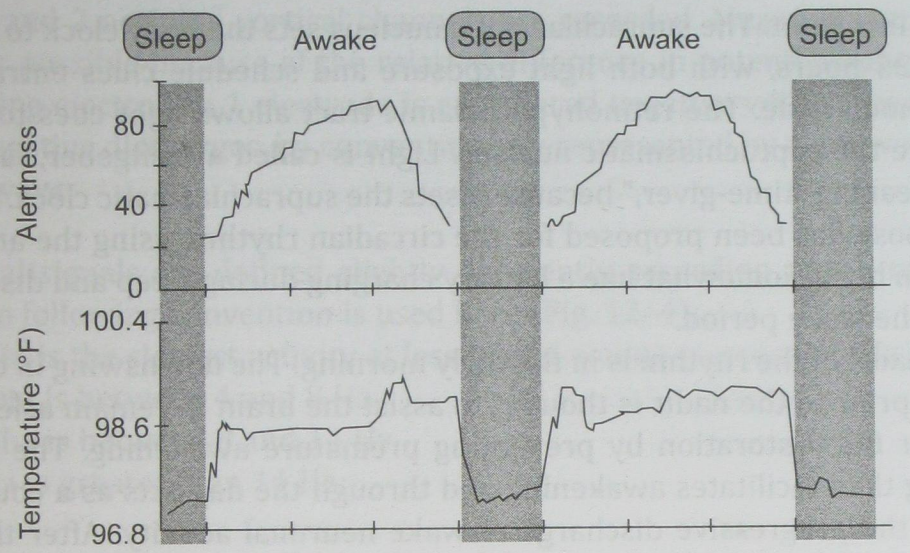


Fig. 41–3. Circadian rhythms in some homeostatically regulated functions during two 24-hour periods. Alertness is measured on an arbitrary scale between sleep and most alert. (From Coleman RM. Wide Awake at 3 : 00 AM. New York: WH Freeman, 1986).

When there is a disruption in rhythmic pattern, such as by sleep deprivation or when passing too rapidly through several time zones, the period required for re-entrainment of the SCN to the new day/night pattern is characterized by a feeling of malaise and physiological distress. This is commonly experienced as jet lag in travellers crossing several time zones or by workers changing from day shift to night shift or from night shift to day shift. In such cases, the hypothalamus requires time to "reset its clock" before the regular rhythms are restored and a feeling of well-being ensues.

The SCN uses the new pattern of light/darkness, as perceived in the retina, to entrain its firing rate to a pattern consistent with the external world. Resetting the clock maybe facilitated by the judicious use of exogenous melatonin and by altering exposure to light.

7. CIRCADIAN RHYTHMS THAT INFLUENCE SLEEP

Circadian sleep rhythm is 1 of the several intrinsic body rhythms modulated by the hypothalamus. The existence of circadian rhythms has been known since 1731, when the French astronomer J. J. D. de Mairan noted a circadian

rhythm in a plant. The suprachiasmatic nucleus sets the body clock to approximately 25 hours, with both light exposure and schedule clues entraining to the 24-hour cycle. The retinohypothalamic tract allows light cues to directly influence the suprachiasmatic nucleus. Light is called a "zeitgeber," a German word meaning "time-giver," because it sets the suprachiasmatic clock. A practical purpose has been proposed for the circadian rhythm, using the analogy of the brain being somewhat like a battery charging during sleep and discharging during the wake period.

The nadir of the rhythm is in the early morning. The downswing in circadian rhythm prior to the nadir is thought to assist the brain to remain asleep overnight for full restoration by preventing premature awakening. The morning upswing then facilitates awakening and through the day acts as a counterbalance to the progressive discharge of wake neuronal activity. After the circadian apex in the early evening, the downswing aids sleep initiation. This model explains the relatively steady cognitive function throughout wakefulness.

Body temperature cycles are also under hypothalamic control. An increase in body temperature is seen during the course of the day and a decrease is observed during the night. The temperature peaks and troughs are thought to mirror the sleep rhythm. People who are alert late in the evening (i.e, evening types) have body temperature peaks late in the evening, while those who find themselves most alert early in the morning (i.e, morning types) have body temperature peaks early in the evening.

Melatonin has been implicated as a modulator of light entrainment. It is secreted maximally during the night by the pineal gland. Prolactin, testosterone, and growth hormone also demonstrate circadian rhythms, with maximal secretion during the night.

8. SLEEP AND THE EEG

Sleep is regulated by the reticular formation. The ascending reticular activating system is periodically shut down by influences from other regions of the reticular formation. The EEG recorded during sleep reveals a persistently changing pattern of wave amplitudes and frequencies, indicating that the brain remains continually active even in the deepest stages of sleep.

The EEG signal is of primary importance in interpreting polysomnographic studies. It records electric potentials generated by the interaction between the cortex and the deeper brain structures, especially the thalamus. Two centro-

cephalic and 2 occipital cortical channels are recorded. Measurement of EEG signals is possible because of the relative difference in potential between the 2 recording electrodes; 1 electrode is considered negative with respect to the other. Negative discharges, by convention, are represented by an upwardly deflecting wave.

Cortical signals are defined slightly differently according to the reference used. The following convention is used here (Fig. 12–4):

- Delta is the slowest activity at less than 4 counts per second (Hz).
- Theta is between 4 and 8 Hz.
- Alpha is between 8 and 14 Hz.
- Beta is greater than 14 Hz.

Another range occasionally mentioned is gamma, which is part of the high end of the beta frequency and has been described to range from 30-45 Hz.

It should be mentioned also that both muscle functional activity (EMG, electromyography) and eye movements (EOG, electrooculography) are recommended to use for the proper sleep stages analysis. The EMG signals are muscle twitch potentials that are of secondary importance in polysomnography. Their utilization is based on the finding that, during sleep, muscle activity decreases. During rapid eye movement sleep, muscle activity is at its nadir. However, in many cases appreciating the decreasing tone is difficult. The relative silence during REM sleep may not be of help in distinguishing REM sleep from the preceding or subsequent sleep stages.

The EOG signals measure changes in the electric potential of the positive anterior aspect of the eye relative to the negative posterior aspect. Horizontal axis electrodes are placed near the outer canthi and vertical axis electrodes below and above the eye to measure transient changes in potential during the actual eye movement. During any eye movement, the cornea (positive) moves toward 1 electrode, while the fundus (negative) moves away. When the eye is not moving, the change in relative position is zero, and the eye leads do not record a signal. Slow, rolling eye movements are recorded as long gentle waves, while rapid jerking movements are represented by sharply contoured fast waves. Blinking of the eyes produces rapid vertical movements. Eye movements during drowsiness and stage I sleep may be jerky, irregular, or gently rolling. In deeper stages of sleep, macro eye movements cease altogether. During REM sleep, eye movements again become active and jerky. The intensity of the bursts of activity is used to describe the density of REM sleep.

The EEG pattern recorded during sleep varies in a cyclic fashion that repeats approximately every 90 minutes, starting from the time of falling asleep to awakening 7 to 8 hrs later (Fig. 41–1). These cycles are associated with two different forms of sleep, which follow each other sequentially:

1. Slow-wave sleep: four stages of progressively deepening sleep (i.e., it becomes harder to wake the subject)

2. REM sleep: back-and-forth movements of the eyes under closed lids, accompanied by autonomic excitation EEG recordings of sleeping subjects in laboratory settings reveal that the brain's electrical activity varies as the subject passes through cycles of slow-wave sleep, then REM sleep, on through the night.

A normal sleep cycle begins with slow-wave sleep, four stages of increasingly deep sleep during which the EEG becomes progressively slower in frequency and higher in amplitude.

The first several epochs of the record will be the wake stage. The EEG will show mixed beta and alpha activities as the eyes open and close and predominantly alpha activity when the eyes remain closed. The EMG will reflect the high-amplitude muscle contractions and movement artifacts. The EOG will show eye blinking and rapid movement. The record will slow in frequency and amplitude as the subject stops moving and becomes drowsy.

Stage I sleep is scored when the alpha activity in the EEG drops to less than 50 %. A transition is observed from alpha activity to a lower frequency activity, such as theta, possibly intermixed with low-amplitude delta activity. Amplitudes are less than 50–75 μV. Paroxysms of 2-7 per s activity up to 75 μV may occur.

Stage I is usually brief, lasting for 1–7 minutes. Vertex sharp waves may occur, but no sleep spindles or K complexes are recorded. The EOG may show slow, rolling eye movements, especially early in the stage. No REMs are observed. The EMG shows less activity than in wake stage, but the transition is gradual and of little assistance in scoring.

Arousals are paroxysms of activity lasting 3 seconds or longer. The minimum arousal is simply a paroxysmal burst in the EEG channel, usually to alpha or theta activity. Arousal from stage I is common and usually is represented by a burst of activity on the EEG, EOG, and EMG. If the burst results in alpha activity for greater than 50 % of the record, then the epoch is scored as wake.

Stage II sleep. The EEG shows predominant theta activity with minimal alpha activity. Delta is permitted for less than 20 % of the record. Amplitude may increase from that seen in stage I.

K-complexes appear for the first time. K-complexes are sharply negative (i.e, updeflecting) monophasic or polyphasic waves followed by a slower, positive (i.e, downdeflecting) wave. The complex must persist for at least 0.5 seconds. No minimum amplitude is defined, but characteristically the waves stand out clearly from the background. K complexes can occur in response to a sudden sound and were so named because they were appreciated as following the knocking sound produced by knuckle rapping. In this respect, they may represent a form of cortical evoked potential in a brain still minimally responsive to external stimuli.

Sleep spindles may appear. These are paroxysms of 12-14 per s activity persisting for at least 0.5 seconds (that is, 6-7 small waves in 0.5 seconds). Although classically described as spindle shaped, they vary in morphology and may attach as a tail to a K complex.

No specific criteria exist for EOG and EMG in this stage.

Arousal from stage II may be into stage I or into wakefulness. If the arousal is 3 seconds or longer in duration, and the resulting alpha activity persists for less than 50 % of the record, the epoch is scored as stage I. If the alpha persists for greater than 50 % of the record, the epoch is scored as stage wake. If the first half of the following epoch demonstrates stage II characteristics (i.e., spindles, K- complexes, high-amplitude theta/delta activity), that epoch is scored as stage II.

Once in stage II, that score is maintained unless a reason to exit presents. One such reason to exit is described as the 3-minute rule. If no specific stage II indicators appear, and in the absence of arousals and muscle tone changes that would alter the staging, continue to score all epochs as stage II for up to 3 minutes. At 3 minutes, if no specific indicators for stage II have occurred, scroll back 3 minutes and score those epochs as stage I.

Stage III sleep. The EEG shows 4 cps or slower activity, with peak-to-peak amplitudes greater than 75 μV for between 20 % and 50 % of the epoch. Both K complexes and sleep spindles may be seen in stage III sleep. No specific criteria exist for EOG and EMG. The transition to stage III from stage II may be gradual, and stage III may alternate with stage I.

Stage IV sleep is reached at the end of about an hour, when delta waves are observed. The EEG shows 4 cps or slower activity, with peak-to-peak amplitudes greater than 75 μV for at least 50 % of the epoch. Both K complexes and sleep spindles may be seen in stage IV sleep. No specific criteria for EOG and EMG exist. In general, muscle tone decreases gradually from stage II to stage IV. The transition from stage III to stage IV may be gradual, and stage III may alternate with stage IV.

The subject then passes through the same stages in reverse order, approaching stage 1 by about 90 minutes, when a REM period begins, followed by a new cycle of slow-wave sleep. Slow-wave sleep is characterized by decreased heart rate and blood pressure, slow and regular breathing, and relaxed muscle tone. Stages 3 and 4 occur only in the first few sleep cycles of the night. In contrast, REM periods increase in duration with each successive cycle, so that the last few cycles consist of approximately equal periods of REM sleep and stage 2 slow-wave sleep.

Stage V sleep or **REM sleep** is also known as **paradoxical sleep,** because of the seeming contradictions in its characteristics:

1. First, the EEG exhibits unsynchronised, high-frequency, low amplitude waves (i.e., a beta rhythm), which is more typical of the awake state than sleep, yet the subject is as difficult to arouse as when in stage 4 slow-wave sleep.
2. The EOG of REM shows paroxysmal, relatively sharply contoured, high-amplitude activity occurring in all eye leads simultaneously. The EOG activity is not needed to mark the start of an REM period. REM epochs may be recognized by EEG activity before EOG movements start. Small REMs on EOG may serve as a harbinger of REM stage and can indicate the actual onset of REM in another area where interscorer concordance is lower.
3. The autonomic nervous system is in a state of excitation; blood pressure and heart rate are increased and breathing is irregular. The increasiness of the cardiac and cerebral blood supply.
4. In males, autonomic excitation in REM sleep includes penile erection. This reflex is used in diagnosing impotence, to determine whether erectile failure is based on a neurological or a vascular defect (in which case, erection does not accompany REM sleep).
5. When subjects are awakened during a REM period, they usually report dreaming. Accordingly, it is customary to consider REM sleep as dream sleep.
6. The increasiness of the muscle tone. It is well known that some types of the epileptiform disorders are closely associated with this stage of sleep.
7. Another curious characteristic of REM sleep is that most voluntary muscles are temporarily paralyzed. Two exceptions, in addition to the muscles of respiration, include the extraocular muscles, which contract rhythmically to produce the rapid eye movements, and the muscles of the middle ear, which protect the inner ear. Muscle paralysis is caused by an active inhibition of motor neurons mediated by a group of neurons located close to the locus ceruleus in the brainstem. Many of us have

experienced this muscle paralysis on waking from a bad dream, feeling momentarily incapable of running from danger. In certain sleep disorders in which skeletal muscle contraction is not temporarily paralyzed in REM sleep, subjects act out dream sequences with disturbing results, with no conscious awareness of this happening.

9. DREAMING

The demonstrably necessary phenomenon of dreaming would suffice to prove the importance of sleep to humans, and perhaps to other animals as well. Dreaming involves an involuntary conjuring up of images in a sequence in which the sleeper/dreamer is usually more a participant than an observer.

Most scientists agree that dreaming is stimulated by the pons and occurs during the REM phase of sleep.

Many functions have been hypothesized for dreaming. Freud postulated that dreams are the symbolic expression of frustrated desires that had been relegated to the subconscious, and used dream interpretation in the form of psychoanalysis he pioneered. Scientists today have generally become more skeptical about details of Freudian interpretation, and place more emphasis on dreaming as a requirement for organization and consolidation of recent memory and experience.

Another theory is that dreaming allows an animal to play out scenarios that may help the animal avoid dangers when awake. For example, a rabbit might dream about being cornered by a fox and may play out different scenarios that might increase chances of survival should he come across a fox in reality.

10. NORMAL SLEEP IN ADULTS, INFANTS AND THE ELDERLY

Sleep requirements change dramatically from infancy to old age. Newborns sleep approximately 16 hrs per day, of which about 50 % is spent in REM sleep.

Infants have an overall greater total sleep time than any other age group; their sleep time can be divided into multiple periods. In newborns, the total sleep duration in a day can be 14-16 hrs. Over the first several months of life, sleep time decreases; by age 5-6 months, sleep consolidates into an overnight period with at least 1 nap during the day. REM sleep in infants represents a larger percentage of the total sleep at the expense of stages III and IV. Until age

3–4 months, newborns transition from wake into REM sleep. Thereafter, wake begins to transition directly into non-REM.

Overall, electrocortical recorded voltage remains high during sleep, as it does during periods of wakefulness. Sleep spindles begin appearing in the second month of life with a density greater than that seen in adults. After the first year, the spindles begin decreasing in density and progress toward adult patterns. K complexes begin by the sixth month of life.

Normal adults sleep 7 to 8 hrs per day, of which about 25 % is spent in REM sleep. The percentage of REM sleep declines further with age, together with a loss of the ability to achieve stages 3 and 4 of slow-wave sleep. In adults, sleep of 8-8.4 hrs is considered fully restorative. In some cultures, total sleep often is divided into an overnight sleep period of 6-7 hrs and a midafternoon nap of 1–2 hrs.

Stage I is considered a transition between wake and sleep. It occurs upon falling asleep and during brief arousal periods within sleep and usually accounts for 5-10 % of total sleep time. Stage II occurs throughout the sleep period and represents 40-50 % of total sleep time. Stages III and IV delta sleep occur mostly in the first third of the night. They are distinguished from each other only by the percentage of delta activity and represent up to 20 % of total sleep time. REM represents 20-25 % of total sleep time.

In elderly persons, the time spent in stages III and IV sleep decreases by 10- 15 % and the time in stage II increases by 5 % compared to young adults, representing an overall decrease in total sleep duration. Latency to fall asleep and the number and duration of overnight arousal periods increase. Thus, to have a fully restorative sleep, the total time in bed must increase. If the elderly person does not increase the total time in bed, complaints of insomnia and chronic sleepiness may occur.

Sleep fragmentation results from the increase in overnight arousals and may be exacerbated by the increasing number of geriatric medical conditions, including sleep apnea, musculoskeletal disorders, and cardiopulmonary disease.

NOTES NOTES NOTES NOTES

Підручник з фізіології людини, викладений англійською мовою для кращого опанування англомовни студентами знань з фізіології, містить основний навчальний матеріал за програмою курсу фізіології. Видан містить інформацію про всі розділи дисципліни, зважаючи на основні вимоги кредитно-модульної систе навчання стосовно покращення самостійного вивчення навчального матеріалу. Разом із висвітленням основн механізмів функціонування органів та систем органів, принципів нервово-гуморальної регуляції функ організму людини за умов норми, адаптаційно-пристосувальних механізмів за умов дії стресорних чи інш чинників, у підручнику наведені початкові дані про порушення функціонування кожного з органів.

Чітка, обдумана рубрифікація змісту кожного розділу з виділенням найважливіших деталей кожної те чітке формулювання основних положень, визначень та класифікацій із додатковим винесенням матеріалу лек в рисунки та схеми дозволяють студентам легко з'ясувати головні аспекти кожної з тем.

Підручник може бути також корисний для викладачів, які працюють в навчальних медичних заклад з англомовними студентами.

Навчальне видання

Мороз Василь Максимович
Шандра Олексій Антонович
Вастьянов Руслан Сергійович та ін.

Фізіологія

Підручник (англ. мовою)

Редактор *О. В. Марчук*
Коректор *О. В. Рябоконь*
Технічний редактор *Ж. С. Швець*
Комп'ютерна верстка: *О. С. Парфенюк*

Підписано до друку 05.08.15. Формат 60×84/16. Папір офсетний.
Гарнітура Камбріа. Друк офсетний.Ум. друк. арк. 42,46.
Тираж 300 пр. Зам. № 823.

ПП "Нова Книга"
21029, м. Вінниця, вул. Квятека, 20
Свідоцтво про внесення до державного реєстру видавців,
виготівників і розповсюджувачів видавничої продукції
ДК № 2646 від 11.10.2006 р.
Тел. (0432) 52-34-80, 52-34-82 Факс 52-34-81
E-mail: info@novaknyha.com.ua
www.novaknyha.com.ua

Надруковано у ПП "Юнісофт"
61036, м. Харків, вул. Морозова, 13б
www.ttornado.com.ua