Functional Organization of the Human Body and Control of the "Internal Environment"

The goal of physiology is to explain the physical and chemical factors that are responsible for the origin, development, and progression of life. Each type of life, from the simple virus to the largest tree or the complicated human being, has its own functional characteristics. Therefore, the vast field of physiology can be divided into *viral physiology, bacterial physiology, cellular physiology, plant physiology, human physiology,* and many more subdivisions.

Human Physiology. In *human physiology,* we attempt to explain the specific characteristics and mechanisms of the human body that make it a living being. The very fact that we remain alive is the result of complex control systems, for hunger makes us seek food and fear makes us seek refuge. Sensations of cold make us look for warmth. Other forces cause us to seek fellowship and to reproduce. Thus, the human being is, in many ways, like an automaton, and the fact that we are sensing, feeling, and knowledgeable beings is part of this automatic sequence of life; these special attributes allow us to exist under widely varying conditions.

Cells as the Living Units of the Body

The basic living unit of the body is the cell. Each organ is an aggregate of many different cells held together by intercellular supporting structures.

Each type of cell is specially adapted to perform one or a few particular functions. For instance, the red blood cells, numbering 25 trillion in each human being, transport oxygen from the lungs to the tissues. Although the red cells are the most abundant of any single type of cell in the body, there are about 75 trillion additional cells of other types that perform functions different from those of the red cell. The entire body, then, contains about 100 trillion cells.

Although the many cells of the body often differ markedly from one another, all of them have certain basic characteristics that are alike. For instance, in all cells, oxygen reacts with carbohydrate, fat, and protein to release the energy required for cell function. Further, the general chemical mechanisms for changing nutrients into energy are basically the same in all cells, and all cells deliver end products of their chemical reactions into the surrounding fluids.

Almost all cells also have the ability to reproduce additional cells of their own kind. Fortunately, when cells of a particular type are destroyed, the remaining cells of this type usually generate new cells until the supply is replenished.

Extracellular Fluid—The "Internal Environment"

About 60 percent of the adult human body is fluid, mainly a water solution of ions and other substances. Although most of this fluid is inside the cells and is called *intracellular fluid,* about one third is in the spaces outside the cells and is called *extracellular fluid.* This extracellular fluid is in constant motion throughout the body. It is transported rapidly in the circulating blood and then mixed between the blood and the tissue fluids by diffusion through the capillary walls.

In the extracellular fluid are the ions and nutrients needed by the cells to maintain cell life. Thus, all cells live in essentially the same environment—the extracellular fluid. For this reason, the extracellular fluid is also called the *internal environment* of the body, or the *milieu intérieur,* a term introduced more than 100 years ago by the great 19th-century French physiologist Claude Bernard.

Cells are capable of living, growing, and performing their special functions as long as the proper concentrations of oxygen, glucose, different ions, amino acids, fatty substances, and other constituents are available in this internal environment.

Differences Between Extracellular and Intracellular Fluids. The extracellular fluid contains large amounts of *sodium, chloride,* and *bicarbonate ions* plus nutrients for the cells, such as *oxygen, glucose, fatty acids,* and *amino acids.* It also contains *carbon dioxide* that is

being transported from the cells to the lungs to be excreted, plus other cellular waste products that are being transported to the kidneys for excretion.

The intracellular fluid differs significantly from the extracellular fluid; for example, it contains large amounts of *potassium, magnesium,* and *phosphate ions* instead of the sodium and chloride ions found in the extracellular fluid. Special mechanisms for transporting ions through the cell membranes maintain the ion concentration differences between the extracellular and intracellular fluids. These transport processes are discussed in Chapter 4.

"Homeostatic" Mechanisms of the Major Functional Systems

Homeostasis

The term *homeostasis* is used by physiologists to mean *maintenance of nearly constant conditions in the internal environment.* Essentially all organs and tissues of the body perform functions that help maintain these relatively constant conditions. For instance, the lungs provide oxygen to the extracellular fluid to replenish the oxygen used by the cells, the kidneys maintain constant ion concentrations, and the gastrointestinal system provides nutrients.

A large segment of this text is concerned with the manner in which each organ or tissue contributes to homeostasis. To begin this discussion, the different functional systems of the body and their contributions to homeostasis are outlined in this chapter; then we briefly outline the basic theory of the body's control systems that allow the functional systems to operate in support of one another.

Extracellular Fluid Transport and Mixing System—The Blood Circulatory System

Extracellular fluid is transported through all parts of the body in two stages. The first stage is movement of blood through the body in the blood vessels, and the second is movement of fluid between the blood capillaries and the *intercellular spaces* between the tissue cells.

Figure 1-1 shows the overall circulation of blood. All the blood in the circulation traverses the entire circulatory circuit an average of once each minute when the body is at rest and as many as six times each minute when a person is extremely active.

As blood passes through the blood capillaries, continual exchange of extracellular fluid also occurs between the plasma portion of the blood and the interstitial fluid that fills the intercellular spaces. This process is shown in Figure 1-2. The walls of the capillaries are permeable to most molecules in the plasma of the blood, with the exception of plasma protein molecules, which are too large to readily pass through the capillaries. Therefore, large amounts of fluid and its dissolved constituents *diffuse* back and forth between the blood and the tissue spaces, as shown by the arrows. This process of diffusion is caused by kinetic motion of the molecules in both the plasma and the interstitial fluid. That is, the fluid and dissolved molecules are continually moving and bouncing in all directions within the plasma and the fluid in the intercellular spaces, as well as through the capillary pores.

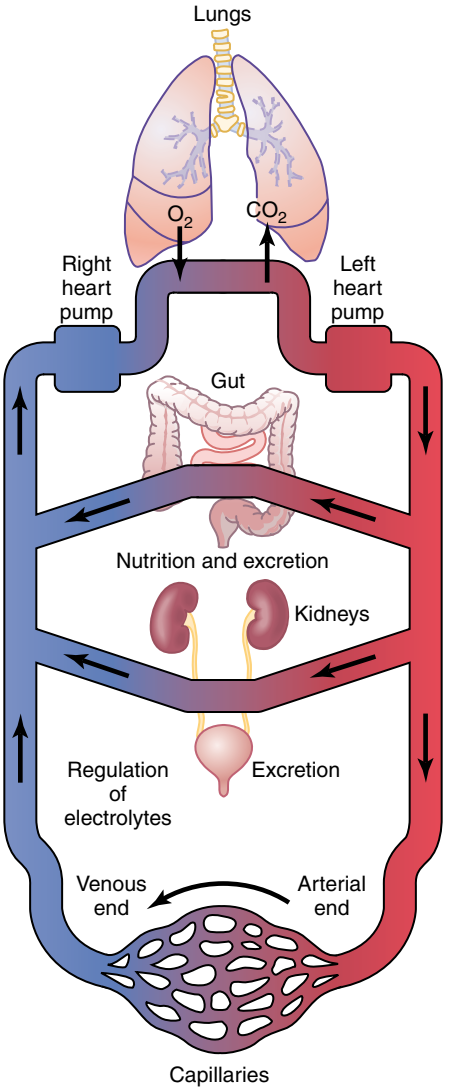


Figure 1-1 General organization of the circulatory system.

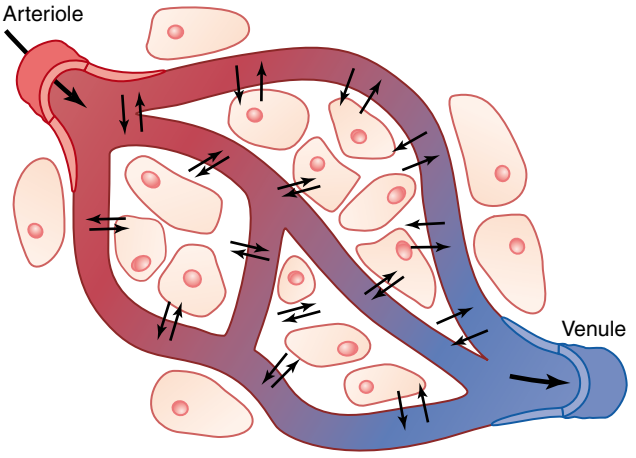


Figure 1-2 Diffusion of fluid and dissolved constituents through the capillary walls and through the interstitial spaces.

Few cells are located more than 50 micrometers from a capillary, which ensures diffusion of almost any substance from the capillary to the cell within a few seconds. Thus, the extracellular fluid everywhere in the body—both that of the plasma and that of the interstitial fluid—is continually being mixed, thereby maintaining homogeneity of the extracellular fluid throughout the body.

Origin of Nutrients in the Extracellular Fluid

Respiratory System. Figure 1-1 shows that each time the blood passes through the body, it also flows through the lungs. The blood picks up oxygen in the alveoli, thus acquiring the *oxygen* needed by the cells. The membrane between the alveoli and the lumen of the pulmonary capillaries, the *alveolar membrane,* is only 0.4 to 2.0 micrometers thick, and oxygen rapidly diffuses by molecular motion through this membrane into the blood.

Gastrointestinal Tract. A large portion of the blood pumped by the heart also passes through the walls of the gastrointestinal tract. Here different dissolved nutrients, including *carbohydrates, fatty acids,* and *amino acids,* are absorbed from the ingested food into the extracellular fluid of the blood.

Liver and Other Organs That Perform Primarily Metabolic Functions. Not all substances absorbed from the gastrointestinal tract can be used in their absorbed form by the cells. The liver changes the chemical compositions of many of these substances to more usable forms, and other tissues of the body—fat cells, gastrointestinal mucosa, kidneys, and endocrine glands—help modify the absorbed substances or store them until they are needed. The liver also eliminates certain waste products produced in the body and toxic substances that are ingested.

Musculoskeletal System. How does the musculoskeletal system contribute to homeostasis? The answer is obvious and simple: Were it not for the muscles, the body could not move to the appropriate place at the appropriate time to obtain the foods required for nutrition. The musculoskeletal system also provides motility for protection against adverse surroundings, without which the entire body, along with its homeostatic mechanisms, could be destroyed instantaneously.

Removal of Metabolic End Products

Removal of Carbon Dioxide by the Lungs. At the same time that blood picks up oxygen in the lungs, *carbon dioxide* is released from the blood into the lung alveoli; the respiratory movement of air into and out of the lungs carries the carbon dioxide to the atmosphere. Carbon dioxide is the most abundant of all the end products of metabolism.

Kidneys. Passage of the blood through the kidneys removes from the plasma most of the other substances besides carbon dioxide that are not needed by the cells. These substances include different end products of cellular metabolism, such as urea and uric acid; they also include excesses of ions and water from the food that might have accumulated in the extracellular fluid.

The kidneys perform their function by first filtering large quantities of plasma through the glomeruli into the tubules and then reabsorbing into the blood those substances needed by the body, such as glucose, amino acids, appropriate amounts of water, and many of the ions. Most of the other substances that are not needed by the body, especially the metabolic end products such as urea, are reabsorbed poorly and pass through the renal tubules into the urine.

Gastrointestinal Tract. Undigested material that enters the gastrointestinal tract and some waste products of metabolism are eliminated in the feces.

Liver. Among the functions of the liver is the detoxification or removal of many drugs and chemicals that are ingested. The liver secretes many of these wastes into the bile to be eventually eliminated in the feces.

Regulation of Body Functions

Nervous System. The nervous system is composed of three major parts: the *sensory input portion,* the *central nervous system* (or *integrative portion*), and the *motor output portion.* Sensory receptors detect the state of the body or the state of the surroundings. For instance, receptors in the skin apprise one whenever an object touches the skin at any point. The eyes are sensory organs that give one a visual image of the surrounding area. The ears are also sensory organs. The central nervous system is composed of the brain and spinal cord. The brain can store information, generate thoughts, create ambition, and determine reactions that the body performs in response to the sensations. Appropriate signals are then transmitted through the motor output portion of the nervous system to carry out one's desires.

An important segment of the nervous system is called the *autonomic system.* It operates at a subconscious level and controls many functions of the internal organs, including the level of pumping activity by the heart, movements of the gastrointestinal tract, and secretion by many of the body's glands.

Hormone Systems. Located in the body are eight major *endocrine glands* that secrete chemical substances called *hormones.* Hormones are transported in the extracellular fluid to all parts of the body to help regulate cellular function. For instance, *thyroid hormone* increases the rates of most chemical reactions in all cells, thus helping to set the tempo of bodily activity. *Insulin* controls glucose metabolism; *adrenocortical hormones* control sodium ion, potassium ion, and protein metabolism; and *parathyroid hormone* controls bone calcium and phosphate. Thus, the hormones provide a system for regulation that complements the nervous system. The nervous

system regulates many muscular and secretory activities of the body, whereas the hormonal system regulates many metabolic functions.

Protection of the Body

Immune System. The immune system consists of the white blood cells, tissue cells derived from white blood cells, the thymus, lymph nodes, and lymph vessels that protect the body from pathogens such as bacteria, viruses, parasites, and fungi. The immune system provides a mechanism for the body to (1) distinguish its own cells from foreign cells and substances and (2) destroy the invader by *phagocytosis* or by producing *sensitized lymphocytes* or specialized proteins (e.g., *antibodies)* that either destroy or neutralize the invader.

Integumentary System. The skin and its various appendages, including the hair, nails, glands, and other structures, cover, cushion, and protect the deeper tissues and organs of the body and generally provide a boundary between the body's internal environment and the outside world. The integumentary system is also important for temperature regulation and excretion of wastes and it provides a sensory interface between the body and the external environment. The skin generally comprises about 12 to 15 percent of body weight.

Reproduction

Sometimes reproduction is not considered a homeostatic function. It does, however, help maintain homeostasis by generating new beings to take the place of those that are dying. This may sound like a permissive usage of the term *homeostasis,* but it illustrates that, in the final analysis, essentially all body structures are organized such that they help maintain the automaticity and continuity of life.

Control Systems of the Body

The human body has thousands of control systems. The most intricate of these are the genetic control systems that operate in all cells to help control intracellular function and extracellular functions. This subject is discussed in Chapter 3.

Many other control systems operate *within the organs* to control functions of the individual parts of the organs; others operate throughout the entire body *to control the interrelations between the organs.* For instance, the respiratory system, operating in association with the nervous system, regulates the concentration of carbon dioxide in the extracellular fluid. The liver and pancreas regulate the concentration of glucose in the extracellular fluid, and the kidneys regulate concentrations of hydrogen, sodium, potassium, phosphate, and other ions in the extracellular fluid.

Examples of Control Mechanisms

Regulation of Oxygen and Carbon Dioxide Concentrations in the Extracellular Fluid. Because oxygen is one of the major substances required for chemical reactions in the cells, the body has a special control mechanism to maintain an almost exact and constant oxygen concentration in the extracellular fluid. This mechanism depends principally on the chemical characteristics of *hemoglobin,* which is present in all red blood cells. Hemoglobin combines with oxygen as the blood passes through the lungs. Then, as the blood passes through the tissue capillaries, hemoglobin, because of its own strong chemical affinity for oxygen, does not release oxygen into the tissue fluid if too much oxygen is already there. But if the oxygen concentration in the tissue fluid is too low, sufficient oxygen is released to re-establish an adequate concentration. Thus, regulation of oxygen concentration in the tissues is vested principally in the chemical characteristics of hemoglobin itself. This regulation is called the *oxygen-buffering function of hemoglobin.*

Carbon dioxide concentration in the extracellular fluid is regulated in a much different way. Carbon dioxide is a major end product of the oxidative reactions in cells. If all the carbon dioxide formed in the cells continued to accumulate in the tissue fluids, all energy-giving reactions of the cells would cease. Fortunately, a higher than normal carbon dioxide concentration in the blood *excites the respiratory center,* causing a person to breathe rapidly and deeply. This increases expiration of carbon dioxide and, therefore, removes excess carbon dioxide from the blood and tissue fluids. This process continues until the concentration returns to normal.

Regulation of Arterial Blood Pressure. Several systems contribute to the regulation of arterial blood pressure. One of these, the *baroreceptor system,* is a simple and excellent example of a rapidly acting control mechanism. In the walls of the bifurcation region of the carotid arteries in the neck, and also in the arch of the aorta in the thorax, are many nerve receptors called *baroreceptors,* which are stimulated by stretch of the arterial wall. When the arterial pressure rises too high, the baroreceptors send barrages of nerve impulses to the medulla of the brain. Here these impulses inhibit the *vasomotor center,* which in turn decreases the number of impulses transmitted from the vasomotor center through the sympathetic nervous system to the heart and blood vessels. Lack of these impulses causes diminished pumping activity by the heart and also dilation of the peripheral blood vessels, allowing increased blood flow through the vessels. Both of these effects decrease the arterial pressure back toward normal.

Conversely, a decrease in arterial pressure below normal relaxes the stretch receptors, allowing the vasomotor center to become more active than usual, thereby causing vasoconstriction and increased heart pumping. The decrease in arterial pressure also raises arterial pressure back toward normal.

Normal Ranges and Physical Characteristics of Important Extracellular Fluid Constituents

Table 1-1 lists some of the important constituents and physical characteristics of extracellular fluid, along with their normal values, normal ranges, and maximum limits without causing death. Note the narrowness of the normal range for each one. Values outside these ranges are usually caused by illness.

Most important are the limits beyond which abnormalities can cause death. For example, an increase in the body temperature of only 11°F (7°C) above normal can lead to a vicious cycle of increasing cellular metabolism that destroys the cells. Note also the narrow range for acid-base balance in the body, with a normal pH value of 7.4 and lethal values only about 0.5 on either side of normal. Another important factor is the potassium ion concentration because whenever it decreases to less than one-third normal, a person is likely to be paralyzed as a result of the nerves' inability to carry signals. Alternatively, if the potassium ion concentration increases to two or more times normal, the heart muscle is likely to be severely depressed. Also, when the calcium ion concentration falls below about one-half normal, a person is likely to experience tetanic contraction of muscles throughout the body because of the spontaneous generation of excess nerve impulses in the peripheral nerves. When the glucose concentration falls below one-half normal, a person frequently develops extreme mental irritability and sometimes even convulsions.

These examples should give one an appreciation for the extreme value and even the necessity of the vast numbers of control systems that keep the body operating in health; in the absence of any one of these controls, serious body malfunction or death can result.

Characteristics of Control Systems

The aforementioned examples of homeostatic control mechanisms are only a few of the many thousands in the body, all of which have certain characteristics in common as explained in this section.

Negative Feedback Nature of Most Control Systems

Most control systems of the body act by *negative feedback*, which can best be explained by reviewing some of the homeostatic control systems mentioned previously. In the regulation of carbon dioxide concentration, a high concentration of carbon dioxide in the extracellular fluid increases pulmonary ventilation. This, in turn, decreases the extracellular fluid carbon dioxide concentration because the lungs expire greater amounts of carbon dioxide from the body. In other words, the high concentration of carbon dioxide initiates events that decrease the concentration toward normal, which is *negative* to the initiating stimulus. Conversely, if the carbon dioxide concentration falls too low, this causes feedback to increase the concentration. This response is also negative to the initiating stimulus.

In the arterial pressure-regulating mechanisms, a high pressure causes a series of reactions that promote a lowered pressure, or a low pressure causes a series of reactions that promote an elevated pressure. In both instances, these effects are negative with respect to the initiating stimulus.

Therefore, in general, if some factor becomes excessive or deficient, a control system initiates *negative feedback*, which consists of a series of changes that return the factor toward a certain mean value, thus maintaining homeostasis.

"Gain" of a Control System. The degree of effectiveness with which a control system maintains constant conditions is determined by the *gain* of the negative feedback. For instance, let us assume that a large volume of blood is transfused into a person whose baroreceptor pressure control system is not functioning, and the arterial pressure rises from the normal level of 100 mm Hg up to 175 mm Hg. Then, let us assume that the same volume of blood is injected into the same person when the baroreceptor system is functioning, and this time the pressure increases only 25 mm Hg. Thus, the feedback control system has caused a "correction" of –50 mm Hg—that is, from

Table 1-1 Important Constituents and Physical Characteristics of Extracellular Fluid

	Normal Value	Normal Range	Approximate Short-Term Nonlethal Limit	Unit
Oxygen	40	35-45	10-1000	mm Hg
Carbon dioxide	40	35-45	5-80	mm Hg
Sodium ion	142	138-146	115-175	mmol/L
Potassium ion	4.2	3.8-5.0	1.5-9.0	mmol/L
Calcium ion	1.2	1.0-1.4	0.5-2.0	mmol/L
Chloride ion	108	103-112	70-130	mmol/L
Bicarbonate ion	28	24-32	8-45	mmol/L
Glucose	85	75-95	20-1500	mg/dl
Body temperature	98.4 (37.0)	98-98.8 (37.0)	65-110 (18.3-43.3)	°F (°C)
Acid-base	7.4	7.3-7.5	6.9-8.0	pH

175 mm Hg to 125 mm Hg. There remains an increase in pressure of +25 mm Hg, called the "error," which means that the control system is not 100 percent effective in preventing change. The gain of the system is then calculated by the following formula:

$$\text{Gain} = \frac{\text{Correction}}{\text{Error}}$$

Thus, in the baroreceptor system example, the correction is −50 mm Hg and the error persisting is +25 mm Hg. Therefore, the gain of the person's baroreceptor system for control of arterial pressure is −50 divided by +25, or −2. That is, a disturbance that increases or decreases the arterial pressure does so only one-third as much as would occur if this control system were not present.

The gains of some other physiologic control systems are much greater than that of the baroreceptor system. For instance, the gain of the system controlling internal body temperature when a person is exposed to moderately cold weather is about −33. Therefore, one can see that the temperature control system is much more effective than the baroreceptor pressure control system.

Positive Feedback Can Sometimes Cause Vicious Cycles and Death

One might ask the question, Why do most control systems of the body operate by negative feedback rather than positive feedback? If one considers the nature of positive feedback, one immediately sees that positive feedback does not lead to stability but to instability and, in some cases, can cause death.

Figure 1-3 shows an example in which death can ensue from positive feedback. This figure depicts the pumping effectiveness of the heart, showing that the heart of a healthy human being pumps about 5 liters of blood per minute. If the person is suddenly bled 2 liters, the amount of blood in the body is decreased to such a low level that

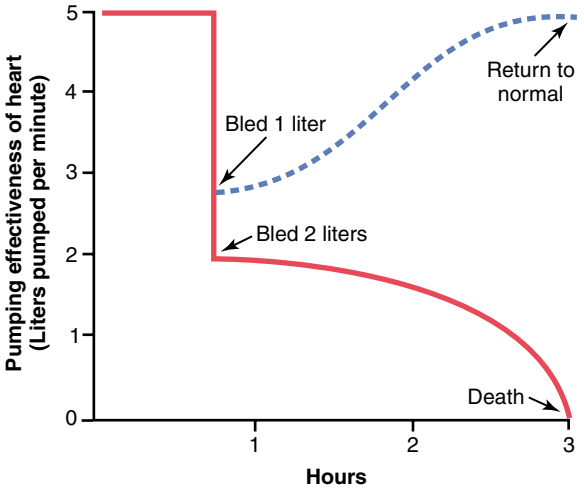


Figure 1-3 Recovery of heart pumping caused by *negative feedback* after 1 liter of blood is removed from the circulation. Death is caused by *positive feedback* when 2 liters of blood are removed.

not enough blood is available for the heart to pump effectively. As a result, the arterial pressure falls and the flow of blood to the heart muscle through the coronary vessels diminishes. This results in weakening of the heart, further diminished pumping, a further decrease in coronary blood flow, and still more weakness of the heart; the cycle repeats itself again and again until death occurs. Note that each cycle in the feedback results in further weakening of the heart. In other words, the initiating stimulus causes more of the same, which is *positive feedback.*

Positive feedback is better known as a "vicious cycle," but a mild degree of positive feedback can be overcome by the negative feedback control mechanisms of the body and the vicious cycle fails to develop. For instance, if the person in the aforementioned example were bled only 1 liter instead of 2 liters, the normal negative feedback mechanisms for controlling cardiac output and arterial pressure would overbalance the positive feedback and the person would recover, as shown by the dashed curve of Figure 1-3.

Positive Feedback Can Sometimes Be Useful. In some instances, the body uses positive feedback to its advantage. Blood clotting is an example of a valuable use of positive feedback. When a blood vessel is ruptured and a clot begins to form, multiple enzymes called *clotting factors* are activated within the clot itself. Some of these enzymes act on other unactivated enzymes of the immediately adjacent blood, thus causing more blood clotting. This process continues until the hole in the vessel is plugged and bleeding no longer occurs. On occasion, this mechanism can get out of hand and cause the formation of unwanted clots. In fact, this is what initiates most acute heart attacks, which are caused by a clot beginning on the inside surface of an atherosclerotic plaque in a coronary artery and then growing until the artery is blocked.

Childbirth is another instance in which positive feedback plays a valuable role. When uterine contractions become strong enough for the baby's head to begin pushing through the cervix, stretch of the cervix sends signals through the uterine muscle back to the body of the uterus, causing even more powerful contractions. Thus, the uterine contractions stretch the cervix and the cervical stretch causes stronger contractions. When this process becomes powerful enough, the baby is born. If it is not powerful enough, the contractions usually die out and a few days pass before they begin again.

Another important use of positive feedback is for the generation of nerve signals. That is, when the membrane of a nerve fiber is stimulated, this causes slight leakage of sodium ions through sodium channels in the nerve membrane to the fiber's interior. The sodium ions entering the fiber then change the membrane potential, which in turn causes more opening of channels, more change of potential, still more opening of channels, and so forth. Thus, a slight leak becomes an explosion of sodium entering the interior of the nerve fiber, which creates the nerve action potential. This action potential in turn causes electrical current to flow along both the outside and the inside

of the fiber and initiates additional action potentials. This process continues again and again until the nerve signal goes all the way to the end of the fiber.

In each case in which positive feedback is useful, the positive feedback itself is part of an overall negative feedback process. For example, in the case of blood clotting, the positive feedback clotting process is a negative feedback process for maintenance of normal blood volume. Also, the positive feedback that causes nerve signals allows the nerves to participate in thousands of negative feedback nervous control systems.

More Complex Types of Control Systems—Adaptive Control

Later in this text, when we study the nervous system, we shall see that this system contains great numbers of interconnected control mechanisms. Some are simple feedback systems similar to those already discussed. Many are not. For instance, some movements of the body occur so rapidly that there is not enough time for nerve signals to travel from the peripheral parts of the body all the way to the brain and then back to the periphery again to control the movement. Therefore, the brain uses a principle called *feed-forward control* to cause required muscle contractions. That is, sensory nerve signals from the moving parts apprise the brain whether the movement is performed correctly. If not, the brain corrects the feed-forward signals that it sends to the muscles the *next* time the movement is required. Then, if still further correction is necessary, this will be done again for subsequent movements. This is called *adaptive control.* Adaptive control, in a sense, is delayed negative feedback.

Thus, one can see how complex the feedback control systems of the body can be. A person's life depends on all of them. Therefore, a major share of this text is devoted to discussing these life-giving mechanisms.

Summary—Automaticity of the Body

The purpose of this chapter has been to point out, first, the overall organization of the body and, second, the means by which the different parts of the body operate in harmony. To summarize, the body is actually a *social order of about 100 trillion cells* organized into different functional structures, some of which are called *organs.* Each functional structure contributes its share to the maintenance of homeostatic conditions in the extracellular fluid, which is called the *internal environment.* As long as normal conditions are maintained in this internal environment, the cells of the body continue to live and function properly. Each cell benefits from homeostasis, and in turn, each cell contributes its share toward the maintenance of homeostasis. This reciprocal interplay provides continuous automaticity of the body until one or more functional systems lose their ability to contribute their share of function. When this happens, all the cells of the body suffer. Extreme dysfunction leads to death; moderate dysfunction leads to sickness.

Bibliography

Adolph EF: Physiological adaptations: hypertrophies and superfunctions, *Am Sci* 60:608, 1972.

Bernard C: *Lectures on the Phenomena of Life Common to Animals and Plants*, Springfield, IL, 1974, Charles C Thomas.

Cannon WB: *The Wisdom of the Body*, New York, 1932, WW Norton.

Chien S: Mechanotransduction and endothelial cell homeostasis: the wisdom of the cell, *Am J Physiol Heart Circ Physiol* 292:H1209, 2007.

Csete ME, Doyle JC: Reverse engineering of biological complexity, *Science* 295:1664, 2002.

Danzler WH, editor: *Handbook of Physiology, Sec 13: Comparative Physiology*, Bethesda, 1997, American Physiological Society.

DiBona GF: Physiology in perspective: the wisdom of the body. Neural control of the kidney, *Am J Physiol Regul Integr Comp Physiol* 289:R633, 2005.

Dickinson MH, Farley CT, Full RJ, et al: How animals move: an integrative view, *Science* 288:100, 2000.

Garland T Jr, Carter PA: Evolutionary physiology, *Annu Rev Physiol* 56:579, 1994.

Gao Q, Horvath TL: Neuronal control of energy homeostasis, *FEBS Lett* 582:132, 2008.

Guyton AC: *Arterial Pressure and Hypertension*, Philadelphia, 1980, WB Saunders.

Guyton AC, Jones CE, Coleman TG: *Cardiac Output and Its Regulation*, Philadelphia, 1973, WB Saunders.

Guyton AC, Taylor AE, Granger HJ: *Dynamics and Control of the Body Fluids*, Philadelphia, 1975, WB Saunders.

Herman MA, Kahn BB: Glucose transport and sensing in the maintenance of glucose homeostasis and metabolic harmony, *J Clin Invest* 116:1767, 2006.

Krahe R, Gabbiani F: Burst firing in sensory systems, *Nat Rev Neurosci* 5:13, 2004.

Orgel LE: The origin of life on the earth, *Sci Am* 271:76, 1994.

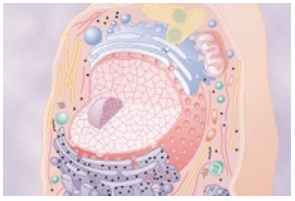
Quarles LD: Endocrine functions of bone in mineral metabolism regulation, *J Clin Invest* 118:3820, 2008.

Smith HW: *From Fish to Philosopher*, New York, 1961, Doubleday.

Tjian R: Molecular machines that control genes, *Sci Am* 272:54, 1995.

CHAPTER 2

The Cell and Its Functions



Each of the 100 trillion cells in a human being is a living structure that can survive for months or many years, provided its surrounding fluids contain appropriate nutrients. To understand the function of organs and other structures of the body, it is essential that we first understand the basic organization of the cell and the functions of its component parts.

Organization of the Cell

A typical cell, as seen by the light microscope, is shown in Figure 2-1. Its two major parts are the *nucleus* and the *cytoplasm.* The nucleus is separated from the cytoplasm by a *nuclear membrane,* and the cytoplasm is separated from the surrounding fluids by a *cell membrane,* also called the *plasma membrane.*

The different substances that make up the cell are collectively called *protoplasm.* Protoplasm is composed mainly of five basic substances: water, electrolytes, proteins, lipids, and carbohydrates.

Water. The principal fluid medium of the cell is water, which is present in most cells, except for fat cells, in a concentration of 70 to 85 percent. Many cellular chemicals are dissolved in the water. Others are suspended in the water as solid particulates. Chemical reactions take place among the dissolved chemicals or at the surfaces of the suspended particles or membranes.

Ions. Important ions in the cell include *potassium, magnesium, phosphate, sulfate, bicarbonate,* and smaller quantities of *sodium, chloride,* and *calcium.* These are all discussed in more detail in Chapter 4, which considers the interrelations between the intracellular and extracellular fluids.

The ions provide inorganic chemicals for cellular reactions. Also, they are necessary for operation of some of the cellular control mechanisms. For instance, ions acting at the cell membrane are required for transmission of electrochemical impulses in nerve and muscle fibers.

Proteins. After water, the most abundant substances in most cells are proteins, which normally constitute 10 to 20 percent of the cell mass. These can be divided into two types: *structural proteins* and *functional proteins.*

Structural proteins are present in the cell mainly in the form of long filaments that are polymers of many individual protein molecules. A prominent use of such intracellular filaments is to form *microtubules* that provide the "cytoskeletons" of such cellular organelles as cilia, nerve axons, the mitotic spindles of mitosing cells, and a tangled mass of thin filamentous tubules that hold the parts of the cytoplasm and nucleoplasm together in their respective compartments. Extracellularly, fibrillar proteins are found especially in the collagen and elastin fibers of connective tissue and in blood vessel walls, tendons, ligaments, and so forth.

The *functional proteins* are an entirely different type of protein, usually composed of combinations of a few molecules in tubular-globular form. These proteins are mainly the *enzymes* of the cell and, in contrast to the fibrillar proteins, are often mobile in the cell fluid. Also, many of them are adherent to membranous structures inside the cell. The enzymes come into direct contact with other substances in the cell fluid and thereby catalyze specific intracellular chemical reactions. For instance, the chemical reactions that split glucose into its component parts and then combine these with oxygen to form carbon dioxide and water while simultaneously providing energy for cellular function are all catalyzed by a series of protein enzymes.

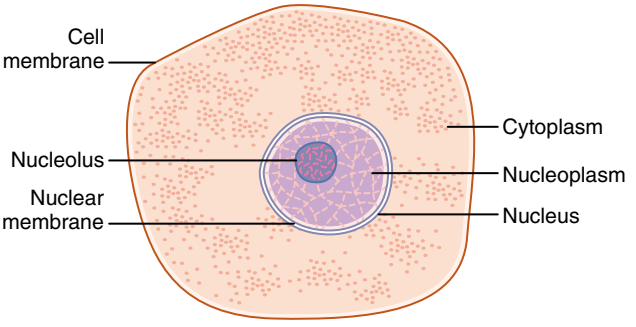


Figure 2-1 Structure of the cell as seen with the light microscope.

Lipids. Lipids are several types of substances that are grouped together because of their common property of being soluble in fat solvents. Especially important lipids are *phospholipids* and *cholesterol,* which together constitute only about 2 percent of the total cell mass. The significance of phospholipids and cholesterol is that they are mainly insoluble in water and, therefore, are used to form the cell membrane and intracellular membrane barriers that separate the different cell compartments.

In addition to phospholipids and cholesterol, some cells contain large quantities of *triglycerides,* also called *neutral fat.* In the *fat cells,* triglycerides often account for as much as 95 percent of the cell mass. The fat stored in these cells represents the body's main storehouse of energy-giving nutrients that can later be dissoluted and used to provide energy wherever in the body it is needed.

Carbohydrates. Carbohydrates have little structural function in the cell except as parts of glycoprotein molecules, but they play a major role in nutrition of the cell. Most human cells do not maintain large stores of carbohydrates; the amount usually averages about 1 percent of their total mass but increases to as much as 3 percent in muscle cells and, occasionally, 6 percent in liver cells. However, carbohydrate in the form of dissolved glucose is always present in the surrounding extracellular fluid so that it is readily available to the cell. Also, a small amount of carbohydrate is stored in the cells in the form of *glycogen,* which is an insoluble polymer of glucose that can be depolymerized and used rapidly to supply the cells' energy needs.

Physical Structure of the Cell

The cell is not merely a bag of fluid, enzymes, and chemicals; it also contains highly organized physical structures, called *intracellular organelles.* The physical nature of each organelle is as important as the cell's chemical constituents for cell function. For instance, without one of the organelles, the *mitochondria,* more than 95 percent of the cell's energy release from nutrients would cease immediately. The most important organelles and other structures of the cell are shown in Figure 2-2.

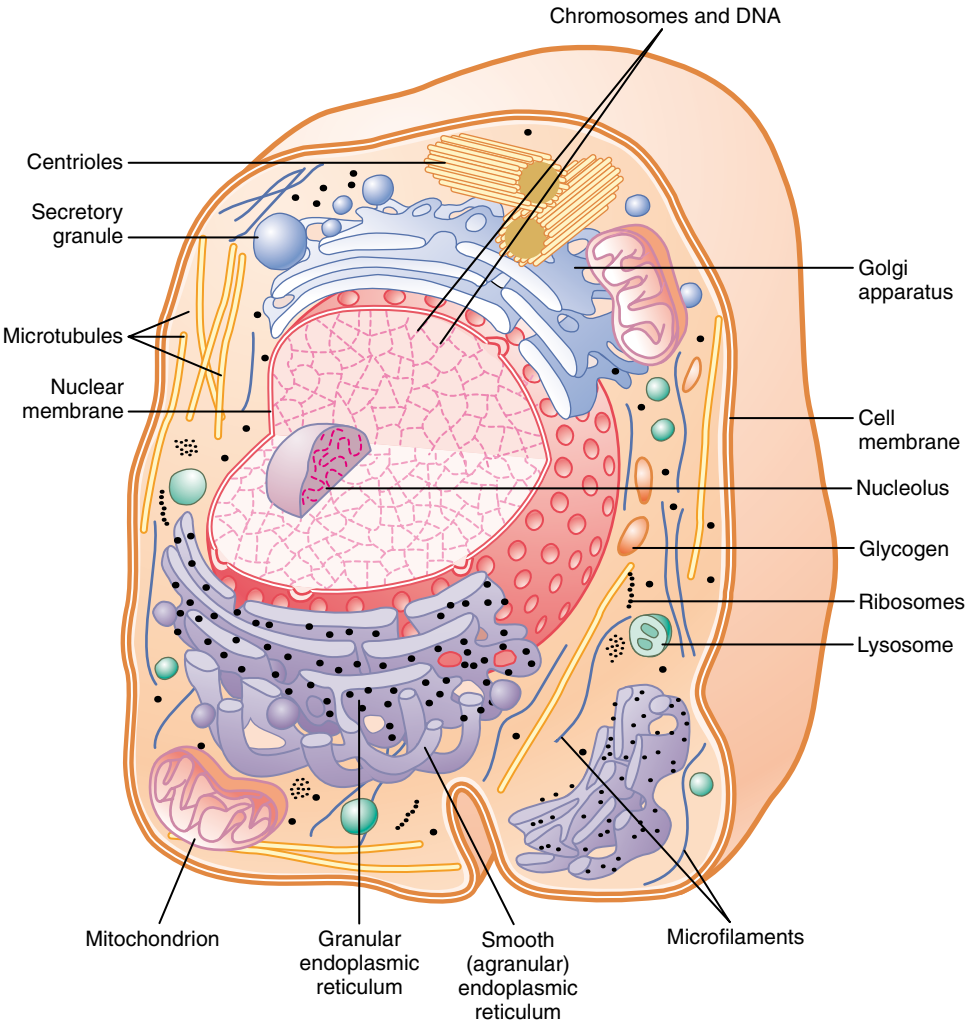


Figure 2-2 Reconstruction of a typical cell, showing the internal organelles in the cytoplasm and in the nucleus.

Membranous Structures of the Cell

Most organelles of the cell are covered by membranes composed primarily of lipids and proteins. These membranes include the *cell membrane, nuclear membrane, membrane of the endoplasmic reticulum,* and *membranes of the mitochondria, lysosomes,* and *Golgi apparatus.*

The lipids of the membranes provide a barrier that impedes the movement of water and water-soluble substances from one cell compartment to another because water is not soluble in lipids. However, protein molecules in the membrane often do penetrate all the way through the membrane, thus providing specialized pathways, often organized into actual *pores,* for passage of specific substances through the membrane. Also, many other membrane proteins are *enzymes* that catalyze a multitude of different chemical reactions, discussed here and in subsequent chapters.

Cell Membrane

The cell membrane (also called the *plasma membrane*), which envelops the cell, is a thin, pliable, elastic structure only 7.5 to 10 nanometers thick. It is composed almost entirely of proteins and lipids. The approximate composition is proteins, 55 percent; phospholipids, 25 percent; cholesterol, 13 percent; other lipids, 4 percent; and carbohydrates, 3 percent.

Lipid Barrier of the Cell Membrane Impedes Water Penetration. Figure 2-3 shows the structure of the cell membrane. Its basic structure is a *lipid bilayer,* which is a thin, double-layered film of lipids—each layer only one molecule thick—that is continuous over the entire cell surface. Interspersed in this lipid film are large globular protein molecules.

The basic lipid bilayer is composed of phospholipid molecules. One end of each phospholipid molecule is soluble in water; that is, it is *hydrophilic.* The other end is soluble only in fats; that is, it is *hydrophobic.* The phosphate end of the phospholipid is hydrophilic, and the fatty acid portion is hydrophobic.

Because the hydrophobic portions of the phospholipid molecules are repelled by water but are mutually attracted to one another, they have a natural tendency to attach to one another in the middle of the membrane, as shown in Figure 2-3. The hydrophilic phosphate portions then constitute the two surfaces of the complete cell membrane, in contact with *intracellular* water on the inside of the membrane and *extracellular* water on the outside surface.

The lipid layer in the middle of the membrane is impermeable to the usual water-soluble substances, such as ions, glucose, and urea. Conversely, fat-soluble substances, such as oxygen, carbon dioxide, and alcohol, can penetrate this portion of the membrane with ease.

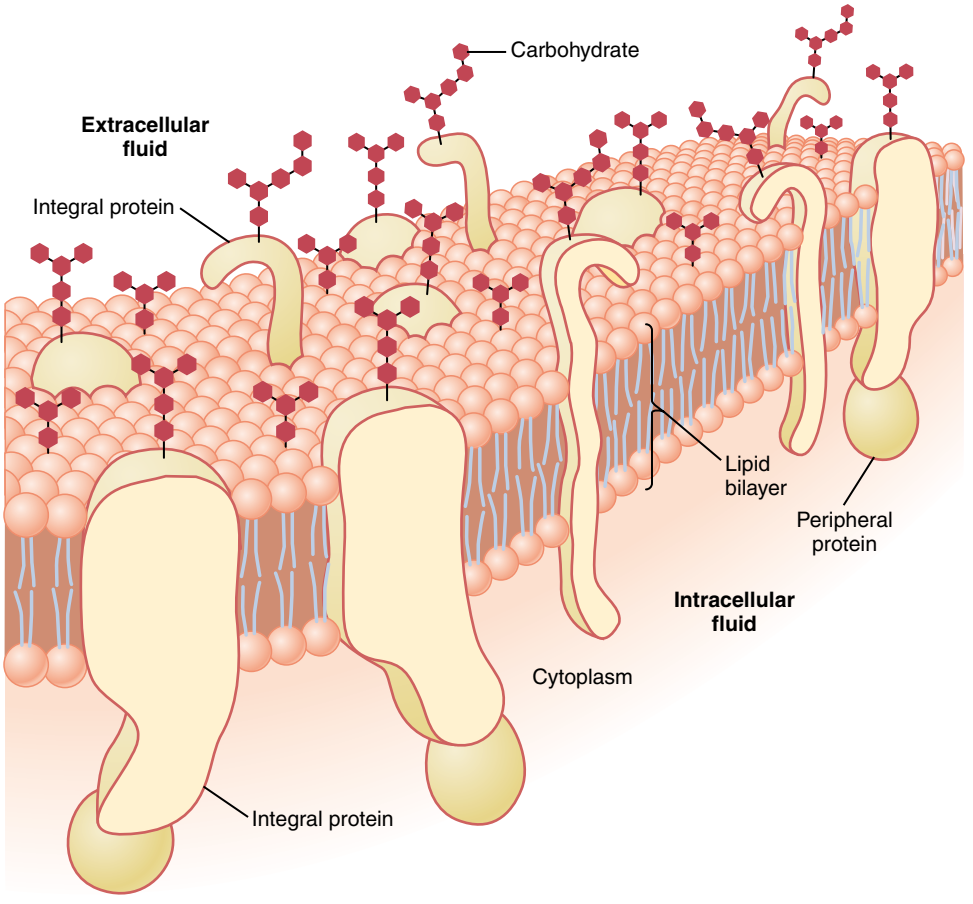


Figure 2-3 Structure of the cell membrane, showing that it is composed mainly of a lipid bilayer of phospholipid molecules, but with large numbers of protein molecules protruding through the layer. Also, carbohydrate moieties are attached to the protein molecules on the outside of the membrane and to additional protein molecules on the inside. (Redrawn from Lodish HF, Rothman JE: The assembly of cell membranes. Sci Am 240:48, 1979. Copyright George V. Kevin.)

The cholesterol molecules in the membrane are also lipid in nature because their steroid nucleus is highly fat soluble. These molecules, in a sense, are dissolved in the bilayer of the membrane. They mainly help determine the degree of permeability (or impermeability) of the bilayer to water-soluble constituents of body fluids. Cholesterol controls much of the fluidity of the membrane as well.

Integral and Peripheral Cell Membrane Proteins. Figure 2-3 also shows globular masses floating in the lipid bilayer. These are membrane proteins, most of which are *glycoproteins.* There are two types of cell membrane proteins: *integral proteins* that protrude all the way through the membrane and *peripheral proteins* that are attached only to one surface of the membrane and do not penetrate all the way through.

Many of the integral proteins provide structural *channels* (or *pores*) through which water molecules and water-soluble substances, especially ions, can diffuse between the extracellular and intracellular fluids. These protein channels also have selective properties that allow preferential diffusion of some substances over others.

Other integral proteins act as *carrier proteins* for transporting substances that otherwise could not penetrate the lipid bilayer. Sometimes these even transport substances in the direction opposite to their electrochemical gradients for diffusion, which is called "active transport." Still others act as *enzymes.*

Integral membrane proteins can also serve as *receptors* for water-soluble chemicals, such as peptide hormones, that do not easily penetrate the cell membrane. Interaction of cell membrane receptors with specific *ligands* that bind to the receptor causes conformational changes in the receptor protein. This, in turn, enzymatically activates the intracellular part of the protein or induces interactions between the receptor and proteins in the cytoplasm that act as *second messengers,* thereby relaying the signal from the extracellular part of the receptor to the interior of the cell. In this way, integral proteins spanning the cell membrane provide a means of conveying information about the environment to the cell interior.

Peripheral protein molecules are often attached to the integral proteins. These peripheral proteins function almost entirely as enzymes or as controllers of transport of substances through the cell membrane "pores."

Membrane Carbohydrates—The Cell "Glycocalyx." Membrane carbohydrates occur almost invariably in combination with proteins or lipids in the form of *glycoproteins* or *glycolipids.* In fact, most of the integral proteins are glycoproteins, and about one tenth of the membrane lipid molecules are glycolipids. The "glyco" portions of these molecules almost invariably protrude to the outside of the cell, dangling outward from the cell surface. Many other carbohydrate compounds, called *proteoglycans*—which are mainly carbohydrate substances bound to small protein cores—are loosely attached to the outer surface of the cell as well. Thus, the entire outside surface of the cell often has a loose carbohydrate coat called the *glycocalyx.*

The carbohydrate moieties attached to the outer surface of the cell have several important functions: (1) Many of them have a negative electrical charge, which gives most cells an overall negative surface charge that repels other negative objects. (2) The glycocalyx of some cells attaches to the glycocalyx of other cells, thus attaching cells to one another. (3) Many of the carbohydrates act as *receptor substances* for binding hormones, such as insulin; when bound, this combination activates attached internal proteins that, in turn, activate a cascade of intracellular enzymes. (4) Some carbohydrate moieties enter into immune reactions, as discussed in Chapter 34.

Cytoplasm and Its Organelles

The cytoplasm is filled with both minute and large dispersed particles and organelles. The clear fluid portion of the cytoplasm in which the particles are dispersed is called *cytosol;* this contains mainly dissolved proteins, electrolytes, and glucose.

Dispersed in the cytoplasm are neutral fat globules, glycogen granules, ribosomes, secretory vesicles, and five especially important organelles: the *endoplasmic reticulum,* the *Golgi apparatus, mitochondria, lysosomes,* and *peroxisomes.*

Endoplasmic Reticulum

Figure 2-2 shows a network of tubular and flat vesicular structures in the cytoplasm; this is the *endoplasmic reticulum.* The tubules and vesicles interconnect with one another. Also, their walls are constructed of lipid bilayer membranes that contain large amounts of proteins, similar to the cell membrane. The total surface area of this structure in some cells—the liver cells, for instance—can be as much as 30 to 40 times the cell membrane area.

The detailed structure of a small portion of endoplasmic reticulum is shown in Figure 2-4. The space inside the tubules and vesicles is filled with *endoplasmic matrix,* a watery medium that is different from the fluid in the cytosol outside the endoplasmic reticulum. Electron micrographs show that the space inside the endoplasmic reticulum is connected with the space between the two membrane surfaces of the nuclear membrane.

Substances formed in some parts of the cell enter the space of the endoplasmic reticulum and are then conducted to other parts of the cell. Also, the vast surface area of this reticulum and the multiple enzyme systems attached to its membranes provide machinery for a major share of the metabolic functions of the cell.

Ribosomes and the Granular Endoplasmic Reticulum. Attached to the outer surfaces of many parts of the endoplasmic reticulum are large numbers of minute granular particles called *ribosomes.* Where these are present, the reticulum is called the *granular endoplasmic reticulum.* The ribosomes are composed of a mixture of RNA and proteins, and they function to synthesize new protein molecules in the cell, as discussed later in this chapter and in Chapter 3.

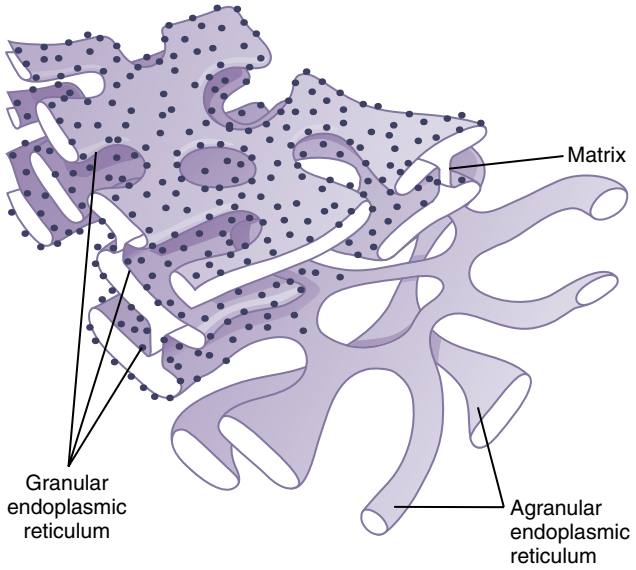


Figure 2-4 Structure of the endoplasmic reticulum. (Modified from DeRobertis EDP, Saez FA, DeRobertis EMF: Cell Biology, 6th ed. Philadelphia: WB Saunders, 1975.)

Agranular Endoplasmic Reticulum. Part of the endoplasmic reticulum has no attached ribosomes. This part is called the *agranular,* or *smooth, endoplasmic reticulum.* The agranular reticulum functions for the synthesis of lipid substances and for other processes of the cells promoted by intrareticular enzymes.

Golgi Apparatus

The Golgi apparatus, shown in Figure 2-5, is closely related to the endoplasmic reticulum. It has membranes similar to those of the agranular endoplasmic reticulum. It is usually composed of four or more stacked layers of thin, flat, enclosed vesicles lying near one side of the nucleus. This apparatus is prominent in secretory cells, where it is located on the side of the cell from which the secretory substances are extruded.

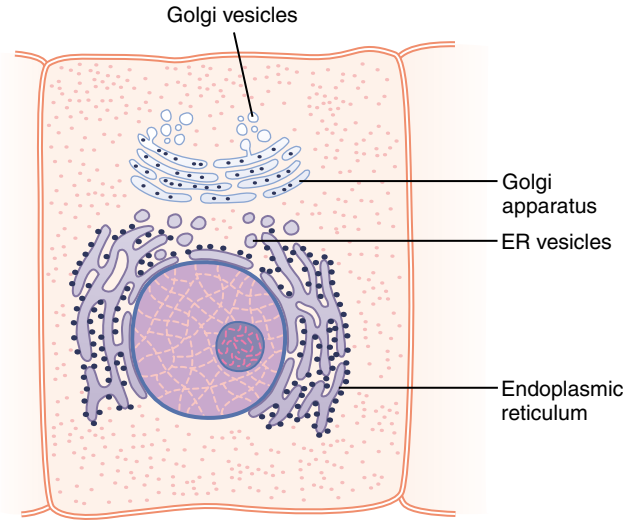


Figure 2-5 A typical Golgi apparatus and its relationship to the endoplasmic reticulum (ER) and the nucleus.

The Golgi apparatus functions in association with the endoplasmic reticulum. As shown in Figure 2-5, small "transport vesicles" (also called endoplasmic reticulum vesicles, or *ER vesicles*) continually pinch off from the endoplasmic reticulum and shortly thereafter fuse with the Golgi apparatus. In this way, substances entrapped in the ER vesicles are transported from the endoplasmic reticulum to the Golgi apparatus. The transported substances are then processed in the Golgi apparatus to form lysosomes, secretory vesicles, and other cytoplasmic components that are discussed later in the chapter.

Lysosomes

Lysosomes, shown in Figure 2-2, are vesicular organelles that form by breaking off from the Golgi apparatus and then dispersing throughout the cytoplasm. The lysosomes provide an *intracellular digestive system* that allows the cell to digest (1) damaged cellular structures, (2) food particles that have been ingested by the cell, and (3) unwanted matter such as bacteria. The lysosome is quite different in different cell types, but it is usually 250 to 750 nanometers in diameter. It is surrounded by a typical lipid bilayer membrane and is filled with large numbers of small granules 5 to 8 nanometers in diameter, which are protein aggregates of as many as 40 different *hydrolase (digestive) enzymes.* A hydrolytic enzyme is capable of splitting an organic compound into two or more parts by combining hydrogen from a water molecule with one part of the compound and combining the hydroxyl portion of the water molecule with the other part of the compound. For instance, protein is hydrolyzed to form amino acids, glycogen is hydrolyzed to form glucose, and lipids are hydrolyzed to form fatty acids and glycerol.

Ordinarily, the membrane surrounding the lysosome prevents the enclosed hydrolytic enzymes from coming in contact with other substances in the cell and, therefore, prevents their digestive actions. However, some conditions of the cell break the membranes of some of the lysosomes, allowing release of the digestive enzymes. These enzymes then split the organic substances with which they come in contact into small, highly diffusible substances such as amino acids and glucose. Some of the specific functions of lysosomes are discussed later in the chapter.

Peroxisomes

Peroxisomes are similar physically to lysosomes, but they are different in two important ways. First, they are believed to be formed by self-replication (or perhaps by budding off from the smooth endoplasmic reticulum) rather than from the Golgi apparatus. Second, they contain oxidases rather than hydrolases. Several of the oxidases are capable of combining oxygen with hydrogen ions derived from different intracellular chemicals to form hydrogen peroxide (H_2O_2). Hydrogen peroxide is a highly oxidizing substance and is used in association with *catalase,* another oxidase enzyme present in large quantities in peroxisomes, to oxidize many substances that might otherwise be poisonous

to the cell. For instance, about half the alcohol a person drinks is detoxified by the peroxisomes of the liver cells in this manner.

Secretory Vesicles

One of the important functions of many cells is secretion of special chemical substances. Almost all such secretory substances are formed by the endoplasmic reticulum–Golgi apparatus system and are then released from the Golgi apparatus into the cytoplasm in the form of storage vesicles called *secretory vesicles* or *secretory granules.* Figure 2-6 shows typical secretory vesicles inside pancreatic acinar cells; these vesicles store protein proenzymes (enzymes that are not yet activated). The proenzymes are secreted later through the outer cell membrane into the pancreatic duct and thence into the duodenum, where they become activated and perform digestive functions on the food in the intestinal tract.

Mitochondria

The mitochondria, shown in Figures 2-2 and 2-7, are called the "powerhouses" of the cell. Without them, cells would be unable to extract enough energy from the nutrients, and essentially all cellular functions would cease.

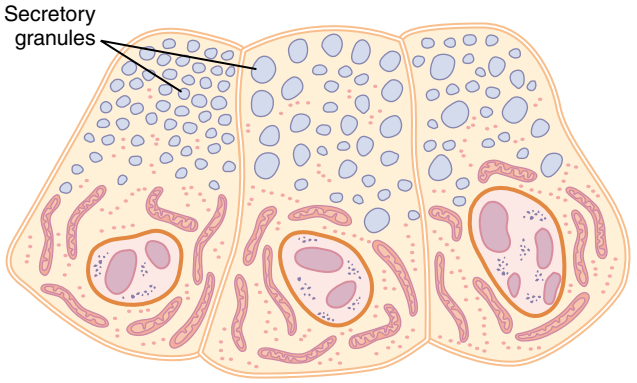


Figure 2-6 Secretory granules (secretory vesicles) in acinar cells of the pancreas.

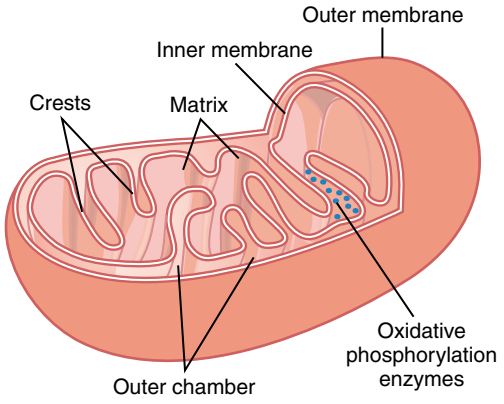


Figure 2-7 Structure of a mitochondrion. (Modified from DeRobertis EDP, Saez FA, DeRobertis EMF: Cell Biology, 6th ed. Philadelphia: WB Saunders, 1975.)

Mitochondria are present in all areas of each cell's cytoplasm, but the total number per cell varies from less than a hundred up to several thousand, depending on the amount of energy required by the cell. Further, the mitochondria are concentrated in those portions of the cell that are responsible for the major share of its energy metabolism. They are also variable in size and shape. Some are only a few hundred nanometers in diameter and globular in shape, whereas others are elongated—as large as 1 micrometer in diameter and 7 micrometers long; still others are branching and filamentous.

The basic structure of the mitochondrion, shown in Figure 2-7, is composed mainly of two lipid bilayer–protein membranes: an *outer membrane* and an *inner membrane.* Many infoldings of the inner membrane form *shelves* onto which oxidative enzymes are attached. In addition, the inner cavity of the mitochondrion is filled with a *matrix* that contains large quantities of dissolved enzymes that are necessary for extracting energy from nutrients. These enzymes operate in association with the oxidative enzymes on the shelves to cause oxidation of the nutrients, thereby forming carbon dioxide and water and at the same time releasing energy. The liberated energy is used to synthesize a "high-energy" substance called *adenosine triphosphate* (ATP). ATP is then transported out of the mitochondrion, and it diffuses throughout the cell to release its own energy wherever it is needed for performing cellular functions. The chemical details of ATP formation by the mitochondrion are given in Chapter 67, but some of the basic functions of ATP in the cell are introduced later in this chapter.

Mitochondria are self-replicative, which means that one mitochondrion can form a second one, a third one, and so on, whenever there is a need in the cell for increased amounts of ATP. Indeed, the mitochondria contain *DNA* similar to that found in the cell nucleus. In Chapter 3 we will see that DNA is the basic chemical of the nucleus that controls replication of the cell. The DNA of the mitochondrion plays a similar role, controlling replication of the mitochondrion.

Cell Cytoskeleton—Filament and Tubular Structures

The fibrillar proteins of the cell are usually organized into filaments or tubules. These originate as precursor protein molecules synthesized by ribosomes in the cytoplasm. The precursor molecules then polymerize to form *filaments.* As an example, large numbers of actin filaments frequently occur in the outer zone of the cytoplasm, called the *ectoplasm,* to form an elastic support for the cell membrane. Also, in muscle cells, actin and myosin filaments are organized into a special contractile machine that is the basis for muscle contraction, as discussed in detail in Chapter 6.

A special type of stiff filament composed of polymerized *tubulin* molecules is used in all cells to construct strong tubular structures, the *microtubules.* Figure 2-8 shows typical microtubules that were teased from the flagellum of a sperm.

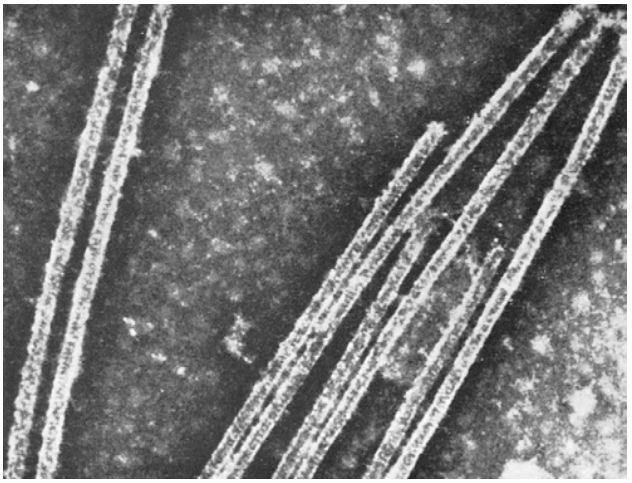


Figure 2-8 Microtubules teased from the flagellum of a sperm. (From Wolstenholme GEW, O'Connor M, and the publisher, JA Churchill, 1967. Figure 4, page 314. Copyright the Novartis Foundation, formerly the Ciba Foundation.)

Another example of microtubules is the tubular skeletal structure in the center of each cilium that radiates upward from the cell cytoplasm to the tip of the cilium. This structure is discussed later in the chapter and is illustrated in Figure 2-17. Also, both the *centrioles* and the *mitotic spindle* of the mitosing cell are composed of stiff microtubules.

Thus, a primary function of microtubules is to act as a *cytoskeleton,* providing rigid physical structures for certain parts of cells.

Nucleus

The nucleus is the control center of the cell. Briefly, the nucleus contains large quantities of DNA, which are the *genes.* The genes determine the characteristics of the cell's proteins, including the structural proteins, as well as the intracellular enzymes that control cytoplasmic and nuclear activities.

The genes also control and promote reproduction of the cell itself. The genes first reproduce to give two identical sets of genes; then the cell splits by a special process called *mitosis* to form two daughter cells, each of which receives one of the two sets of DNA genes. All these activities of the nucleus are considered in detail in the next chapter.

Unfortunately, the appearance of the nucleus under the microscope does not provide many clues to the mechanisms by which the nucleus performs its control activities. Figure 2-9 shows the light microscopic appearance of the *interphase* nucleus (during the period between mitoses), revealing darkly staining *chromatin material* throughout the nucleoplasm. During mitosis, the chromatin material organizes in the form of highly structured *chromosomes,* which can then be easily identified using the light microscope, as illustrated in the next chapter.

Nuclear Membrane

The *nuclear membrane,* also called the *nuclear envelope,* is actually two separate bilayer membranes, one inside the other. The outer membrane is continuous with the endoplasmic reticulum of the cell cytoplasm, and the space between the two nuclear membranes is also continuous with the space inside the endoplasmic reticulum, as shown in Figure 2-9.

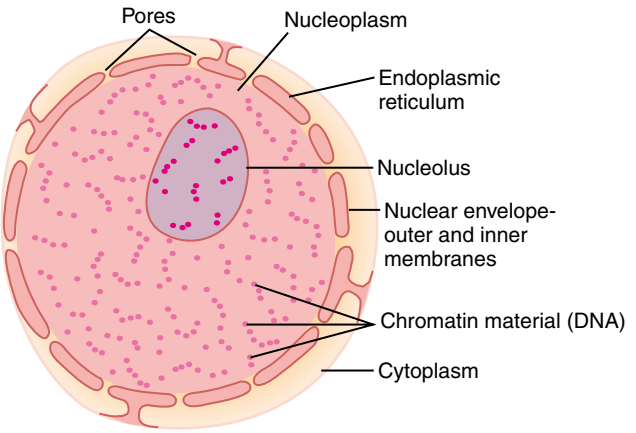


Figure 2-9 Structure of the nucleus.

The nuclear membrane is penetrated by several thousand *nuclear pores.* Large complexes of protein molecules are attached at the edges of the pores so that the central area of each pore is only about 9 nanometers in diameter. Even this size is large enough to allow molecules up to 44,000 molecular weight to pass through with reasonable ease.

Nucleoli and Formation of Ribosomes

The nuclei of most cells contain one or more highly staining structures called *nucleoli.* The nucleolus, unlike most other organelles discussed here, does not have a limiting membrane. Instead, it is simply an accumulation of large amounts of RNA and proteins of the types found in ribosomes. The nucleolus becomes considerably enlarged when the cell is actively synthesizing proteins.

Formation of the nucleoli (and of the ribosomes in the cytoplasm outside the nucleus) begins in the nucleus. First, specific DNA genes in the chromosomes cause RNA to be synthesized. Some of this is stored in the nucleoli, but most of it is transported outward through the nuclear pores into cytoplasm. Here, it is used in conjunction with specific proteins to assemble "mature" ribosomes that play an essential role in forming cytoplasmic proteins, as discussed more fully in Chapter 3.

Comparison of the Animal Cell with Precellular Forms of Life

The cell is a complicated organism that required many hundreds of millions of years to develop after the earliest form of life, an organism similar to the present-day *virus,* first appeared on earth. Figure 2-10 shows the relative sizes of (1) the smallest known virus, (2) a large virus, (3) a *rickettsia,* (4) a *bacterium,* and (5) a *nucleated cell,* demonstrating that the cell has a diameter about 1000 times that of the smallest virus and, therefore, a volume about

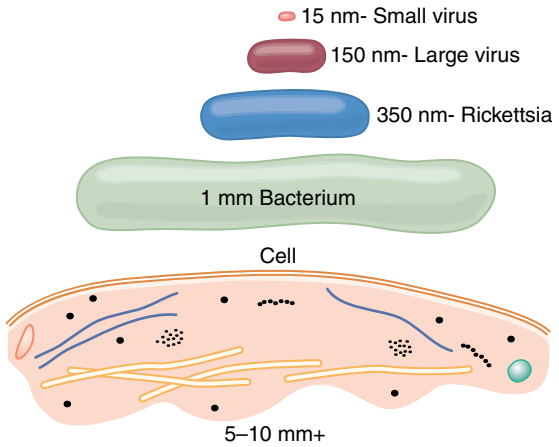


Figure 2-10 Comparison of sizes of precellular organisms with that of the average cell in the human body.

1 billion times that of the smallest virus. Correspondingly, the functions and anatomical organization of the cell are also far more complex than those of the virus.

The essential life-giving constituent of the small virus is a *nucleic acid* embedded in a coat of protein. This nucleic acid is composed of the same basic nucleic acid constituents (DNA or RNA) found in mammalian cells, and it is capable of reproducing itself under appropriate conditions. Thus, the virus propagates its lineage from generation to generation and is therefore a living structure in the same way that the cell and the human being are living structures.

As life evolved, other chemicals besides nucleic acid and simple proteins became integral parts of the organism, and specialized functions began to develop in different parts of the virus. A membrane formed around the virus, and inside the membrane, a fluid matrix appeared. Specialized chemicals then developed inside the fluid to perform special functions; many protein enzymes appeared that were capable of catalyzing chemical reactions and, therefore, determining the organism's activities.

In still later stages of life, particularly in the rickettsial and bacterial stages, *organelles* developed inside the organism, representing physical structures of chemical aggregates that perform functions in a more efficient manner than can be achieved by dispersed chemicals throughout the fluid matrix.

Finally, in the nucleated cell, still more complex organelles developed, the most important of which is the *nucleus* itself. The nucleus distinguishes this type of cell from all lower forms of life; the nucleus provides a control center for all cellular activities, and it provides for exact reproduction of new cells generation after generation, each new cell having almost exactly the same structure as its progenitor.

Functional Systems of the Cell

In the remainder of this chapter, we discuss several representative functional systems of the cell that make it a living organism.

Ingestion by the Cell—Endocytosis

If a cell is to live and grow and reproduce, it must obtain nutrients and other substances from the surrounding fluids. Most substances pass through the cell membrane by *diffusion* and *active transport.*

Diffusion involves simple movement through the membrane caused by the random motion of the molecules of the substance; substances move either through cell membrane pores or, in the case of lipid-soluble substances, through the lipid matrix of the membrane.

Active transport involves the actual carrying of a substance through the membrane by a physical protein structure that penetrates all the way through the membrane. These active transport mechanisms are so important to cell function that they are presented in detail in Chapter 4.

Very large particles enter the cell by a specialized function of the cell membrane called *endocytosis.* The principal forms of endocytosis are *pinocytosis* and *phagocytosis.* Pinocytosis means ingestion of minute particles that form vesicles of extracellular fluid and particulate constituents inside the cell cytoplasm. Phagocytosis means ingestion of large particles, such as bacteria, whole cells, or portions of degenerating tissue.

Pinocytosis. Pinocytosis occurs continually in the cell membranes of most cells, but it is especially rapid in some cells. For instance, it occurs so rapidly in macrophages that about 3 percent of the total macrophage membrane is engulfed in the form of vesicles each minute. Even so, the pinocytotic vesicles are so small—usually only 100 to 200 nanometers in diameter—that most of them can be seen only with the electron microscope.

Pinocytosis is the only means by which most large macromolecules, such as most protein molecules, can enter cells. In fact, the rate at which pinocytotic vesicles form is usually enhanced when such macromolecules attach to the cell membrane.

Figure 2-11 demonstrates the successive steps of pinocytosis, showing three molecules of protein attaching to the membrane. These molecules usually attach to

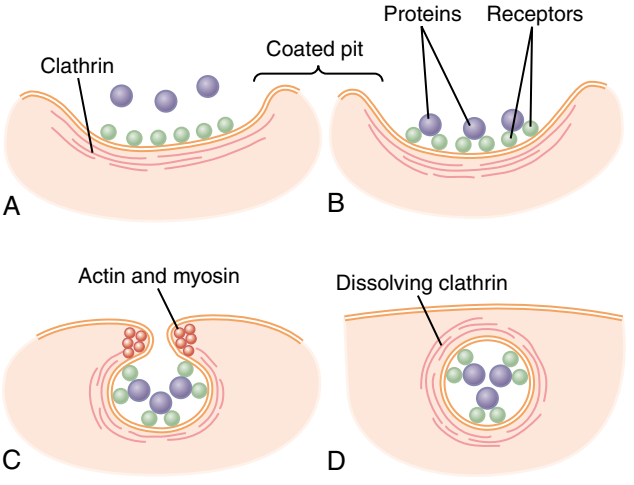


Figure 2-11 Mechanism of pinocytosis.

specialized protein *receptors* on the surface of the membrane that are specific for the type of protein that is to be absorbed. The receptors generally are concentrated in small pits on the outer surface of the cell membrane, called *coated pits.* On the inside of the cell membrane beneath these pits is a latticework of fibrillar protein called *clathrin,* as well as other proteins, perhaps including contractile filaments of *actin* and *myosin.* Once the protein molecules have bound with the receptors, the surface properties of the local membrane change in such a way that the entire pit invaginates inward and the fibrillar proteins surrounding the invaginating pit cause its borders to close over the attached proteins, as well as over a small amount of extracellular fluid. Immediately thereafter, the invaginated portion of the membrane breaks away from the surface of the cell, forming a *pinocytotic vesicle* inside the cytoplasm of the cell.

What causes the cell membrane to go through the necessary contortions to form pinocytotic vesicles is still unclear. This process requires energy from within the cell; this is supplied by ATP, a high-energy substance discussed later in the chapter. Also, it requires the presence of calcium ions in the extracellular fluid, which probably react with contractile protein filaments beneath the coated pits to provide the force for pinching the vesicles away from the cell membrane.

Phagocytosis. Phagocytosis occurs in much the same way as pinocytosis, except that it involves large particles rather than molecules. Only certain cells have the capability of phagocytosis, most notably the tissue macrophages and some of the white blood cells.

Phagocytosis is initiated when a particle such as a bacterium, a dead cell, or tissue debris binds with receptors on the surface of the phagocyte. In the case of bacteria, each bacterium is usually already attached to a specific antibody, and it is the antibody that attaches to the phagocyte receptors, dragging the bacterium along with it. This intermediation of antibodies is called *opsonization,* which is discussed in Chapters 33 and 34.

Phagocytosis occurs in the following steps:

1. The cell membrane receptors attach to the surface ligands of the particle.
2. The edges of the membrane around the points of attachment evaginate outward within a fraction of a second to surround the entire particle; then, progressively more and more membrane receptors attach to the particle ligands. All this occurs suddenly in a zipper-like manner to form a closed *phagocytic vesicle.*
3. Actin and other contractile fibrils in the cytoplasm surround the phagocytic vesicle and contract around its outer edge, pushing the vesicle to the interior.
4. The contractile proteins then pinch the stem of the vesicle so completely that the vesicle separates from the cell membrane, leaving the vesicle in the cell interior in the same way that pinocytotic vesicles are formed.

Digestion of Pinocytotic and Phagocytic Foreign Substances Inside the Cell—Function of the Lysosomes

Almost immediately after a pinocytotic or phagocytic vesicle appears inside a cell, one or more *lysosomes* become attached to the vesicle and empty their *acid hydrolases* to the inside of the vesicle, as shown in Figure 2-12. Thus, a *digestive vesicle* is formed inside the cell cytoplasm in which the vesicular hydrolases begin hydrolyzing the proteins, carbohydrates, lipids, and other substances in the vesicle. The products of digestion are small molecules of amino acids, glucose, phosphates, and so forth that can diffuse through the membrane of the vesicle into the cytoplasm. What is left of the digestive vesicle, called the *residual body,* represents indigestible substances. In most instances, this is finally excreted through the cell membrane by a process called *exocytosis,* which is essentially the opposite of endocytosis.

Thus, the pinocytotic and phagocytic vesicles containing lysosomes can be called the *digestive organs* of the cells.

Regression of Tissues and Autolysis of Cells. Tissues of the body often regress to a smaller size. For instance, this occurs in the uterus after pregnancy, in muscles during long periods of inactivity, and in mammary glands at the end of lactation. Lysosomes are responsible for much of this regression. The mechanism by which lack of activity in a tissue causes the lysosomes to increase their activity is unknown.

Another special role of the lysosomes is removal of damaged cells or damaged portions of cells from tissues. Damage to the cell—caused by heat, cold, trauma, chemicals, or any other factor—induces lysosomes to rupture. The released hydrolases immediately begin to digest the surrounding organic substances. If the damage is slight, only a portion of the cell is removed and the cell is then repaired. If the damage is severe, the entire cell is

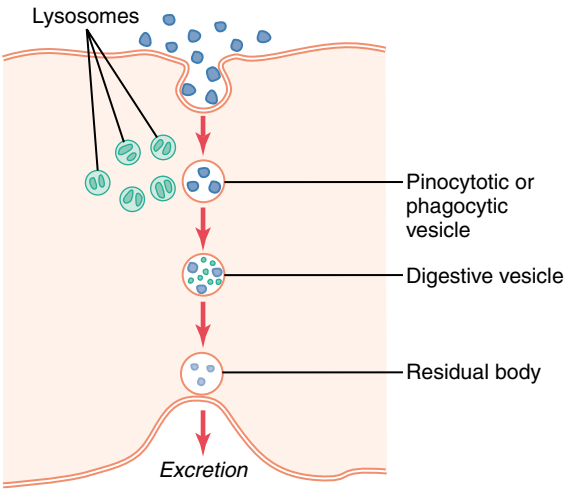


Figure 2-12 Digestion of substances in pinocytotic or phagocytic vesicles by enzymes derived from lysosomes.

digested, a process called *autolysis.* In this way, the cell is completely removed and a new cell of the same type ordinarily is formed by mitotic reproduction of an adjacent cell to take the place of the old one.

The lysosomes also contain bactericidal agents that can kill phagocytized bacteria before they can cause cellular damage. These agents include (1) *lysozyme,* which dissolves the bacterial cell membrane; (2) *lysoferrin,* which binds iron and other substances before they can promote bacterial growth; and (3) acid at a pH of about 5.0, which activates the hydrolases and inactivates bacterial metabolic systems.

Synthesis and Formation of Cellular Structures by Endoplasmic Reticulum and Golgi Apparatus

Specific Functions of the Endoplasmic Reticulum

The extensiveness of the endoplasmic reticulum and the Golgi apparatus in secretory cells has already been emphasized. These structures are formed primarily of lipid bilayer membranes similar to the cell membrane, and their walls are loaded with protein enzymes that catalyze the synthesis of many substances required by the cell.

Most synthesis begins in the endoplasmic reticulum. The products formed there are then passed on to the Golgi apparatus, where they are further processed before being released into the cytoplasm. But first, let us note the specific products that are synthesized in specific portions of the endoplasmic reticulum and the Golgi apparatus.

Proteins Are Formed by the Granular Endoplasmic Reticulum. The granular portion of the endoplasmic reticulum is characterized by large numbers of ribosomes attached to the outer surfaces of the endoplasmic reticulum membrane. As discussed in Chapter 3, protein molecules are synthesized within the structures of the ribosomes. The ribosomes extrude some of the synthesized protein molecules directly into the cytosol, but they also extrude many more through the wall of the endoplasmic reticulum to the interior of the endoplasmic vesicles and tubules, into the *endoplasmic matrix.*

Synthesis of Lipids by the Smooth Endoplasmic Reticulum. The endoplasmic reticulum also synthesizes lipids, especially phospholipids and cholesterol. These are rapidly incorporated into the lipid bilayer of the endoplasmic reticulum itself, thus causing the endoplasmic reticulum to grow more extensive. This occurs mainly in the smooth portion of the endoplasmic reticulum.

To keep the endoplasmic reticulum from growing beyond the needs of the cell, small vesicles called *ER vesicles* or *transport vesicles* continually break away from the smooth reticulum; most of these vesicles then migrate rapidly to the Golgi apparatus.

Other Functions of the Endoplasmic Reticulum. Other significant functions of the endoplasmic reticulum, especially the smooth reticulum, include the following:

1. It provides the enzymes that control glycogen breakdown when glycogen is to be used for energy.
2. It provides a vast number of enzymes that are capable of detoxifying substances, such as drugs, that might damage the cell. It achieves detoxification by coagulation, oxidation, hydrolysis, conjugation with glycuronic acid, and in other ways.

Specific Functions of the Golgi Apparatus

Synthetic Functions of the Golgi Apparatus. Although the major function of the Golgi apparatus is to provide additional processing of substances already formed in the endoplasmic reticulum, it also has the capability of synthesizing certain carbohydrates that cannot be formed in the endoplasmic reticulum. This is especially true for the formation of large saccharide polymers bound with small amounts of protein; important examples include *hyaluronic acid* and *chondroitin sulfate.*

A few of the many functions of hyaluronic acid and chondroitin sulfate in the body are as follows: (1) they are the major components of proteoglycans secreted in mucus and other glandular secretions; (2) they are the major components of the *ground substance* outside the cells in the interstitial spaces, acting as fillers between collagen fibers and cells; (3) they are principal components of the organic matrix in both cartilage and bone; and (4) they are important in many cell activities including migration and proliferation.

Processing of Endoplasmic Secretions by the Golgi Apparatus—Formation of Vesicles. Figure 2-13 summarizes the major functions of the endoplasmic reticulum and Golgi apparatus. As substances are formed in the endoplasmic reticulum, especially the proteins, they are transported through the tubules toward portions of the smooth endoplasmic reticulum that lie nearest the

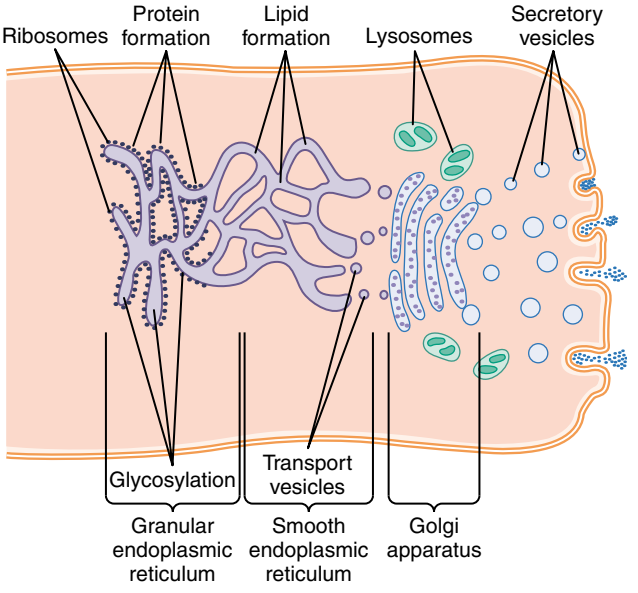


Figure 2-13 Formation of proteins, lipids, and cellular vesicles by the endoplasmic reticulum and Golgi apparatus.

Golgi apparatus. At this point, small *transport vesicles* composed of small envelopes of smooth endoplasmic reticulum continually break away and diffuse to the *deepest layer* of the Golgi apparatus. Inside these vesicles are the synthesized proteins and other products from the endoplasmic reticulum.

The transport vesicles instantly fuse with the Golgi apparatus and empty their contained substances into the vesicular spaces of the Golgi apparatus. Here, additional carbohydrate moieties are added to the secretions. Also, an important function of the Golgi apparatus is to compact the endoplasmic reticular secretions into highly concentrated packets. As the secretions pass toward the outermost layers of the Golgi apparatus, the compaction and processing proceed. Finally, both small and large vesicles continually break away from the Golgi apparatus, carrying with them the compacted secretory substances, and in turn, the vesicles diffuse throughout the cell.

To give an idea of the timing of these processes: When a glandular cell is bathed in radioactive amino acids, newly formed radioactive protein molecules can be detected in the granular endoplasmic reticulum within 3 to 5 minutes. Within 20 minutes, newly formed proteins are already present in the Golgi apparatus, and within 1 to 2 hours, radioactive proteins are secreted from the surface of the cell.

Types of Vesicles Formed by the Golgi Apparatus—Secretory Vesicles and Lysosomes. In a highly secretory cell, the vesicles formed by the Golgi apparatus are mainly *secretory vesicles* containing protein substances that are to be secreted through the surface of the cell membrane. These secretory vesicles first diffuse to the cell membrane, then fuse with it and empty their substances to the exterior by the mechanism called *exocytosis.* Exocytosis, in most cases, is stimulated by the entry of calcium ions into the cell; calcium ions interact with the vesicular membrane in some way that is not understood and cause its fusion with the cell membrane, followed by exocytosis—that is, opening of the membrane's outer surface and extrusion of its contents outside the cell.

Some vesicles, however, are destined for intracellular use.

Use of Intracellular Vesicles to Replenish Cellular Membranes. Some of the intracellular vesicles formed by the Golgi apparatus fuse with the cell membrane or with the membranes of intracellular structures such as the mitochondria and even the endoplasmic reticulum. This increases the expanse of these membranes and thereby replenishes the membranes as they are used up. For instance, the cell membrane loses much of its substance every time it forms a phagocytic or pinocytotic vesicle, and the vesicular membranes of the Golgi apparatus continually replenish the cell membrane.

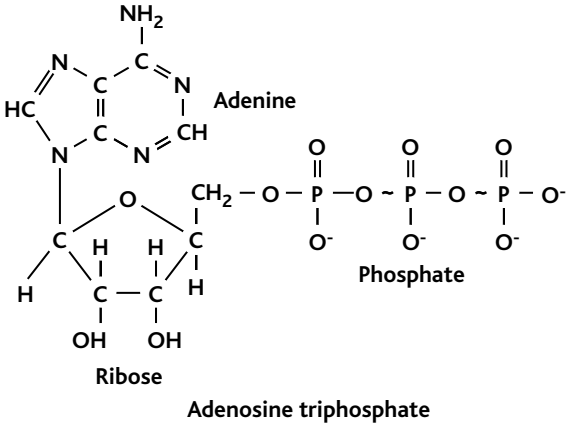
In summary, the membranous system of the endoplasmic reticulum and Golgi apparatus represents a highly metabolic organ capable of forming new intracellular structures, as well as secretory substances to be extruded from the cell.

Extraction of Energy from Nutrients—Function of the Mitochondria

The principal substances from which cells extract energy are foodstuffs that react chemically with oxygen—carbohydrates, fats, and proteins. In the human body, essentially all carbohydrates are converted into *glucose* by the digestive tract and liver before they reach the other cells of the body. Similarly, proteins are converted into *amino acids* and fats into *fatty acids.* Figure 2-14 shows oxygen and the foodstuffs—glucose, fatty acids, and amino acids—all entering the cell. Inside the cell, the foodstuffs react chemically with oxygen, under the influence of enzymes that control the reactions and channel the energy released in the proper direction. The details of all these digestive and metabolic functions are given in Chapters 62 through 72.

Briefly, almost all these oxidative reactions occur inside the mitochondria and the energy that is released is used to form the high-energy compound *ATP.* Then, ATP, not the original foodstuffs, is used throughout the cell to energize almost all the subsequent intracellular metabolic reactions.

Functional Characteristics of ATP



ATP is a nucleotide composed of (1) the nitrogenous base *adenine,* (2) the pentose sugar *ribose,* and (3) three *phosphate radicals.* The last two phosphate radicals are connected with the remainder of the molecule by so-called *high-energy phosphate bonds,* which are represented in the formula shown by the symbol ~. *Under the physical and chemical conditions of the body,* each of these high-energy bonds contains about 12,000 calories of energy per mole of ATP, which is many times greater than the energy stored in the average chemical bond, thus giving rise to the term *high-energy bond.* Further, the high-energy phosphate bond is very labile so that it can be split instantly on demand whenever energy is required to promote other intracellular reactions.

When ATP releases its energy, a phosphoric acid radical is split away and *adenosine diphosphate* (ADP) is formed. This released energy is used to energize virtually many of the cell's other functions, such as synthesis of substances and muscular contraction.

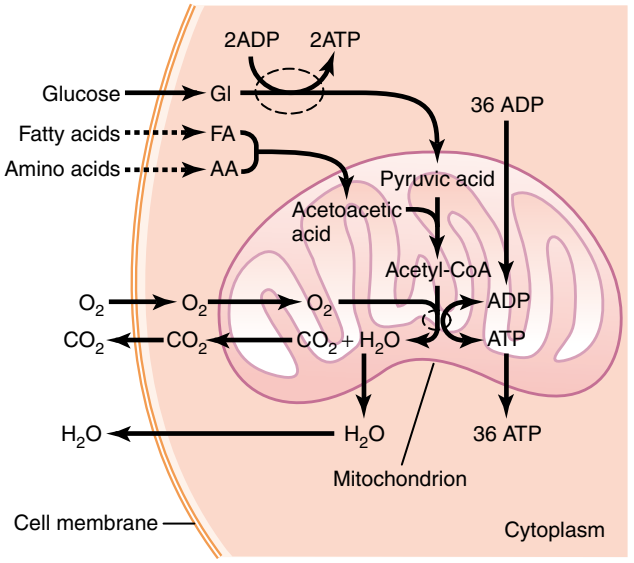


Figure 2-14 Formation of adenosine triphosphate (ATP) in the cell, showing that most of the ATP is formed in the mitochondria. ADP, adenosine diphosphate.

To reconstitute the cellular ATP as it is used up, energy derived from the cellular nutrients causes ADP and phosphoric acid to recombine to form new ATP, and the entire process repeats over and over again. For these reasons, ATP has been called the *energy currency* of the cell because it can be spent and remade continually, having a turnover time of only a few minutes.

Chemical Processes in the Formation of ATP—Role of the Mitochondria. On entry into the cells, glucose is subjected to enzymes in the *cytoplasm* that convert it into *pyruvic acid* (a process called *glycolysis*). A small amount of ADP is changed into ATP by the energy released during this conversion, but this amount accounts for less than 5 percent of the overall energy metabolism of the cell.

About 95 percent of the cell's ATP formation occurs in the mitochondria. The pyruvic acid derived from carbohydrates, fatty acids from lipids, and amino acids from proteins is eventually converted into the compound *acetyl-CoA* in the matrix of the mitochondrion. This substance, in turn, is further dissoluted (for the purpose of extracting its energy) by another series of enzymes in the mitochondrion matrix, undergoing dissolution in a sequence of chemical reactions called the *citric acid cycle,* or *Krebs cycle.* These chemical reactions are so important that they are explained in detail in Chapter 67.

In this citric acid cycle, acetyl-CoA is split into its component parts, *hydrogen atoms* and *carbon dioxide.* The carbon dioxide diffuses out of the mitochondria and eventually out of the cell; finally, it is excreted from the body through the lungs.

The hydrogen atoms, conversely, are highly reactive, and they combine instantly with oxygen that has also diffused into the mitochondria. This releases a tremendous amount of energy, which is used by the mitochondria to convert large amounts of ADP to ATP. The processes of these reactions are complex, requiring the participation of many protein enzymes that are integral parts of mitochondrial *membranous shelves* that protrude into the mitochondrial matrix. The initial event is removal of an electron from the hydrogen atom, thus converting it to a hydrogen ion. The terminal event is combination of hydrogen ions with oxygen to form water plus the release of tremendous amounts of energy to large globular proteins, called *ATP synthetase,* that protrude like knobs from the membranes of the mitochondrial shelves. Finally, the enzyme ATP synthetase uses the energy from the hydrogen ions to cause the conversion of ADP to ATP. The newly formed ATP is transported out of the mitochondria into all parts of the cell cytoplasm and nucleoplasm, where its energy is used to energize multiple cell functions.

This overall process for formation of ATP is called the *chemiosmotic mechanism* of ATP formation. The chemical and physical details of this mechanism are presented in Chapter 67, and many of the detailed metabolic functions of ATP in the body are presented in Chapters 67 through 71.

Uses of ATP for Cellular Function. Energy from ATP is used to promote three major categories of cellular functions: (1) *transport* of substances through multiple membranes in the cell, (2) *synthesis of chemical compounds* throughout the cell, and (3) *mechanical work.* These uses of ATP are illustrated by examples in Figure 2-15: (1) to supply energy for the transport of sodium through the cell membrane, (2) to promote protein synthesis by the ribosomes, and (3) to supply the energy needed during muscle contraction.

In addition to membrane transport of sodium, energy from ATP is required for membrane transport of potassium ions, calcium ions, magnesium ions, phosphate ions,

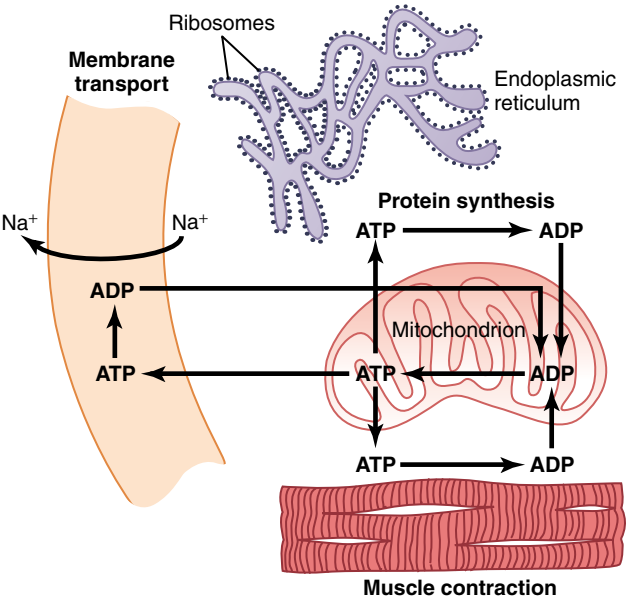


Figure 2-15 Use of adenosine triphosphate (ATP) (formed in the mitochondrion) to provide energy for three major cellular functions: membrane transport, protein synthesis, and muscle contraction. ADP, adenosine diphosphate.

chloride ions, urate ions, hydrogen ions, and many other ions and various organic substances. Membrane transport is so important to cell function that some cells—the renal tubular cells, for instance—use as much as 80 percent of the ATP that they form for this purpose alone.

In addition to synthesizing proteins, cells make phospholipids, cholesterol, purines, pyrimidines, and a host of other substances. Synthesis of almost any chemical compound requires energy. For instance, a single protein molecule might be composed of as many as several thousand amino acids attached to one another by peptide linkages; the formation of each of these linkages requires energy derived from the breakdown of four high-energy bonds; thus, many thousand ATP molecules must release their energy as each protein molecule is formed. Indeed, some cells use as much as 75 percent of all the ATP formed in the cell simply to synthesize new chemical compounds, especially protein molecules; this is particularly true during the growth phase of cells.

The final major use of ATP is to supply energy for special cells to perform mechanical work. We see in Chapter 6 that each contraction of a muscle fiber requires expenditure of tremendous quantities of ATP energy. Other cells perform mechanical work in other ways, especially by *ciliary* and *ameboid motion,* described later in this chapter. The source of energy for all these types of mechanical work is ATP.

In summary, ATP is always available to release its energy rapidly and almost explosively wherever in the cell it is needed. To replace the ATP used by the cell, much slower chemical reactions break down carbohydrates, fats, and proteins and use the energy derived from these to form new ATP. More than 95 percent of this ATP is formed in the mitochondria, which accounts for the mitochondria being called the "powerhouses" of the cell.

Locomotion of Cells

By far the most important type of movement that occurs in the body is that of the muscle cells in skeletal, cardiac, and smooth muscle, which constitute almost 50 percent of the entire body mass. The specialized functions of these cells are discussed in Chapters 6 through 9. Two other types of movement—*ameboid locomotion* and *ciliary movement*—occur in other cells.

Ameboid Movement

Ameboid movement is movement of an entire cell in relation to its surroundings, such as movement of white blood cells through tissues. It receives its name from the fact that amebae move in this manner and have provided an excellent tool for studying the phenomenon.

Typically, ameboid locomotion begins with protrusion of a *pseudopodium* from one end of the cell. The pseudopodium projects far out, away from the cell body, and partially secures itself in a new tissue area. Then the remainder of the cell is pulled toward the pseudopodium. Figure 2-16 demonstrates this process, showing an elongated cell, the right-hand end of which is a protruding pseudopodium. The membrane of this end of the cell is continually moving forward, and the membrane at the left-hand end of the cell is continually following along as the cell moves.

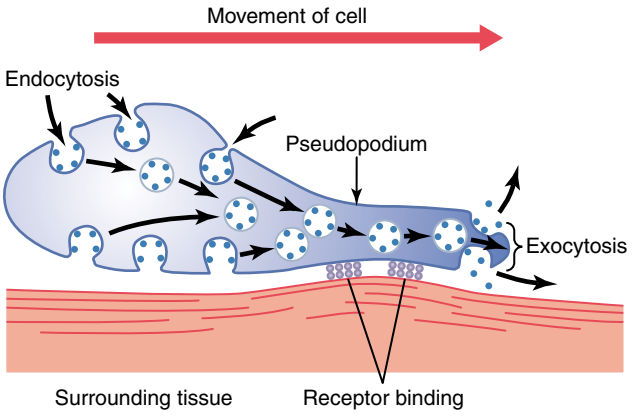


Figure 2-16 Ameboid motion by a cell.

Mechanism of Ameboid Locomotion. Figure 2-16 shows the general principle of ameboid motion. Basically, it results from continual formation of new cell membrane at the leading edge of the pseudopodium and continual absorption of the membrane in mid and rear portions of the cell. Also, two other effects are essential for forward movement of the cell. The first effect is attachment of the pseudopodium to surrounding tissues so that it becomes fixed in its leading position, while the remainder of the cell body is pulled forward toward the point of attachment. This attachment is effected by *receptor proteins* that line the insides of exocytotic vesicles. When the vesicles become part of the pseudopodial membrane, they open so that their insides evert to the outside, and the receptors now protrude to the outside and attach to ligands in the surrounding tissues.

At the opposite end of the cell, the receptors pull away from their ligands and form new endocytotic vesicles. Then, inside the cell, these vesicles stream toward the pseudopodial end of the cell, where they are used to form still new membrane for the pseudopodium.

The second essential effect for locomotion is to provide the energy required to pull the cell body in the direction of the pseudopodium. Experiments suggest the following as an explanation: In the cytoplasm of all cells is a moderate to large amount of the protein *actin.* Much of the actin is in the form of single molecules that do not provide any motive power; however, these polymerize to form a filamentous network, and the network contracts when it binds with an actin-binding protein such as *myosin.* The whole process is energized by the high-energy compound ATP. This is what happens in the pseudopodium of a moving cell, where such a network of actin filaments forms anew inside the enlarging pseudopodium. Contraction also occurs in the ectoplasm of the cell body, where a preexisting actin network is already present beneath the cell membrane.

Types of Cells That Exhibit Ameboid Locomotion. The most common cells to exhibit ameboid locomotion in the human body are the *white blood cells* when they move out of the blood into the tissues to form *tissue macrophages.* Other types of cells can also move by ameboid locomotion under certain circumstances. For instance, fibroblasts move into a damaged area to help repair the damage and even the germinal cells of the skin, though ordinarily completely sessile cells, move toward a cut area to repair the opening. Finally, cell locomotion is especially important in development of the embryo and fetus after fertilization of an ovum. For instance, embryonic cells often must migrate long distances from their sites of origin to new areas during development of special structures.

Control of Ameboid Locomotion—Chemotaxis. The most important initiator of ameboid locomotion is the process called *chemotaxis.* This results from the appearance of certain chemical substances in the tissues. Any chemical substance that causes chemotaxis to occur is called a *chemotactic substance.* Most cells that exhibit ameboid locomotion move toward the source of a chemotactic substance—that is, from an area of lower concentration toward an area of higher concentration—which is called *positive chemotaxis.* Some cells move away from the source, which is called *negative chemotaxis.*

But how does chemotaxis control the direction of ameboid locomotion? Although the answer is not certain, it is known that the side of the cell most exposed to the chemotactic substance develops membrane changes that cause pseudopodial protrusion.

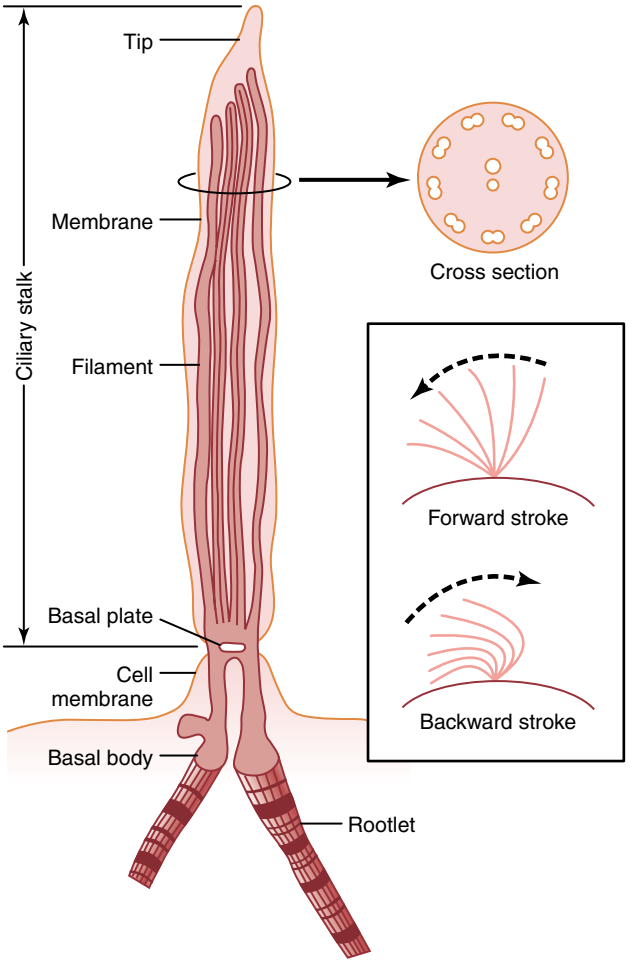


Figure 2-17 Structure and function of the cilium. (Modified from Satir P: Cilia. Sci Am 204:108, 1961. Copyright Donald Garber: Executor of the estate of Bunji Tagawa.)

Cilia and Ciliary Movements

A second type of cellular motion, *ciliary movement,* is a whiplike movement of cilia on the surfaces of cells. This occurs in only two places in the human body: on the surfaces of the respiratory airways and on the inside surfaces of the uterine tubes (fallopian tubes) of the reproductive tract. In the nasal cavity and lower respiratory airways, the whiplike motion of cilia causes a layer of mucus to move at a rate of about 1 cm/min toward the pharynx, in this way continually clearing these passageways of mucus and particles that have become trapped in the mucus. In the uterine tubes, the cilia cause slow movement of fluid from the ostium of the uterine tube toward the uterus cavity; this movement of fluid transports the ovum from the ovary to the uterus.

As shown in Figure 2-17, a cilium has the appearance of a sharp-pointed straight or curved hair that projects 2 to 4 micrometers from the surface of the cell. Many cilia often project from a single cell—for instance, as many as 200 cilia on the surface of each epithelial cell inside the respiratory passageways. The cilium is covered by an outcropping of the cell membrane, and it is supported by 11 microtubules—9 double tubules located around the periphery of the cilium and 2 single tubules down the center, as demonstrated in the cross section shown in Figure 2-17. Each cilium is an outgrowth of a structure that lies immediately beneath the cell membrane, called the *basal body* of the cilium.

The *flagellum of a sperm* is similar to a cilium; in fact, it has much the same type of structure and same type of contractile mechanism. The flagellum, however, is much longer and moves in quasi-sinusoidal waves instead of whiplike movements.

In the inset of Figure 2-17, movement of the cilium is shown. The cilium moves forward with a sudden, rapid whiplike stroke 10 to 20 times per second, bending sharply where it projects from the surface of the cell. Then it moves backward slowly to its initial position. The rapid forward-thrusting, whiplike movement pushes the fluid lying adjacent to the cell in the direction that the cilium moves; the slow, dragging movement in the backward direction has almost no effect on fluid movement. As a result, the fluid is continually propelled in the direction of the fast-forward stroke. Because most ciliated cells have large numbers of cilia on their surfaces and because all the cilia are oriented in the same direction, this is an effective means for moving fluids from one part of the surface to another.

Mechanism of Ciliary Movement. Although not all aspects of ciliary movement are clear, we do know the following: First, the nine double tubules and the two single tubules are all linked to one another by a complex of protein cross-linkages; this total complex of tubules and cross-linkages is called the *axoneme.* Second, even after removal of the membrane and destruction of other elements of the cilium besides the axoneme, the cilium can still beat under appropriate conditions. Third, there are two necessary conditions for continued beating of the axoneme after removal of the other structures of the cilium: (1) the availability of ATP and (2) appropriate ionic conditions, especially appropriate concentrations of magnesium and calcium. Fourth, during forward motion of the cilium, the double tubules on the front edge of the cilium slide outward toward the tip of the cilium, while those on the back edge remain in place. Fifth, multiple protein arms composed of the protein *dynein,* which has ATPase enzymatic activity, project from each double tubule toward an adjacent double tubule.

Given this basic information, it has been determined that the release of energy from ATP in contact with the ATPase dynein arms causes the heads of these arms to "crawl" rapidly along the surface of the adjacent double tubule. If the front tubules crawl outward while the back tubules remain stationary, this will cause bending.

The way in which cilia contraction is controlled is not understood. The cilia of some genetically abnormal cells do not have the two central single tubules, and these cilia fail to beat. Therefore, it is presumed that some signal, perhaps an electrochemical signal, is transmitted along these two central tubules to activate the dynein arms.

Bibliography

Alberts B, Johnson A, Lewis J, et al: *Molecular Biology of the Cell*, 6th ed, New York, 2007, Garland Science.

Bonifacino JS, Glick BS: The mechanisms of vesicle budding and fusion, *Cell* 116:153, 2004.

Chacinska A, Koehler CM, Milenkovic D, Lithgow T, Pfanner N: Importing mitochondrial proteins: machineries and mechanisms, *Cell* 138:628, 2009.

Cohen AW, Hnasko R, Schubert W, Lisanti MP: Role of caveolae and caveolins in health and disease, *Physiol Rev* 84:1341, 2004.

Danial NN, Korsmeyer SJ: Cell death: critical control points, *Cell* 116:205, 2004.

Dröge W: Free radicals in the physiological control of cell function, *Physiol Rev* 82:47, 2002.

Edidin M: Lipids on the frontier: a century of cell-membrane bilayers, *Nat Rev Mol Cell Biol* 4:414, 2003.

Ginger ML, Portman N, McKean PG: Swimming with protists: perception, motility and flagellum assembly, *Nat Rev Microbiol* 6:838, 2008.

Grant BD, Donaldson JG: Pathways and mechanisms of endocytic recycling, *Nat Rev Mol Cell Biol* 10:597, 2009.

Güttinger S, Laurell E, Kutay U: Orchestrating nuclear envelope disassembly and reassembly during mitosis, *Nat Rev Mol Cell Biol* 10:178, 2009.

Hamill OP, Martinac B: Molecular basis of mechanotransduction in living cells, *Physiol Rev* 81:685, 2001.

Hock MB, Kralli A: Transcriptional control of mitochondrial biogenesis and function, *Annu Rev Physiol* 71:177, 2009.

Liesa M, Palacín M, Zorzano A: Mitochondrial dynamics in mammalian health and disease, *Physiol Rev* 89:799, 2009.

Mattaj IW: Sorting out the nuclear envelope from the endoplasmic reticulum, *Nat Rev Mol Cell Biol* 5:65, 2004.

Parton RG, Simons K: The multiple faces of caveolae, *Nat Rev Mol Cell Biol* 8:185, 2007.

Raiborg C, Stenmark H: The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins, *Nature* 458:445, 2009.

Ridley AJ, Schwartz MA, Burridge K, et al: Cell migration: integrating signals from front to back, *Science* 302:1704, 2003.

Saftig P, Klumperman J: Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function, *Nat Rev Mol Cell Biol* 10:623, 2009.

Scarpulla RC: Transcriptional paradigms in mammalian mitochondrial biogenesis and function, *Physiol Rev* 88:611, 2008.

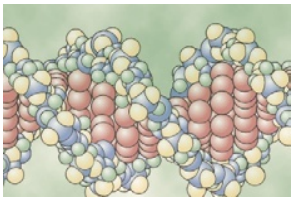
Stenmark H: Rab GTPases as coordinators of vesicle traffic, *Nat Rev Mol Cell Biol* 10:513, 2009.

Traub LM: Tickets to ride: selecting cargo for clathrin-regulated internalization, *Nat Rev Mol Cell Biol* 10:583, 2009.

Vereb G, Szollosi J, Matko J, et al: Dynamic, yet structured: the cell membrane three decades after the Singer-Nicolson model, *Proc Natl Acad Sci U S A* 100:8053, 2003.

CHAPTER 3

Genetic Control of Protein Synthesis, Cell Function, and Cell Reproduction



Virtually everyone knows that the genes, located in the nuclei of all cells of the body, control heredity from parents to children, but most people do not realize that these same genes also control day-to-day function of all the body's cells. The genes control cell function by determining which substances are synthesized within the cell—which structures, which enzymes, which chemicals.

Figure 3-1 shows the general schema of genetic control. Each gene, which is a nucleic acid called *deoxyribonucleic acid* (DNA), automatically controls the formation of another nucleic acid, *ribonucleic acid* (RNA); this RNA then spreads throughout the cell to control the formation of a specific protein. The entire process, from *transcription* of the genetic code in the nucleus to *translation* of the RNA code and formation or proteins in the cell cytoplasm, is often referred to as *gene expression.*

Because there are approximately 30,000 different genes in each cell, it is theoretically possible to form a large number of different cellular proteins.

Some of the cellular proteins are *structural proteins,* which, in association with various lipids and carbohydrates, form the structures of the various intracellular organelles discussed in Chapter 2. However, the majority of the proteins are *enzymes* that catalyze the different chemical reactions in the cells. For instance, enzymes promote all the oxidative reactions that supply energy to the cell, and they promote synthesis of all the cell chemicals, such as lipids, glycogen, and adenosine triphosphate (ATP).

Genes in the Cell Nucleus

In the cell nucleus, large numbers of genes are attached end on end in extremely long double-stranded helical molecules of DNA having molecular weights measured in the billions. A very short segment of such a molecule is shown in Figure 3-2. This molecule is composed of several simple chemical compounds bound together in a regular pattern, details of which are explained in the next few paragraphs.

Basic Building Blocks of DNA. Figure 3-3 shows the basic chemical compounds involved in the formation of DNA. These include (1) *phosphoric acid,* (2) a sugar called *deoxyribose,* and (3) four nitrogenous *bases* (two purines, *adenine* and *guanine,* and two pyrimidines, *thymine* and *cytosine*). The phosphoric acid and deoxyribose form the two helical strands that are the backbone of the DNA molecule, and the nitrogenous bases lie between the two strands and connect them, as illustrated in Figure 3-6.

Nucleotides. The first stage in the formation of DNA is to combine one molecule of phosphoric acid, one molecule of deoxyribose, and one of the four bases to form an acidic nucleotide. Four separate nucleotides are thus formed, one for each of the four bases: *deoxyadenylic, deoxythymidylic, deoxyguanylic,* and *deoxycytidylic acids.* Figure 3-4 shows the chemical structure of deoxyadenylic

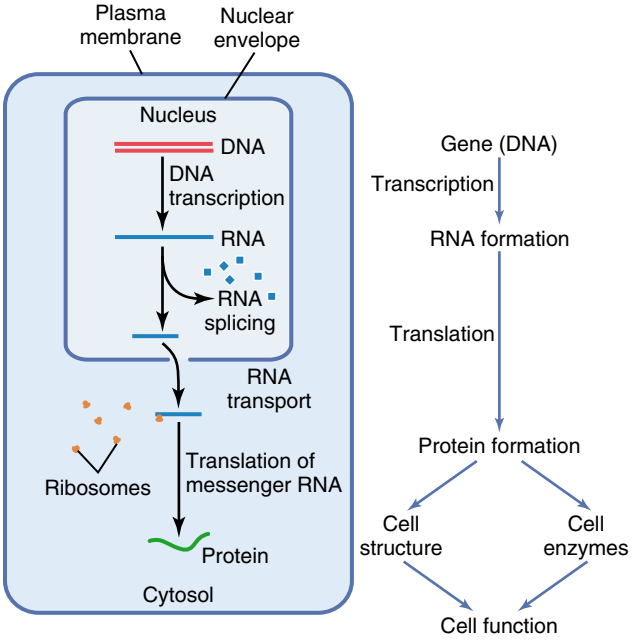


Figure 3-1. General schema by which the genes control cell function.

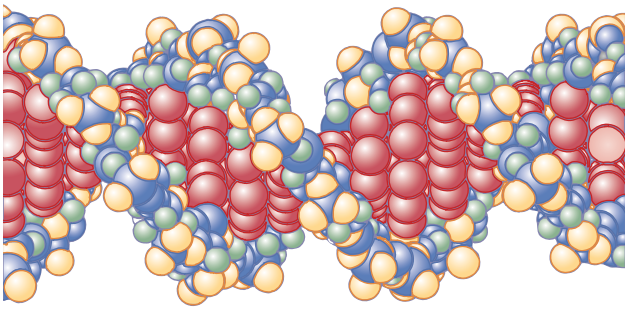


Figure 3-2. The helical, double-stranded structure of the gene. The outside strands are composed of phosphoric acid and the sugar deoxyribose. The internal molecules connecting the two strands of the helix are purine and pyrimidine bases; these determine the "code" of the gene.

acid, and Figure 3-5 shows simple symbols for the four nucleotides that form DNA.

Organization of the Nucleotides to Form Two Strands of DNA Loosely Bound to Each Other. Figure 3-6 shows the manner in which multiple numbers of nucleotides are bound together to form two strands of DNA. The two strands are, in turn, loosely bonded with each other by weak cross-linkages, illustrated in Figure 3-6 by the central dashed lines. Note that the backbone of

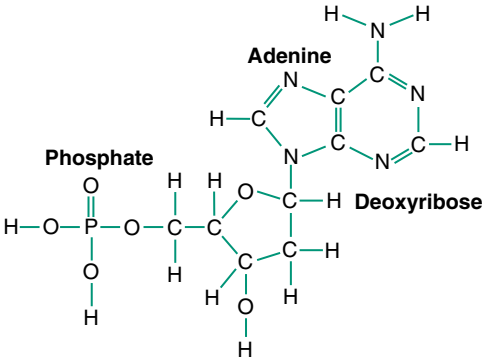
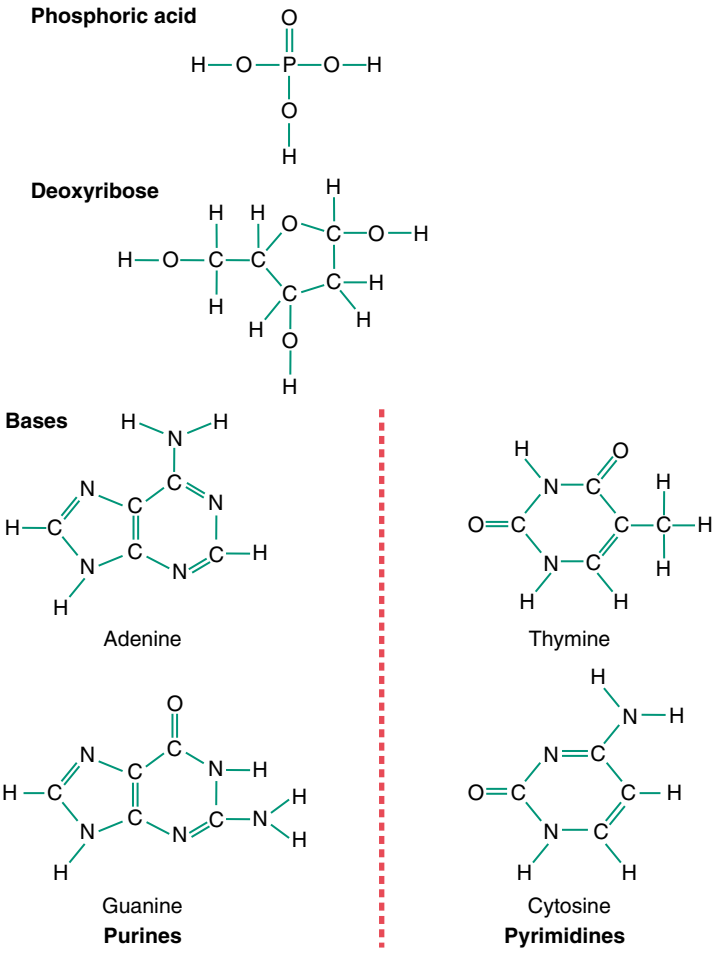


Figure 3-4. Deoxyadenylic acid, one of the nucleotides that make up DNA.

each DNA strand is composed of alternating phosphoric acid and deoxyribose molecules. In turn, purine and pyrimidine bases are attached to the sides of the deoxyribose molecules. Then, by means of loose *hydrogen bonds* (dashed lines) between the purine and pyrimidine bases, the two respective DNA strands are held together. But note the following:

1. Each purine base *adenine* of one strand always bonds with a pyrimidine base *thymine* of the other strand, and
2. Each purine base *guanine* always bonds with a pyrimidine base *cytosine.*

Figure 3-3. The basic building blocks of DNA.



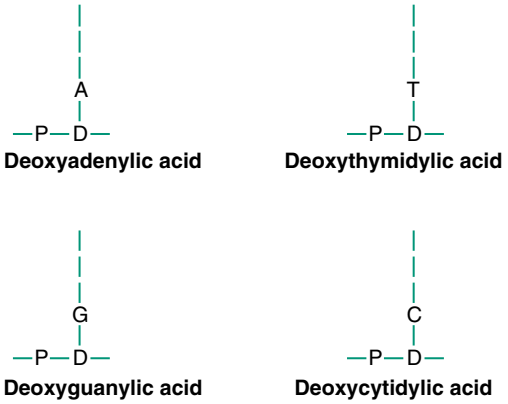


Figure 3-5. Symbols for the four nucleotides that combine to form DNA. Each nucleotide contains phosphoric acid (*P*), deoxyribose (*D*), and one of the four nucleotide bases: *A,* adenine; *T,* thymine; *G,* guanine; or *C,* cytosine.

Thus, in Figure 3-6, the sequence of complementary pairs of bases is CG, CG, GC, TA, CG, TA, GC, AT, and AT. Because of the looseness of the hydrogen bonds, the two strands can pull apart with ease, and they do so many times during the course of their function in the cell.

To put the DNA of Figure 3-6 into its proper physical perspective, one could merely pick up the two ends and twist them into a helix. Ten pairs of nucleotides are present in each full turn of the helix in the DNA molecule, as shown in Figure 3-2.

Genetic Code

The importance of DNA lies in its ability to control the formation of proteins in the cell. It does this by means of a *genetic code.* That is, when the two strands of a DNA molecule are split apart, this exposes the purine and pyrimidine bases projecting to the side of each DNA strand, as shown by the top strand in Figure 3-7. It is these projecting bases that form the genetic code.

The genetic code consists of successive "triplets" of bases—that is, each three successive bases is a *code word.* The successive triplets eventually control the sequence of amino acids in a protein molecule that is to be synthesized in the cell. Note in Figure 3-6 that the top strand of DNA, reading from left to right, has the genetic code GGC, AGA, CTT, the triplets being separated from one another by the arrows. As we follow this genetic code through Figures 3-7 and 3-8, we see that these three respective triplets are responsible for successive placement of the three amino acids, *proline, serine,* and *glutamic acid,* in a newly formed molecule of protein.

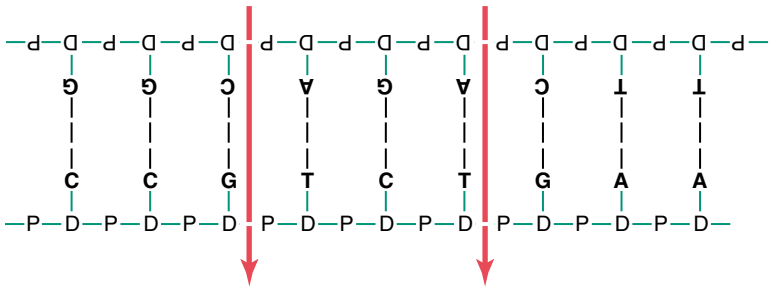


Figure 3-6. Arrangement of deoxyribose nucleotides in a double strand of DNA.

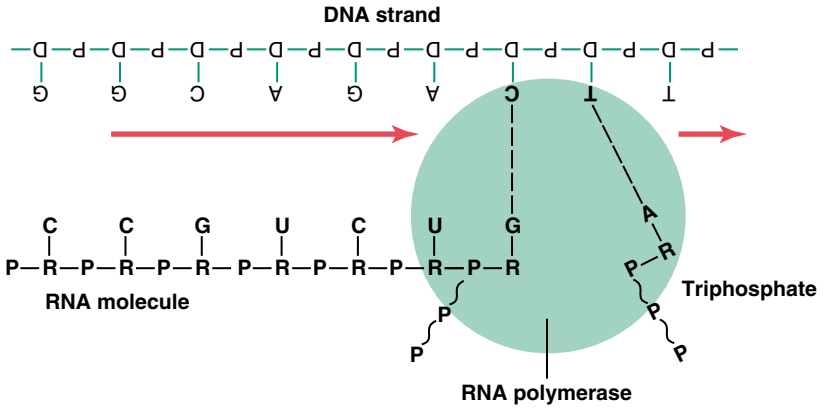


Figure 3-7. Combination of ribose nucleotides with a strand of DNA to form a molecule of RNA that carries the genetic code from the gene to the cytoplasm. The *RNA polymerase* enzyme moves along the DNA strand and builds the RNA molecule.

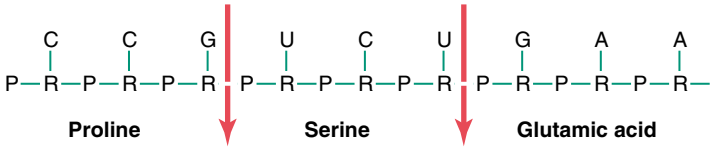


Figure 3-8. Portion of an RNA molecule, showing three RNA "codons"—CCG, UCU, and GAA—that control attachment of the three amino acids, *proline, serine,* and *glutamic acid,* respectively, to the growing RNA chain.

The DNA Code in the Cell Nucleus Is Transferred to an RNA Code in the Cell Cytoplasm—The Process of Transcription

Because the DNA is located in the nucleus of the cell, yet most of the functions of the cell are carried out in the cytoplasm, there must be some means for the DNA genes of the nucleus to control the chemical reactions of the cytoplasm. This is achieved through the intermediary of another type of nucleic acid, RNA, the formation of which is controlled by the DNA of the nucleus. Thus, as shown in Figure 3-7, the code is transferred to the RNA; this process is called *transcription.* The RNA, in turn, diffuses from the nucleus through nuclear pores into the cytoplasmic compartment, where it controls protein synthesis.

Synthesis of RNA

During synthesis of RNA, the two strands of the DNA molecule separate temporarily; one of these strands is used as a template for synthesis of an RNA molecule. The code triplets in the DNA cause formation of *complementary* code triplets (called *codons*) in the RNA; these codons, in turn, will control the sequence of amino acids in a protein to be synthesized in the cell cytoplasm.

Basic Building Blocks of RNA. The basic building blocks of RNA are almost the same as those of DNA, except for two differences. First, the sugar deoxyribose is not used in the formation of RNA. In its place is another sugar of slightly different composition, *ribose,* containing an extra hydroxyl ion appended to the ribose ring structure. Second, thymine is replaced by another pyrimidine, *uracil.*

Formation of RNA Nucleotides. The basic building blocks of RNA form *RNA nucleotides,* exactly as previously described for DNA synthesis. Here again, four separate nucleotides are used in the formation of RNA. These nucleotides contain the bases *adenine, guanine, cytosine,* and *uracil.* Note that these are the same bases as in DNA, except that uracil in RNA replaces thymine in DNA.

"Activation" of the RNA Nucleotides. The next step in the synthesis of RNA is "activation" of the RNA nucleotides by an enzyme, *RNA polymerase.* This occurs by adding to each nucleotide two extra phosphate radicals to form triphosphates (shown in Figure 3-7 by the two RNA nucleotides to the far right during RNA chain formation). These last two phosphates are combined with the nucleotide by *high-energy phosphate bonds* derived from ATP in the cell.

The result of this activation process is that large quantities of ATP energy are made available to each of the nucleotides, and this energy is used to promote the chemical reactions that add each new RNA nucleotide at the end of the developing RNA chain.

Assembly of the RNA Chain from Activated Nucleotides Using the DNA Strand as a Template—The Process of "Transcription"

Assembly of the RNA molecule is accomplished in the manner shown in Figure 3-7 under the influence of an enzyme, *RNA polymerase.* This is a large protein enzyme that has many functional properties necessary for formation of the RNA molecule. They are as follows:

1. In the DNA strand immediately ahead of the initial gene is a sequence of nucleotides called the *promoter.* The RNA polymerase has an appropriate complementary structure that recognizes this promoter and becomes attached to it. This is the essential step for initiating formation of the RNA molecule.
2. After the RNA polymerase attaches to the promoter, the polymerase causes unwinding of about two turns of the DNA helix and separation of the unwound portions of the two strands.
3. Then the polymerase moves along the DNA strand, temporarily unwinding and separating the two DNA strands at each stage of its movement. As it moves along, it adds at each stage a new activated RNA nucleotide to the end of the newly forming RNA chain by the following steps:
   - **a.** First, it causes a hydrogen bond to form between the end base of the DNA strand and the base of an RNA nucleotide in the nucleoplasm.
   - **b.** Then, one at a time, the RNA polymerase breaks two of the three phosphate radicals away from each of these RNA nucleotides, liberating large amounts of energy from the broken high-energy phosphate bonds; this energy is used to cause covalent linkage of the remaining phosphate on the nucleotide with the ribose on the end of the growing RNA chain.
   - **c.** When the RNA polymerase reaches the end of the DNA gene, it encounters a new sequence of DNA nucleotides called the *chain-terminating sequence;* this causes the polymerase and the newly formed RNA chain to break away from the DNA strand. Then the polymerase can be used again and again to form still more new RNA chains.
   - **d.** As the new RNA strand is formed, its weak hydrogen bonds with the DNA template break away, because the DNA has a high affinity for rebonding with its own complementary DNA strand. Thus, the RNA chain is forced away from the DNA and is released into the nucleoplasm.

Thus, the code that is present in the DNA strand is eventually transmitted in *complementary* form to the RNA chain. The ribose nucleotide bases always combine with the deoxyribose bases in the following combinations:

DNA Base	RNA Base
guanine	cytosine
cytosine	guanine
adenine	uracil
thymine	adenine

Four Different Types of RNA. Each type of RNA plays an independent and entirely different role in protein formation:

1. *Messenger RNA* (mRNA), which carries the genetic code to the cytoplasm for controlling the type of protein formed.
2. *Transfer RNA* (tRNA), which transports activated amino acids to the ribosomes to be used in assembling the protein molecule.
3. *Ribosomal RNA,* which, along with about 75 different proteins, forms *ribosomes,* the physical and chemical structures on which protein molecules are actually assembled.
4. *MicroRNA* (miRNA), which are single-stranded RNA molecules of 21 to 23 nucleotides that can regulate gene transcription and translation.

Messenger RNA—The Codons

mRNA molecules are long, single RNA strands that are suspended in the cytoplasm. These molecules are composed of several hundred to several thousand RNA nucleotides in unpaired strands, and they contain *codons* that are exactly complementary to the code triplets of the DNA genes. Figure 3-8 shows a small segment of a molecule of messenger RNA. Its codons are CCG, UCU, and GAA. These are the codons for the amino acids proline, serine, and glutamic acid. The transcription of these codons from the DNA molecule to the RNA molecule is shown in Figure 3-7.

RNA Codons for the Different Amino Acids. Table 3-1 gives the RNA codons for the 22 common amino acids found in protein molecules. Note that most of the amino acids are represented by more than one codon;

Table 3-1. RNA Codons for Amino Acids and for Start and Stop

Amino Acid	RNA Codons					
Alanine	GCU	GCC	GCA	GCG		
Arginine	CGU	CGC	CGA	CGG	AGA	AGG
Asparagine	AAU	AAC				
Aspartic acid	GAU	GAC				
Cysteine	UGU	UGC				
Glutamic acid	GAA	GAG				
Glutamine	CAA	CAG				
Glycine	GGU	GGC	GGA	GGG		
Histidine	CAU	CAC				
Isoleucine	AUU	AUC	AUA			
Leucine	CUU	CUC	CUA	CUG	UUA	UUG
Lysine	AAA	AAG				
Methionine	AUG					
Phenylalanine	UUU	UUC				
Proline	CCU	CCC	CCA	CCG		
Serine	UCU	UCC	UCA	UCG	AGC	AGU
Threonine	ACU	ACC	ACA	ACG		
Tryptophan	UGG					
Tyrosine	UAU	UAC				
Valine	GUU	GUC	GUA	GUG		
Start (CI)	AUG					
Stop (CT)	UAA	UAG	UGA			

CI, chain-initiating; CT, chain-terminating.

also, one codon represents the signal "start manufacturing the protein molecule," and three codons represent "stop manufacturing the protein molecule." In Table 3-1, these two types of codons are designated CI for "chain-initiating" and CT for "chain-terminating."

Transfer RNA—The Anticodons

Another type of RNA that plays an essential role in protein synthesis is called *tRNA* because it transfers amino acid molecules to protein molecules as the protein is being synthesized. Each type of tRNA combines specifically with 1 of the 20 amino acids that are to be incorporated into proteins. The tRNA then acts as a *carrier* to transport its specific type of amino acid to the ribosomes, where protein molecules are forming. In the ribosomes, each specific type of transfer RNA recognizes a particular codon on the mRNA (described later) and thereby delivers the appropriate amino acid to the appropriate place in the chain of the newly forming protein molecule.

Transfer RNA, which contains only about 80 nucleotides, is a relatively small molecule in comparison with mRNA. It is a folded chain of nucleotides with a cloverleaf appearance similar to that shown in Figure 3-9. At one end of the molecule is always an adenylic acid; it is to this that the transported amino acid attaches at a hydroxyl group of the ribose in the adenylic acid.

Because the function of tRNA is to cause attachment of a specific amino acid to a forming protein chain, it is essential that each type of tRNA also have specificity for a particular codon in the mRNA. The specific code in the tRNA that allows it to recognize a specific codon is again a triplet of nucleotide bases and is called an *anticodon.* This is located approximately in the middle of the tRNA molecule (at the bottom of the cloverleaf configuration shown in Figure 3-9). During formation of the protein molecule, the anticodon bases combine loosely by hydrogen bonding with the codon bases of the mRNA. In this way, the respective amino acids are lined up one after another along the mRNA chain, thus establishing the appropriate sequence of amino acids in the newly forming protein molecule.

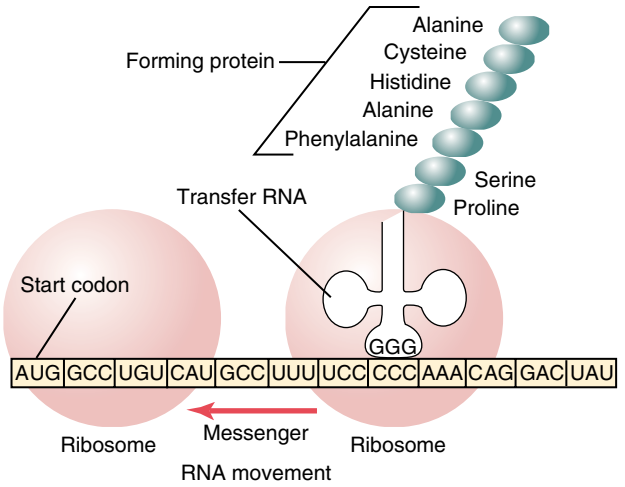


Figure 3-9. A messenger RNA strand is moving through two ribosomes. As each "codon" passes through, an amino acid is added to the growing protein chain, which is shown in the right-hand ribosome. The transfer RNA molecule transports each specific amino acid to the newly forming protein.

Ribosomal RNA

The third type of RNA in the cell is ribosomal RNA; it constitutes about 60 percent of the *ribosome.* The remainder of the ribosome is protein, containing about 75 types of proteins that are both structural proteins and enzymes needed in the manufacture of protein molecules.

The ribosome is the physical structure in the cytoplasm on which protein molecules are actually synthesized. However, it always functions in association with the other two types of RNA as well: *tRNA* transports amino acids to the ribosome for incorporation into the developing protein molecule, whereas *mRNA* provides the information necessary for sequencing the amino acids in proper order for each specific type of protein to be manufactured.

Thus, the ribosome acts as a manufacturing plant in which the protein molecules are formed.

Formation of Ribosomes in the Nucleolus. The DNA genes for formation of ribosomal RNA are located in five pairs of chromosomes in the nucleus, and each of these chromosomes contains many duplicates of these particular genes because of the large amounts of ribosomal RNA required for cellular function.

As the ribosomal RNA forms, it collects in the *nucleolus,* a specialized structure lying adjacent to the chromosomes. When large amounts of ribosomal RNA are being synthesized, as occurs in cells that manufacture large amounts of protein, the nucleolus is a large structure, whereas in cells that synthesize little protein, the nucleolus may not even be seen. Ribosomal RNA is specially processed in the nucleolus, where it binds with "ribosomal proteins" to form granular condensation products that are primordial subunits of ribosomes. These subunits are then released from the nucleolus and transported through the large pores of the nuclear envelope to almost all parts of the cytoplasm. After the subunits enter the cytoplasm, they are assembled to form mature, functional ribosomes. Therefore, proteins are formed in the cytoplasm of the cell, but not in the cell nucleus, because the nucleus does not contain mature ribosomes.

MicroRNA

A fourth type of RNA in the cell is *miRNA.* These are short (21 to 23 nucleotides) single-stranded RNA fragments that regulate gene expression (Figure 3-10). The miRNAs are encoded from the transcribed DNA of genes, but they are not translated into proteins and are therefore often called *noncoding RNA.* The miRNAs are processed by the cell into molecules that are complementary to mRNA and act to decrease gene expression. Generation of miRNAs involves special processing of longer primary precursor

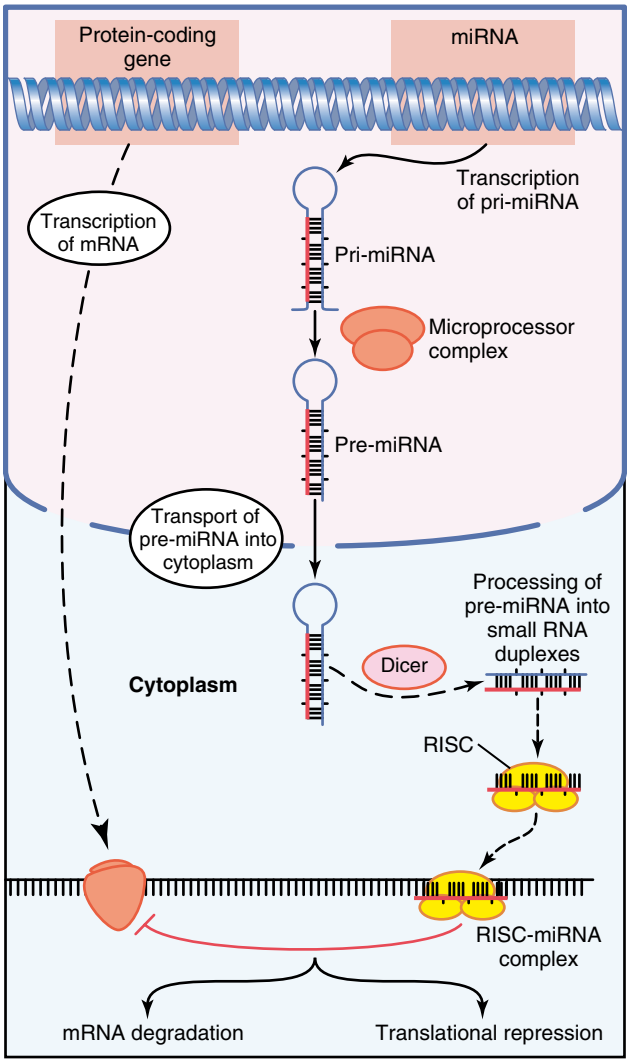


Figure 3-10. Regulation of gene expression by microRNA (miRNA). Primary miRNA (pri-miRNA), the primary transcripts of a gene processed in the cell nucleus by the microprocessor complex to pre-miRNAs. These pre-miRNAs are then further processed in the cytoplasm by *dicer,* an enzyme that helps assemble an RNA-induced silencing complex (RISC) and generates miRNAs. The miRNAs regulate gene expression by binding to the complementary region of the RNA and repressing translation or promoting degradation of the mRNA before it can be translated by the ribosome.

RNAs called *pri-miRNAs,* which are the primary transcripts of the gene. The pri-miRNAs are then processed in the cell nucleus by the *microprocessor complex* to pre-miRNAs, which are 70 nucleotide stem-loop structures. These pre-miRNAs are then further processed in the cytoplasm by a specific *dicer enzyme* that helps assemble an *RNA-induced silencing complex* (RISC) and generates miRNAs.

The miRNAs regulate gene expression by binding to the complementary region of the RNA and promoting repression of translation or degradation of the mRNA before it can be translated by the ribosome. miRNAs are believed to play an important role in the normal regulation of cell function, and alterations in miRNA function have been associated with diseases such as cancer and heart disease.

Another type of microRNA is *small interfering RNA* (siRNA), also called *silencing RNA* or *short interfering RNA.* The siRNAs are short, double-stranded RNA molecules, 20 to 25 nucleotides in length, that interfere with the expression of specific genes. siRNAs generally refer to synthetic miRNAs and can be administered to silence expression of specific genes. They are designed to avoid the nuclear processing by the microprocessor complex, and after the siRNA enters the cytoplasm it activates the RISC silencing complex, blocking the translation of mRNA. Because siRNAs can be tailored for any specific sequence in the gene, they can be used to block translation of any mRNA and therefore expression by any gene for which the nucleotide sequence is known. Some researchers have proposed that siRNAs may become useful therapeutic tools to silence genes that contribute to the pathophysiology of diseases.

Formation of Proteins on the Ribosomes—The Process of "Translation"

When a molecule of messenger RNA comes in contact with a ribosome, it travels through the ribosome, beginning at a predetermined end of the RNA molecule specified by an appropriate sequence of RNA bases called the "chain-initiating" codon. Then, as shown in Figure 3-9, while the messenger RNA travels through the ribosome, a protein molecule is formed—a process called *translation.* Thus, the ribosome reads the codons of the messenger RNA in much the same way that a tape is "read" as it passes through the playback head of a tape recorder. Then, when a "stop" (or "chain-terminating") codon slips past the ribosome, the end of a protein molecule is signaled and the protein molecule is freed into the cytoplasm.

Polyribosomes. A single messenger RNA molecule can form protein molecules in several ribosomes at the same time because the initial end of the RNA strand can pass to a successive ribosome as it leaves the first, as shown at the bottom left in Figure 3-9 and in Figure 3-11. The protein molecules are in different stages of development in each ribosome. As a result, clusters of ribosomes frequently occur, 3 to 10 ribosomes being attached to a single messenger RNA at the same time. These clusters are called *polyribosomes.*

It is especially important to note that a messenger RNA can cause the formation of a protein molecule in any ribosome; that is, there is no specificity of ribosomes for given types of protein. The ribosome is simply the physical manufacturing plant in which the chemical reactions take place.

Many Ribosomes Attach to the Endoplasmic Reticulum. In Chapter 2, it was noted that many ribosomes become attached to the endoplasmic reticulum. This occurs because the initial ends of many forming protein molecules have amino acid sequences that immediately attach to specific receptor sites on the endoplasmic reticulum; this causes these molecules to penetrate the

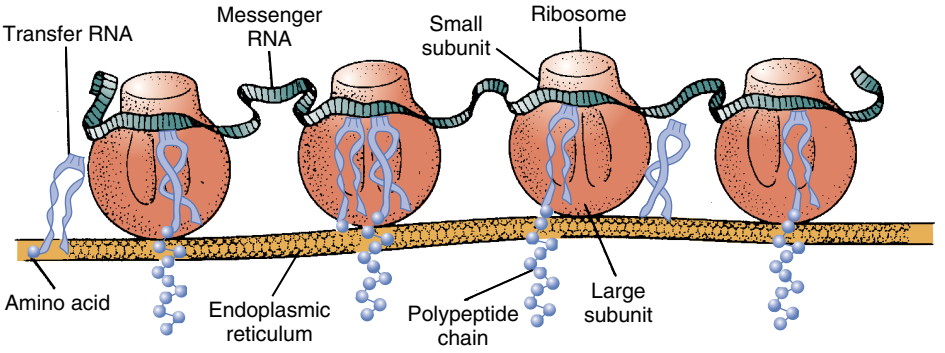


Figure 3-11. Physical structure of the ribosomes, as well as their functional relation to messenger RNA, transfer RNA, and the endoplasmic reticulum during the formation of protein molecules. (Courtesy Dr. Don W. Fawcett, Montana.)

reticulum wall and enter the endoplasmic reticulum matrix. This gives a granular appearance to those portions of the reticulum where proteins are being formed and entering the matrix of the reticulum.

Figure 3-11 shows the functional relation of messenger RNA to the ribosomes and the manner in which the ribosomes attach to the membrane of the endoplasmic reticulum. Note the process of translation occurring in several ribosomes at the same time in response to the same strand of messenger RNA. Note also the newly forming polypeptide (protein) chains passing through the endoplasmic reticulum membrane into the endoplasmic matrix.

Yet it should be noted that except in glandular cells in which large amounts of protein-containing secretory vesicles are formed, most proteins synthesized by the ribosomes are released directly into the cytosol instead of into the endoplasmic reticulum. These proteins are enzymes and internal structural proteins of the cell.

Chemical Steps in Protein Synthesis. Some of the chemical events that occur in synthesis of a protein molecule are shown in Figure 3-12. This figure shows representative reactions for three separate amino acids, AA_1, AA_2, and AA_{20}. The stages of the reactions are the following: (1) Each amino acid is *activated* by a chemical process in which ATP combines with the amino acid to form an *adenosine monophosphate complex with the amino acid,* giving up two high-energy phosphate bonds in the process. (2) The activated amino acid, having an excess of energy, then *combines with its specific transfer RNA to form an amino acid–tRNA complex* and, at the same time, releases the adenosine monophosphate. (3) The transfer RNA carrying the amino acid complex then comes in contact with the messenger RNA molecule in the ribosome, where the anticodon of the transfer RNA attaches temporarily to its specific codon of the messenger RNA, thus lining up the amino acid in appropriate sequence to form a protein molecule. Then, under the influence of the enzyme *peptidyl transferase* (one of the proteins in the ribosome), *peptide bonds* are formed between the successive amino acids, thus adding progressively to the protein chain. These chemical events require energy from two additional high-energy phosphate bonds, making a total of four high-energy bonds used for each amino acid added to the protein chain. Thus, the synthesis of proteins is one of the most energy-consuming processes of the cell.

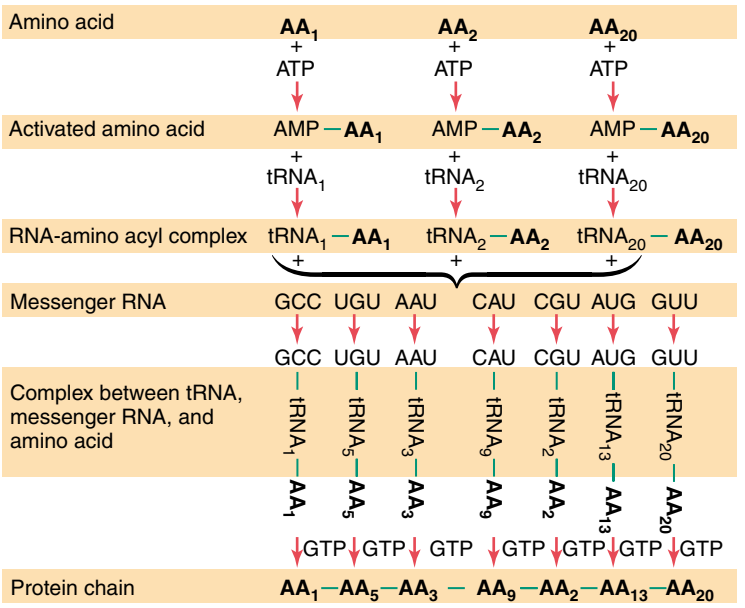
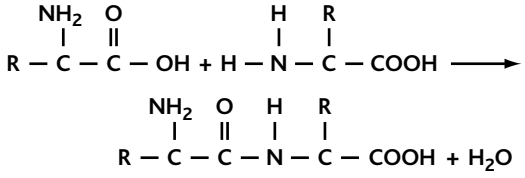


Figure 3-12. Chemical events in the formation of a protein molecule.

Peptide Linkage. The successive amino acids in the protein chain combine with one another according to the typical reaction:



In this chemical reaction, a hydroxyl radical (OH^-) is removed from the COOH portion of the first amino acid and a hydrogen (H^+) of the NH_2 portion of the other amino acid is removed. These combine to form water, and the two reactive sites left on the two successive amino acids bond with each other, resulting in a single molecule. This process is called *peptide linkage.* As each additional amino acid is added, an additional peptide linkage is formed.

Synthesis of Other Substances in the Cell

Many thousand protein enzymes formed in the manner just described control essentially all the other chemical reactions that take place in cells. These enzymes promote synthesis of lipids, glycogen, purines, pyrimidines, and hundreds of other substances. We discuss many of these synthetic processes in relation to carbohydrate, lipid, and protein metabolism in Chapters 67 through 69. It is by means of all these substances that the many functions of the cells are performed.

Control of Gene Function and Biochemical Activity in Cells

From our discussion thus far, it is clear that the genes control both the physical and chemical functions of the cells. However, the degree of activation of respective genes must be controlled as well; otherwise, some parts of the cell might overgrow or some chemical reactions might overact until they kill the cell. Each cell has powerful internal feedback control mechanisms that keep the various functional operations of the cell in step with one another. For each gene (approximately 30,000 genes in all), there is at least one such feedback mechanism.

There are basically two methods by which the biochemical activities in the cell are controlled: (1) *genetic regulation,* in which the degree of activation of the genes and the formation of gene products are themselves controlled and (2) *enzyme regulation,* in which the activity levels of already formed enzymes in the cell are controlled.

Genetic Regulation

Genetic regulation, or regulation of *gene expression,* covers the entire process from transcription of the genetic code in the nucleus to the formation or proteins in the cytoplasm. Regulation of gene expression provides all living organisms the ability to respond to changes in their environment. In animals that have many different types of cells, tissues, and organs, differential regulation of gene expression also permits the many different cell types in the body to each perform their specialized functions. Although a cardiac myocyte contains the same genetic code as a renal tubular epithelia cell, many genes are expressed in cardiac cells that are not expressed in renal tubular cells. The ultimate measure of gene "expression" is whether (and how much) of the gene products (proteins) are produced because proteins carry out cell functions specified by the genes. Regulation of gene expression can occur at any point in the pathways of transcription, RNA processing, and translation.

The Promoter Controls Gene Expression. Synthesis of cellular proteins is a complex process that starts with the transcription of DNA into RNA. The transcription of DNA is controlled by regulatory elements found in the promoter of a gene (Figure 3-13). In eukaryotes, which includes all mammals, the basal promoter consists of a sequence of seven bases (TATAAAA) called the *TATA box,* the binding site for the *TATA-binding protein* (TBP) and several other important *transcription factors* that are collectively referred to as the *transcription factor IID complex.* In addition to the transcription factor IID complex, this region is where transcription factor IIB binds to both the DNA and RNA polymerase 2 to facilitate transcription of the DNA into RNA. This basal promoter is found in all protein-coding genes and the polymerase must bind with this basal promoter before it can begin traveling along the DNA strand to synthesize RNA. The *upstream promoter* is located farther upstream from the transcription start site and contains several binding sites for positive or negative transcription factors that can effect transcription through interactions with proteins bound to the basal promoter. The structure and transcription factor binding sites in the upstream promoter vary from gene to gene to give rise to the different expression patterns of genes in different tissues.

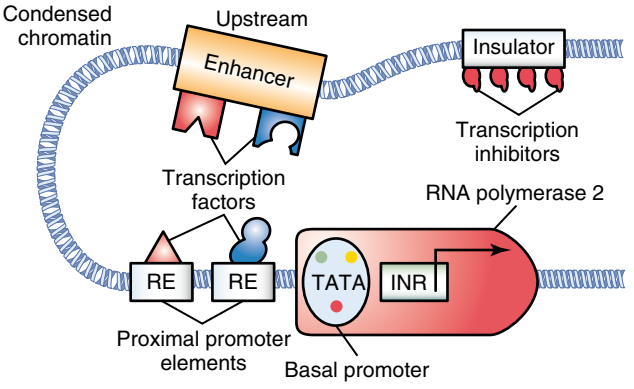


Figure 3-13. Gene transcriptional in eukaryotic cells. A complex arrangement of multiple clustered enhancer modules interspersed with insulator elements, which can be located either upstream or downstream of a basal promoter containing TATA box (TATA), proximal promoter elements (response elements, RE), and Initiator sequences (INR).

Transcription of genes in eukaryotes is also influenced by *enhancers,* which are regions of DNA that can bind transcription factors. Enhancers can be located a great distance from the gene they act on or even on a different chromosome. They can also be located either upstream or downstream of the gene that they regulate. Although enhancers may be located a great distance away from their target gene, they may be relatively close when DNA is coiled in the nucleus. It is estimated that there are 110,000 gene enhancer sequences in the human genome.

In the organization of the chromosome, it is important to separate active genes that are being transcribed from genes that are repressed. This can be challenging because multiple genes may be located close together on the chromosome. This is achieved by chromosomal *insulators.* These insulators are gene sequences that provide a barrier so that a specific gene is isolated against transcriptional influences from surrounding genes. Insulators can vary greatly in their DNA sequence and the proteins that bind to them. One way an insulator activity can be modulated is by *DNA methylation*. This is the case for the mammalian insulin-like growth factor 2 (IGF-2) gene. The mother's allele has an insulator between the enhancer and promoter of the gene that allows for the binding of a transcriptional repressor. However, the paternal DNA sequence is methylated such that the transcriptional repressor cannot bind to the insulator and the IGF-2 gene is expressed from the paternal copy of the gene.

Other Mechanisms for Control of Transcription by the Promoter. Variations in the basic mechanism for control of the promoter have been discovered with rapidity in the past 2 decades. Without giving details, let us list some of them:

1. A promoter is frequently controlled by transcription factors located elsewhere in the genome. That is, the regulatory gene causes the formation of a regulatory protein that in turn acts either as an activator or a repressor of transcription.
2. Occasionally, many different promoters are controlled at the same time by the same regulatory protein. In some instances, the same regulatory protein functions as an activator for one promoter and as a repressor for another promoter.
3. Some proteins are controlled not at the starting point of transcription on the DNA strand but farther along the strand. Sometimes the control is not even at the DNA strand itself but during the processing of the RNA molecules in the nucleus before they are released into the cytoplasm; rarely, control might occur at the level of protein formation in the cytoplasm during RNA translation by the ribosomes.
4. In nucleated cells, the nuclear DNA is packaged in specific structural units, the *chromosomes.* Within each chromosome, the DNA is wound around small proteins called *histones,* which in turn are held tightly together in a compacted state by still other proteins. As long as the DNA is in this compacted state, it cannot function to form RNA. However, multiple control mechanisms are beginning to be discovered that can cause selected areas of chromosomes to become decompacted one part at a time so that partial RNA transcription can occur. Even then, specific *transcriptor factor*s control the actual rate of transcription by the promoter in the chromosome. Thus, still higher orders of control are used for establishing proper cell function. In addition, signals from outside the cell, such as some of the body's hormones, can activate specific chromosomal areas and specific transcription factors, thus controlling the chemical machinery for function of the cell.

Because there are more than 30,000 different genes in each human cell, the large number of ways in which genetic activity can be controlled is not surprising. The gene control systems are especially important for controlling intracellular concentrations of amino acids, amino acid derivatives, and intermediate substrates and products of carbohydrate, lipid, and protein metabolism.

Control of Intracellular Function by Enzyme Regulation

In addition to control of cell function by genetic regulation, some cell activities are controlled by intracellular inhibitors or activators that act directly on specific intracellular enzymes. Thus, enzyme regulation represents a second category of mechanisms by which cellular biochemical functions can be controlled.

Enzyme Inhibition. Some chemical substances formed in the cell have direct feedback effects in inhibiting the specific enzyme systems that synthesize them. Almost always the synthesized product acts on the first enzyme in a sequence, rather than on the subsequent enzymes, usually binding directly with the enzyme and causing an allosteric conformational change that inactivates it. One can readily recognize the importance of inactivating the first enzyme: this prevents buildup of intermediary products that are not used.

Enzyme inhibition is another example of negative feedback control; it is responsible for controlling intracellular concentrations of multiple amino acids, purines, pyrimidines, vitamins, and other substances.

Enzyme Activation. Enzymes that are normally inactive often can be activated when needed. An example of this occurs when most of the ATP has been depleted in a cell. In this case, a considerable amount of cyclic adenosine monophosphate (cAMP) begins to be formed as a breakdown product of the ATP; the presence of this cAMP, in turn, immediately activates the glycogen-splitting enzyme phosphorylase, liberating glucose molecules that are rapidly metabolized and their energy used for replenishment of the ATP stores. Thus, cAMP acts as an enzyme activator for the enzyme phosphorylase and thereby helps control intracellular ATP concentration.

Another interesting instance of both enzyme inhibition and enzyme activation occurs in the formation of the purines and pyrimidines. These substances are needed by the cell in approximately equal quantities for formation of DNA and RNA. When purines are formed, they *inhibit* the enzymes that are required for formation of additional purines. However, they *activate* the enzymes for formation of pyrimidines. Conversely, the pyrimidines inhibit their own enzymes but activate the purine enzymes. In this way, there is continual cross-feed between the synthesizing systems for these two substances, resulting in almost exactly equal amounts of the two substances in the cells at all times.

Summary. In summary, there are two principal methods by which cells control proper proportions and proper quantities of different cellular constituents: (1) the mechanism of genetic regulation and (2) the mechanism of enzyme regulation. The genes can be either activated or inhibited, and likewise, the enzyme systems can be either activated or inhibited. These regulatory mechanisms most often function as feedback control systems that continually monitor the cell's biochemical composition and make corrections as needed. But on occasion, substances from without the cell (especially some of the hormones discussed throughout this text) also control the intracellular biochemical reactions by activating or inhibiting one or more of the intracellular control systems.

The DNA-Genetic System Also Controls Cell Reproduction

Cell reproduction is another example of the ubiquitous role that the DNA-genetic system plays in all life processes. The genes and their regulatory mechanisms determine the growth characteristics of the cells and also when or whether these cells will divide to form new cells. In this way, the all-important genetic system controls each stage in the development of the human being, from the single-cell fertilized ovum to the whole functioning body. Thus, if there is any central theme to life, it is the DNA-genetic system.

Life Cycle of the Cell. The life cycle of a cell is the period from cell reproduction to the next cell reproduction. When mammalian cells *are not inhibited and are reproducing as rapidly as they can,* this life cycle may be as little as 10 to 30 hours. It is terminated by a series of distinct physical events called *mitosis* that cause division of the cell into two new daughter cells. The events of mitosis are shown in Figure 3-14 and are described later. The actual stage of mitosis, however, lasts for only about 30 minutes, so more than 95 percent of the life cycle of even rapidly reproducing cells is represented by the interval between mitosis, called *interphase.*

Except in special conditions of rapid cellular reproduction, inhibitory factors almost always slow or stop the uninhibited life cycle of the cell. Therefore, different cells of the body actually have life cycle periods that vary from as little as 10 hours for highly stimulated bone marrow cells to an entire lifetime of the human body for most nerve cells.

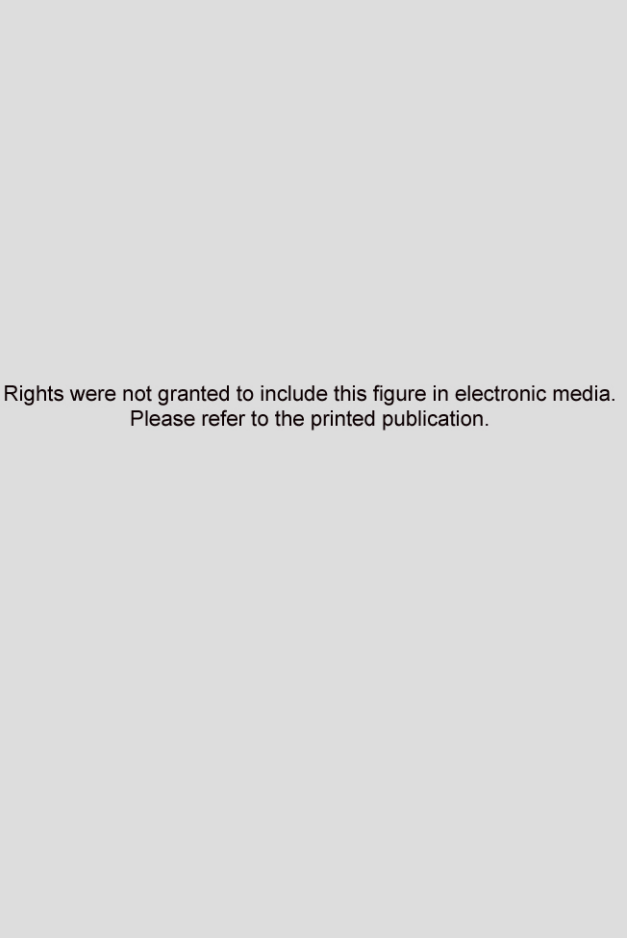


Figure 3-14. Stages of cell reproduction. *A, B,* and *C,* Prophase. *D,* Prometaphase. *E,* Metaphase. *F,* Anaphase. *G* and *H,* Telophase. (From Margaret C. Gladbach, Estate of Mary E. and Dan Todd, Kansas.)

Cell Reproduction Begins with Replication of DNA

As is true of almost all other important events in the cell, reproduction begins in the nucleus itself. The first step is *replication (duplication) of all DNA in the chromosomes.* Only after this has occurred can mitosis take place.

The DNA begins to be duplicated some 5 to 10 hours before mitosis, and this is completed in 4 to 8 hours. The net result is two exact *replicas* of all DNA. These replicas become the DNA in the two new daughter cells that will be formed at mitosis. After replication of the DNA, there is another period of 1 to 2 hours before mitosis begins abruptly. Even during this period, preliminary changes that will lead to the mitotic process are beginning to take place.

Chemical and Physical Events of DNA Replication. DNA is replicated in much the same way that RNA is transcribed in response to DNA, except for a few important differences:

1. Both strands of the DNA in each chromosome are replicated, not simply one of them.

2. Both entire strands of the DNA helix are replicated from end to end, rather than small portions of them, as occurs in the transcription of RNA.
3. The principal enzymes for replicating DNA are a complex of multiple enzymes called *DNA polymerase,* which is comparable to RNA polymerase. It attaches to and moves along the DNA template strand while another enzyme, *DNA ligase,* causes bonding of successive DNA nucleotides to one another, using high-energy phosphate bonds to energize these attachments.
4. Formation of each new DNA strand occurs simultaneously in hundreds of segments along each of the two strands of the helix until the entire strand is replicated. Then the ends of the subunits are joined together by the DNA ligase enzyme.
5. Each newly formed strand of DNA remains attached by loose hydrogen bonding to the original DNA strand that was used as its template. Therefore, two DNA helixes are coiled together.
6. Because the DNA helixes in each chromosome are approximately 6 centimeters in length and have millions of helix turns, it would be impossible for the two newly formed DNA helixes to uncoil from each other were it not for some special mechanism. This is achieved by enzymes that periodically cut each helix along its entire length, rotate each segment enough to cause separation, and then resplice the helix. Thus, the two new helixes become uncoiled.

DNA Repair, DNA "Proofreading," and "Mutation." During the hour or so between DNA replication and the beginning of mitosis, there is a period of active repair and "proofreading" of the DNA strands. That is, wherever inappropriate DNA nucleotides have been matched up with the nucleotides of the original template strand, special enzymes cut out the defective areas and replace these with appropriate complementary nucleotides. This is achieved by the same DNA polymerases and DNA ligases that are used in replication. This repair process is referred to as *DNA proofreading.*

Because of repair and proofreading, the transcription process rarely makes a mistake. But when a mistake is made, this is called a *mutation.* The mutation causes formation of some abnormal protein in the cell rather than a needed protein, often leading to abnormal cellular function and sometimes even cell death. Yet given that there are 30,000 or more genes in the human genome and that the period from one human generation to another is about 30 years, one would expect as many as 10 or many more mutations in the passage of the genome from parent to child. As a further protection, however, each human genome is represented by two separate sets of chromosomes with almost identical genes. Therefore, one functional gene of each pair is almost always available to the child despite mutations.

Chromosomes and Their Replication

The DNA helixes of the nucleus are packaged in chromosomes. The human cell contains 46 chromosomes arranged in 23 pairs. Most of the genes in the two chromosomes of each pair are identical or almost identical to each other, so it is usually stated that the different genes also exist in pairs, although occasionally this is not the case.

In addition to DNA in the chromosome, there is a large amount of protein in the chromosome, composed mainly of many small molecules of electropositively charged *histones.* The histones are organized into vast numbers of small, bobbin-like cores. Small segments of each DNA helix are coiled sequentially around one core after another.

The histone cores play an important role in the regulation of DNA activity because as long as the DNA is packaged tightly, it cannot function as a template for either the formation of RNA or the replication of new DNA. Further, some of the regulatory proteins have been shown to *decondense* the histone packaging of the DNA and to allow small segments at a time to form RNA.

Several nonhistone proteins are also major components of chromosomes, functioning both as chromosomal structural proteins and, in connection with the genetic regulatory machinery, as activators, inhibitors, and enzymes.

Replication of the chromosomes in their entirety occurs during the next few minutes after replication of the DNA helixes has been completed; the new DNA helixes collect new protein molecules as needed. The two newly formed chromosomes remain attached to each other (until time for mitosis) at a point called the *centromere* located near their center. These duplicated but still attached chromosomes are called *chromatids.*

Cell Mitosis

The actual process by which the cell splits into two new cells is called *mitosis.* Once each chromosome has been replicated to form the two chromatids, in many cells, mitosis follows automatically within 1 or 2 hours.

Mitotic Apparatus: Function of the Centrioles. One of the first events of mitosis takes place in the cytoplasm, occurring during the latter part of interphase in or around the small structures called *centrioles.* As shown in Figure 3-14, two pairs of centrioles lie close to each other near one pole of the nucleus. These centrioles, like the DNA and chromosomes, are also replicated during interphase, usually shortly before replication of the DNA. Each centriole is a small cylindrical body about 0.4 micrometer long and about 0.15 micrometer in diameter, consisting mainly of nine parallel tubular structures arranged in the form of a cylinder. The two centrioles of each pair lie at right angles to each other. Each pair of centrioles, along with attached *pericentriolar material,* is called a *centrosome.*

Shortly before mitosis is to take place, the two pairs of centrioles begin to move apart from each other. This is caused by polymerization of protein microtubules growing between the respective centriole pairs and

actually pushing them apart. At the same time, other microtubules grow radially away from each of the centriole pairs, forming a spiny star, called the *aster,* in each end of the cell. Some of the spines of the aster penetrate the nuclear membrane and help separate the two sets of chromatids during mitosis. The complex of microtubules extending between the two new centriole pairs is called the *spindle,* and the entire set of microtubules plus the two pairs of centrioles is called the *mitotic apparatus.*

Prophase. The first stage of mitosis, called *prophase,* is shown in Figure 3-14*A, B,* and *C.* While the spindle is forming, the chromosomes of the nucleus (which in interphase consist of loosely coiled strands) become condensed into well-defined chromosomes.

Prometaphase. During this stage (see Figure 3-14*D*), the growing microtubular spines of the aster fragment the nuclear envelope. At the same time, multiple microtubules from the aster attach to the chromatids at the centromeres, where the paired chromatids are still bound to each other; the tubules then pull one chromatid of each pair toward one cellular pole and its partner toward the opposite pole.

Metaphase. During metaphase (see Figure 3-14*E*), the two asters of the mitotic apparatus are pushed farther apart. This is believed to occur because the microtubular spines from the two asters, where they interdigitate with each other to form the mitotic spindle, actually push each other away. There is reason to believe that minute contractile protein molecules called "*molecular motors,*" perhaps composed of the muscle protein *actin,* extend between the respective spines and, using a stepping action as in muscle, actively slide the spines in a reverse direction along each other. Simultaneously, the chromatids are pulled tightly by their attached microtubules to the very center of the cell, lining up to form the *equatorial plate* of the mitotic spindle.

Anaphase. During this phase (see Figure 3-14*F*), the two chromatids of each chromosome are pulled apart at the centromere. All 46 pairs of chromatids are separated, forming two separate sets of 46 *daughter chromosomes.* One of these sets is pulled toward one mitotic aster and the other toward the other aster as the two respective poles of the dividing cell are pushed still farther apart.

Telophase. In telophase (see Figure 3-14*G* and *H*), the two sets of daughter chromosomes are pushed completely apart. Then the mitotic apparatus dissolutes, and a new nuclear membrane develops around each set of chromosomes. This membrane is formed from portions of the endoplasmic reticulum that are already present in the cytoplasm. Shortly thereafter, the cell pinches in two, midway between the two nuclei. This is caused by formation of a contractile ring of *microfilaments* composed of *actin* and probably *myosin* (the two contractile proteins of muscle) at the juncture of the newly developing cells that pinches them off from each other.

Control of Cell Growth and Cell Reproduction

We know that certain cells grow and reproduce all the time, such as the blood-forming cells of the bone marrow, the germinal layers of the skin, and the epithelium of the gut. Many other cells, however, such as smooth muscle cells, may not reproduce for many years. A few cells, such as the neurons and most striated muscle cells, do not reproduce during the entire life of a person, except during the original period of fetal life.

In certain tissues, an insufficiency of some types of cells causes these to grow and reproduce rapidly until appropriate numbers of them are again available. For instance, in some young animals, seven eighths of the liver can be removed surgically, and the cells of the remaining one eighth will grow and divide until the liver mass returns to almost normal. The same occurs for many glandular cells and most cells of the bone marrow, subcutaneous tissue, intestinal epithelium, and almost any other tissue except highly differentiated cells such as nerve and muscle cells.

We know little about the mechanisms that maintain proper numbers of the different types of cells in the body. However, experiments have shown at least three ways in which growth can be controlled. First, growth often is controlled by *growth factors* that come from other parts of the body. Some of these circulate in the blood, but others originate in adjacent tissues. For instance, the epithelial cells of some glands, such as the pancreas, fail to grow without a growth factor from the sublying connective tissue of the gland. Second, most normal cells stop growing when they have run out of space for growth. This occurs when cells are grown in tissue culture; the cells grow until they contact a solid object, and then growth stops. Third, cells grown in tissue culture often stop growing when minute amounts of their own secretions are allowed to collect in the culture medium. This, too, could provide a means for negative feedback control of growth.

Regulation of Cell Size. Cell size is determined almost entirely by the amount of functioning DNA in the nucleus. If replication of the DNA does not occur, the cell grows to a certain size and thereafter remains at that size. Conversely, it is possible, by use of the chemical *colchicine,* to prevent formation of the mitotic spindle and therefore to prevent mitosis, even though replication of the DNA continues. In this event, the nucleus contains far greater quantities of DNA than it normally does, and the cell grows proportionately larger. It is assumed that this results simply from increased production of RNA and cell proteins, which in turn cause the cell to grow larger.

Cell Differentiation

A special characteristic of cell growth and cell division is *cell differentiation,* which refers to changes in physical and functional properties of cells as they proliferate in the embryo to form the different bodily structures and organs.

The description of an especially interesting experiment that helps explain these processes follows.

When the nucleus from an intestinal mucosal cell of a frog is surgically implanted into a frog ovum from which the original ovum nucleus was removed, the result is often the formation of a normal frog. This demonstrates that even the intestinal mucosal cell, which is a well-differentiated cell, carries all the necessary genetic information for development of all structures required in the frog's body.

Therefore, it has become clear that differentiation results not from loss of genes but from selective repression of different gene promoters. In fact, electron micrographs suggest that some segments of DNA helixes wound around histone cores become so condensed that they no longer uncoil to form RNA molecules. One explanation for this is as follows: It has been supposed that the cellular genome begins at a certain stage of cell differentiation to produce a regulatory *protein* that forever after represses a select group of genes. Therefore, the repressed genes never function again. Regardless of the mechanism, mature human cells produce a maximum of about 8000 to 10,000 proteins rather than the potential 30,000 or more if all genes were active.

Embryological experiments show that certain cells in an embryo control differentiation of adjacent cells. For instance, the *primordial chorda-mesoderm* is called the *primary organizer* of the embryo because it forms a focus around which the rest of the embryo develops. It differentiates into a *mesodermal axis* that contains segmentally arranged *somites* and, as a result of *inductions* in the surrounding tissues, causes formation of essentially all the organs of the body.

Another instance of induction occurs when the developing eye vesicles come in contact with the ectoderm of the head and cause the ectoderm to thicken into a lens plate that folds inward to form the lens of the eye. Therefore, a large share of the embryo develops as a result of such inductions, one part of the body affecting another part, and this part affecting still other parts.

Thus, although our understanding of cell differentiation is still hazy, we know many control mechanisms by which differentiation *could* occur.

Apoptosis—Programmed Cell Death

The 100 trillion cells of the body are members of a highly organized community in which the total number of cells is regulated not only by controlling the rate of cell division but also by controlling the rate of cell death. When cells are no longer needed or become a threat to the organism, they undergo a suicidal *programmed cell death,* or *apoptosis.* This process involves a specific proteolytic cascade that causes the cell to shrink and condense, to disassemble its cytoskeleton, and to alter its cell surface so that a neighboring phagocytic cell, such as a macrophage, can attach to the cell membrane and digest the cell.

In contrast to programmed death, cells that die as a result of an acute injury usually swell and burst due to loss of cell membrane integrity, a process called cell *necrosis.* Necrotic cells may spill their contents, causing inflammation and injury to neighboring cells. Apoptosis, however, is an orderly cell death that results in disassembly and phagocytosis of the cell before any leakage of its contents occurs, and neighboring cells usually remain healthy.

Apoptosis is initiated by activation of a family of proteases called *caspases.* These are enzymes that are synthesized and stored in the cell as inactive *procaspases.* The mechanisms of activation of caspases are complex, but once activated, the enzymes cleave and activate other procaspases, triggering a cascade that rapidly breaks down proteins within the cell. The cell thus dismantles itself, and its remains are rapidly digested by neighboring phagocytic cells.

A tremendous amount of apoptosis occurs in tissues that are being remodeled during development. Even in adult humans, billions of cells die each hour in tissues such as the intestine and bone marrow and are replaced by new cells. Programmed cell death, however, is normally balanced with the formation of new cells in healthy adults. Otherwise, the body's tissues would shrink or grow excessively. Recent studies suggest that abnormalities of apoptosis may play a key role in neurodegenerative diseases such as Alzheimer's disease, as well as in cancer and autoimmune disorders. Some drugs that have been used successfully for chemotherapy appear to induce apoptosis in cancer cells.

Cancer

Cancer is caused in all or almost all instances by *mutation* or by some other *abnormal activation* of cellular genes that control cell growth and cell mitosis. The abnormal genes are called *oncogenes.* As many as 100 different oncogenes have been discovered.

Also present in all cells are *antioncogenes,* which suppress the activation of specific oncogenes. Therefore, loss or inactivation of antioncogenes can allow activation of oncogenes that lead to cancer.

Only a minute fraction of the cells that mutate in the body ever lead to cancer. There are several reasons for this. First, most mutated cells have less survival capability than normal cells and simply die. Second, only a few of the mutated cells that do survive become cancerous, because even most mutated cells still have normal feedback controls that prevent excessive growth.

Third, those cells that are potentially cancerous are often destroyed by the body's immune system before they grow into a cancer. This occurs in the following way: Most mutated cells form abnormal proteins within their cell bodies because of their altered genes, and these proteins activate the body's immune system, causing it to form antibodies or sensitized lymphocytes that react against the cancerous cells, destroying them. In support

of this is the fact that in people whose immune systems have been suppressed, such as in those taking immunosuppressant drugs after kidney or heart transplantation, the probability of a cancer's developing is multiplied as much as fivefold.

Fourth, usually several different activated oncogenes are required simultaneously to cause a cancer. For instance, one such gene might promote rapid reproduction of a cell line, but no cancer occurs because there is not a simultaneous mutant gene to form the needed blood vessels.

But what is it that causes the altered genes? Considering that many trillions of new cells are formed each year in humans, a better question might be, why is it that all of us do not develop millions or billions of mutant cancerous cells? The answer is the incredible precision with which DNA chromosomal strands are replicated in each cell before mitosis can take place, and also the proofreading process that cuts and repairs any abnormal DNA strand before the mitotic process is allowed to proceed. Yet despite all these inherited cellular precautions, probably one newly formed cell in every few million still has significant mutant characteristics.

Thus, chance alone is all that is required for mutations to take place, so we can suppose that a large number of cancers are merely the result of an unlucky occurrence.

However, the probability of mutations can be increased manyfold when a person is exposed to certain chemical, physical, or biological factors, including the following:

1. It is well known that *ionizing radiation,* such as x-rays, gamma rays, and particle radiation from radioactive substances, and even ultraviolet light can predispose individuals to cancer. Ions formed in tissue cells under the influence of such radiation are highly reactive and can rupture DNA strands, thus causing many mutations.
2. *Chemical substances* of certain types also have a high propensity for causing mutations. It was discovered long ago that various aniline dye derivatives are likely to cause cancer, so workers in chemical plants producing such substances, if unprotected, have a special predisposition to cancer. Chemical substances that can cause mutation are called *carcinogens.* The carcinogens that currently cause the greatest number of deaths are those in cigarette smoke. They cause about one quarter of all cancer deaths.
3. *Physical irritants* can also lead to cancer, such as continued abrasion of the linings of the intestinal tract by some types of food. The damage to the tissues leads to rapid mitotic replacement of the cells. The more rapid the mitosis, the greater the chance for mutation.
4. In many families, there is a strong *hereditary tendency* to cancer. This results from the fact that most cancers require not one mutation but two or more mutations before cancer occurs. In those families that are particularly predisposed to cancer, it is presumed that one or more cancerous genes are already mutated in the inherited genome. Therefore, far fewer additional mutations must take place in such family members before a cancer begins to grow.
5. In laboratory animals, certain types of viruses can cause some kinds of cancer, including leukemia. This usually results in one of two ways. In the case of DNA viruses, the DNA strand of the virus can insert itself directly into one of the chromosomes and thereby cause a mutation that leads to cancer. In the case of RNA viruses, some of these carry with them an enzyme called *reverse transcriptase* that causes DNA to be transcribed from the RNA. The transcribed DNA then inserts itself into the animal cell genome, leading to cancer.

Invasive Characteristic of the Cancer Cell. The major differences between the cancer cell and the normal cell are the following: (1) The cancer cell does not respect usual cellular growth limits; the reason for this is that these cells presumably do not require all the same growth factors that are necessary to cause growth of normal cells. (2) Cancer cells are often far less adhesive to one another than are normal cells. Therefore, they tend to wander through the tissues, enter the blood stream, and be transported all through the body, where they form nidi for numerous new cancerous growths. (3) Some cancers also produce *angiogenic factors* that cause many new blood vessels to grow into the cancer, thus supplying the nutrients required for cancer growth.

Why Do Cancer Cells Kill?

The answer to this question is usually simple. Cancer tissue competes with normal tissues for nutrients. Because cancer cells continue to proliferate indefinitely, their number multiplying day by day, cancer cells soon demand essentially all the nutrition available to the body or to an essential part of the body. As a result, normal tissues gradually suffer nutritive death.

Bibliography

Alberts B, Johnson A, Lewis J, et al: *Molecular Biology of the Cell*, ed 5, New York, 2008, Garland Science.

Aranda A, Pascal A: Nuclear hormone receptors and gene expression, *Physiol Rev* 81:1269, 2001.

Brodersen P, Voinnet O: Revisiting the principles of microRNA target recognition and mode of action, *Nat Rev Mol Cell Biol* 10:141, 2009.

Cairns BR: The logic of chromatin architecture and remodelling at promoters, *Nature* 461:193, 2009.

Carthew RW, Sontheimer EJ: Origins and mechanisms of miRNAs and siRNAs, *Cell* 136:642, 2009.

Castanotto D, Rossi JJ: The promises and pitfalls of RNA-interference-based therapeutics, *Nature* 457:426, 2009.

Cedar H, Bergman Y: Linking DNA methylation and histone modification: patterns and paradigms, *Nat Rev Genet* 10:295, 2009.

Croce CM: Causes and consequences of microRNA dysregulation in cancer, *Nat Rev Genet* 10:704, 2009.

Frazer KA, Murray SS, Schork NJ, et al: Human genetic variation and its contribution to complex traits, *Nat Rev Genet* 10:241, 2009.

Fuda NJ, Ardehali MB, Lis JT: Defining mechanisms that regulate RNA polymerase II transcription in vivo, *Nature* 461:186, 2009.

Hahn S: Structure and mechanism of the RNA polymerase II transcription machinery, *Nat Struct Mol Biol* 11:394, 2004.

Hastings PJ, Lupski JR, Rosenberg SM, et al: Mechanisms of change in gene copy number, *Nat Rev Genet* 10:551, 2009.

Hoeijmakers JH: DNA damage, aging, and cancer, *N Engl J Med* 361:1475, 2009.

Hotchkiss RS, Strasser A, McDunn JE, et al: Cell death, *N Engl J Med* 361:1570, 2009.

Jinek M, Doudna JA: A three-dimensional view of the molecular machinery of RNA interference, *Nature* 457:40, 2009.

Jockusch BM, Hüttelmaier S, Illenberger S: From the nucleus toward the cell periphery: a guided tour for mRNAs, *News Physiol Sci* 18:7, 2003.

Kim VN, Han J, Siomi MC: Biogenesis of small RNAs in animals, *Nat Rev Mol Cell Biol* 10:126, 2009.

Misteli T, Soutoglou E: The emerging role of nuclear architecture in DNA repair and genome maintenance, *Nat Rev Mol Cell Biol* 10:243, 2009.

Moazed D: Small RNAs in transcriptional gene silencing and genome defence, *Nature* 457:413, 2009.

Siller KH, Doe CQ: Spindle orientation during asymmetric cell division, *Nat Cell Biol* 11:365, 2009.

Sims RJ 3rd, Reinberg D: Is there a code embedded in proteins that is based on post-translational modifications? *Nat Rev Mol Cell Biol* 9:815, 2008.

Stappenbeck TS, Miyoshi H: The role of stromal stem cells in tissue regeneration and wound repair. *Science* 324:1666, 2009.

Sutherland H, Bickmore WA: Transcription factories: gene expression in unions?, *Nat Rev Genet* 10:457, 2009.

Transport of Substances Through Cell Membranes

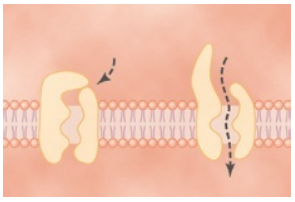


Figure 4-1 gives the approximate concentrations of important electrolytes and other substances in the *extracellular fluid* and *intracellular fluid.* Note that the extracellular fluid contains a large amount of *sodium* but only a small amount of *potassium.* Exactly the opposite is true of the intracellular fluid. Also, the extracellular fluid contains a large amount of *chloride* ions, whereas the intracellular fluid contains very little. But the concentrations of *phosphates* and *proteins* in the intracellular fluid are considerably greater than those in the extracellular fluid. These differences are extremely important to the life of the cell. The purpose of this chapter is to explain how the differences are brought about by the transport mechanisms of the cell membranes.

The Lipid Barrier of the Cell Membrane, and Cell Membrane Transport Proteins

The structure of the membrane covering the outside of every cell of the body is discussed in Chapter 2 and illustrated in Figures 2-3 and 4-2. This membrane consists almost entirely of a *lipid bilayer,* but it also contains large numbers of protein molecules in the lipid, many of which penetrate all the way through the membrane, as shown in Figure 4-2.

The lipid bilayer is not miscible with either the extracellular fluid or the intracellular fluid. Therefore, it constitutes a barrier against movement of water molecules and water-soluble substances between the extracellular and intracellular fluid compartments. However, as demonstrated in Figure 4-2 by the leftmost arrow, a few substances can penetrate this lipid bilayer, diffusing directly through the lipid substance itself; this is true mainly of lipid-soluble substances, as described later.

The protein molecules in the membrane have entirely different properties for transporting substances. Their molecular structures interrupt the continuity of the lipid bilayer, constituting an alternative pathway through the cell membrane. Most of these penetrating proteins, therefore, can function as *transport proteins.* Different proteins function differently. Some have watery spaces all the way through the molecule and allow free movement of water, as well as selected ions or molecules; these are called *channel proteins.* Others, called *carrier proteins,* bind with molecules or ions that are to be transported; conformational changes in the protein molecules then move the substances through the interstices of the protein to the other side of the membrane. Both the channel proteins and the carrier proteins are usually highly selective for the types of molecules or ions that are allowed to cross the membrane.

"Diffusion" Versus "Active Transport." Transport through the cell membrane, either directly through the lipid bilayer or through the proteins, occurs by one of two basic processes: *diffusion* or *active transport.*

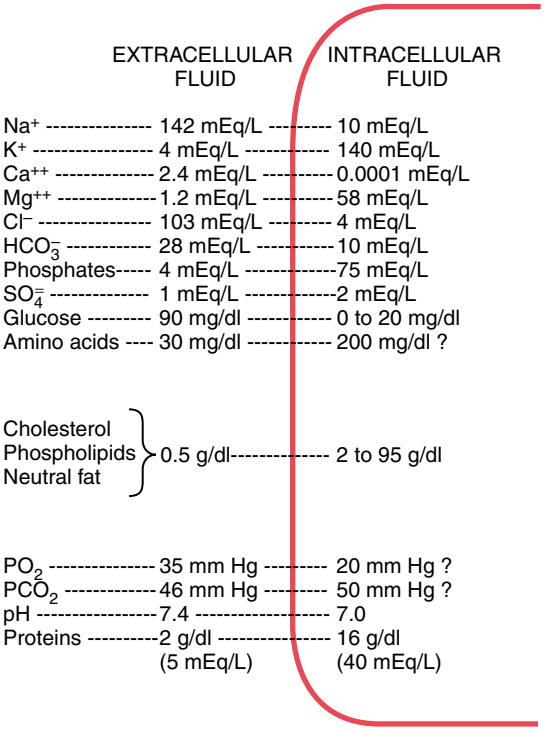


Figure 4-1 Chemical compositions of extracellular and intracellular fluids.

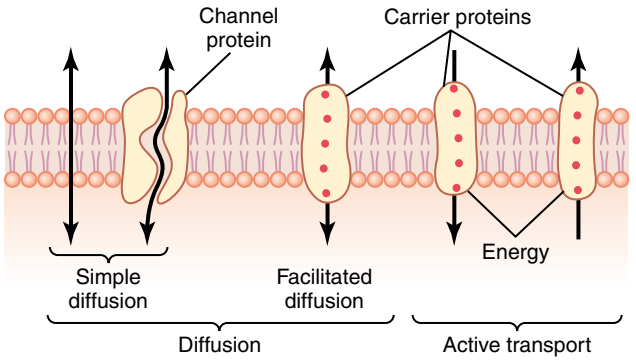


Figure 4-2 Transport pathways through the cell membrane, and the basic mechanisms of transport.

Although there are many variations of these basic mechanisms, diffusion means random molecular movement of substances molecule by molecule, either through intermolecular spaces in the membrane or in combination with a carrier protein. The energy that causes diffusion is the energy of the normal kinetic motion of matter.

By contrast, active transport means movement of ions or other substances across the membrane in combination with a carrier protein in such a way that the carrier protein causes the substance to move against an energy gradient, such as from a low-concentration state to a high-concentration state. This movement requires an additional source of energy besides kinetic energy. Following is a more detailed explanation of the basic physics and physical chemistry of these two processes.

Diffusion

All molecules and ions in the body fluids, including water molecules and dissolved substances, are in constant motion, each particle moving its own separate way. Motion of these particles is what physicists call "heat"—the greater the motion, the higher the temperature—and the motion never ceases under any condition except at absolute zero temperature. When a moving molecule, A, approaches a stationary molecule, B, the electrostatic and other nuclear forces of molecule A repel molecule B, transferring some of the energy of motion of molecule A to molecule B. Consequently, molecule B gains kinetic energy of motion, while molecule A slows down, losing some of its kinetic energy. Thus, as shown in Figure 4-3, a single molecule

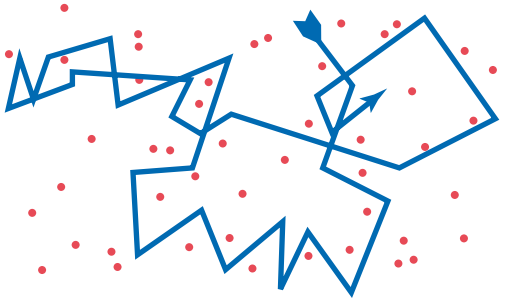


Figure 4-3 Diffusion of a fluid molecule during a thousandth of a second.

in a solution bounces among the other molecules first in one direction, then another, then another, and so forth, randomly bouncing thousands of times each second. This continual movement of molecules among one another in liquids or in gases is called *diffusion.*

Ions diffuse in the same manner as whole molecules, and even suspended colloid particles diffuse in a similar manner, except that the colloids diffuse far less rapidly than molecular substances because of their large size.

Diffusion Through the Cell Membrane

Diffusion through the cell membrane is divided into two subtypes called *simple diffusion* and *facilitated diffusion.* Simple diffusion means that kinetic movement of molecules or ions occurs through a membrane opening or through intermolecular spaces without any interaction with carrier proteins in the membrane. The rate of diffusion is determined by the amount of substance available, the velocity of kinetic motion, and the number and sizes of openings in the membrane through which the molecules or ions can move.

Facilitated diffusion requires interaction of a carrier protein. The carrier protein aids passage of the molecules or ions through the membrane by binding chemically with them and shuttling them through the membrane in this form.

Simple diffusion can occur through the cell membrane by two pathways: (1) through the interstices of the lipid bilayer if the diffusing substance is lipid soluble and (2) through watery channels that penetrate all the way through some of the large transport proteins, as shown to the left in Figure 4-2.

Diffusion of Lipid-Soluble Substances Through the Lipid Bilayer. One of the most important factors that determines how rapidly a substance diffuses through the lipid bilayer is the *lipid solubility* of the substance. For instance, the lipid solubilities of oxygen, nitrogen, carbon dioxide, and alcohols are high, so all these can dissolve directly in the lipid bilayer and diffuse through the cell membrane in the same manner that diffusion of water solutes occurs in a watery solution. For obvious reasons, the rate of diffusion of each of these substances through the membrane is directly proportional to its lipid solubility. Especially large amounts of oxygen can be transported in this way; therefore, oxygen can be delivered to the interior of the cell almost as though the cell membrane did not exist.

Diffusion of Water and Other Lipid-Insoluble Molecules Through Protein Channels. Even though water is highly insoluble in the membrane lipids, it readily passes through channels in protein molecules that penetrate all the way through the membrane. The rapidity with which water molecules can move through most cell membranes is astounding. As an example, the total amount of water that diffuses in each direction through the red cell membrane during each second is about 100 times as great as the volume of the red cell itself.

Other lipid-insoluble molecules can pass through the protein pore channels in the same way as water molecules if they are water soluble and small enough. However, as they become larger, their penetration falls off rapidly. For instance, the diameter of the urea molecule is only 20 percent greater than that of water, yet its penetration through the cell membrane pores is about 1000 times less than that of water. Even so, given the astonishing rate of water penetration, this amount of urea penetration still allows rapid transport of urea through the membrane within minutes.

Diffusion Through Protein Pores and Channels—Selective Permeability and "Gating" of Channels

Computerized three-dimensional reconstructions of protein pores and channels have demonstrated tubular pathways all the way from the extracellular to the intracellular fluid. Therefore, substances can move by simple diffusion directly along these pores and channels from one side of the membrane to the other.

Pores are composed of integral cell membrane proteins that form open tubes through the membrane and are always open. However, the diameter of a pore and its electrical charges provide selectivity that permits only certain molecules to pass through. For example, protein pores, called *aquaporins* or *water channels,* permit rapid passage of water through cell membranes but exclude other molecules. At least 13 different types of aquaporins have been found in various cells of the human body. Aquaporins have a narrow pore that permits water molecules to diffuse through the membrane in single file. The pore is too narrow to permit passage of any hydrated ions. As discussed in Chapters 29 and 75, the density of some aquaporins (e.g., aquaporin-2) in cell membranes is not static but is altered in different physiological conditions.

The protein channels are distinguished by two important characteristics: (1) They are often *selectively permeable* to certain substances, and (2) many of the channels can be opened or closed by *gates* that are regulated by electrical signals *(voltage-gated channels)* or chemicals that bind to the channel proteins *(ligand-gated channels).*

Selective Permeability of Protein Channels. Many of the protein channels are highly selective for transport of one or more specific ions or molecules. This results from the characteristics of the channel itself, such as its diameter, its shape, and the nature of the electrical charges and chemical bonds along its inside surfaces.

Potassium channels permit passage of potassium ions across the cell membrane about 1000 times more readily than they permit passage of sodium ions. This high degree of selectivity, however, cannot be explained entirely by molecular diameters of the ions since potassium ions are slightly larger than sodium ions. What is the mechanism for this remarkable ion selectivity? This question was partially answered when the structure of a *bacterial potassium channel* was determined by x-ray crystallography. Potassium channels were found to have a *tetrameric*

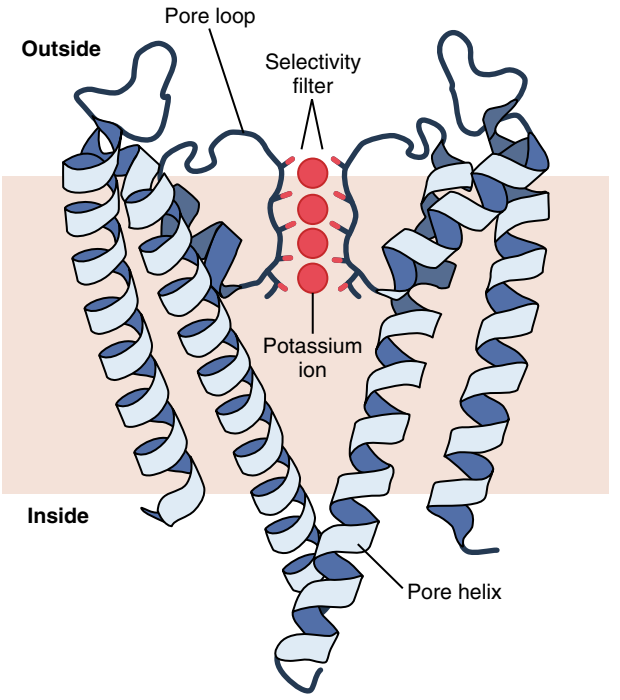


Figure 4-4 The structure of a potassium channel. The channel is composed of four subunits (only two are shown), each with two transmembrane helices. A narrow selectivity filter is formed from the pore loops and carbonyl oxygens line the walls of the selectivity filter, forming sites for transiently binding dehydrated potassium ions. The interaction of the potassium ions with carbonyl oxygens causes the potassium ions to shed their bound water molecules, permitting the dehydrated potassium ions to pass through the pore.

structure consisting of four identical protein subunits surrounding a central pore (Figure 4-4). At the top of the channel pore are *pore loops* that form a narrow *selectivity filter*. Lining the selectivity filter are *carbonyl oxygens.* When hydrated potassium ions enter the selectivity filter, they interact with the carbonyl oxygens and shed most of their bound water molecules, permitting the dehydrated potassium ions to pass through the channel. The carbonyl oxygens are too far apart, however, to enable them to interact closely with the smaller sodium ions, which are therefore effectively excluded by the selectivity filter from passing through the pore.

Different selectivity filters for the various ion channels are believed to determine, in large part, the specificity of the channel for cations or anions or for particular ions, such as Na^+, K^+, and Ca^{++}, that gain access to the channel.

One of the most important of the protein channels, the *sodium channel,* is only 0.3 by 0.5 nanometer in diameter, but more important, the inner surfaces of this channel are lined with amino acids that are *strongly negatively charged,* as shown by the negative signs inside the channel proteins in the top panel of Figure 4-5. These strong negative charges can pull small *dehydrated* sodium ions into these channels, actually pulling the sodium ions away from their hydrating water molecules. Once in the channel, the sodium ions diffuse in either direction according to the usual laws of diffusion. Thus, the sodium channel is specifically selective for passage of sodium ions.

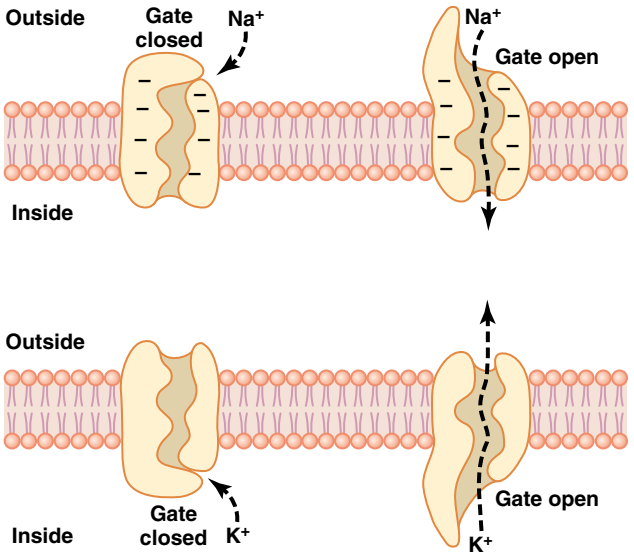


Figure 4-5 Transport of sodium and potassium ions through protein channels. Also shown are conformational changes in the protein molecules to open or close "gates" guarding the channels.

Gating of Protein Channels. Gating of protein channels provides a means of controlling ion permeability of the channels. This is shown in both panels of Figure 4-5 for selective gating of sodium and potassium ions. It is believed that some of the gates are actual gatelike extensions of the transport protein molecule, which can close the opening of the channel or can be lifted away from the opening by a conformational change in the shape of the protein molecule itself.

The opening and closing of gates are controlled in two principal ways:

1. *Voltage gating.* In this instance, the molecular conformation of the gate or of its chemical bonds responds to the electrical potential across the cell membrane. For instance, in the top panel of Figure 4-5, when there is a strong negative charge on the inside of the cell membrane, this presumably could cause the outside sodium gates to remain tightly closed; conversely, when the inside of the membrane loses its negative charge, these gates would open suddenly and allow tremendous quantities of sodium to pass inward through the sodium pores. This is the basic mechanism for eliciting action potentials in nerves that are responsible for nerve signals. In the bottom panel of Figure 4-5, the potassium gates are on the intracellular ends of the potassium channels, and they open when the inside of the cell membrane becomes positively charged. The opening of these gates is partly responsible for terminating the action potential, as is discussed more fully in Chapter 5.
2. *Chemical (ligand) gating.* Some protein channel gates are opened by the binding of a chemical substance (a ligand) with the protein; this causes a conformational or chemical bonding change in the protein molecule that opens or closes the gate. This is called *chemical gating* or *ligand gating.* One of the most important instances of chemical gating is the effect of acetylcholine on the so-called *acetylcholine channel.* Acetylcholine opens the gate of this channel, providing a negatively charged pore about 0.65 nanometer in diameter that allows uncharged molecules or positive ions smaller than this diameter to pass through. This gate is exceedingly important for the transmission of nerve signals from one nerve cell to another (see Chapter 45) and from nerve cells to muscle cells to cause muscle contraction (see Chapter 7).

Open-State Versus Closed-State of Gated Channels. Figure 4-6*A* shows an especially interesting characteristic of most voltage-gated channels. This figure shows two recordings of electrical current flowing through a single sodium channel when there was an approximate 25-millivolt potential gradient across the membrane. Note that the channel conducts current either "all or none." That is, the gate of the channel snaps open and then snaps closed, each open state lasting for only a fraction of a millisecond up to several milliseconds. This demonstrates the rapidity with which changes can occur during the opening and closing of the protein molecular gates. At one voltage potential, the channel may remain closed all the time or almost all the time, whereas at another voltage level, it may remain open either all or most of the time. At in-between voltages, as shown in the figure, the gates tend to snap open and closed intermittently, giving an average current flow somewhere between the minimum and the maximum.

Patch-Clamp Method for Recording Ion Current Flow Through Single Channels. One might wonder how it is technically possible to record ion current flow through single protein channels as shown in Figure 4-6*A*. This has been achieved by using the "patch-clamp" method illustrated in Figure 4-6*B*. Very simply, a micropipette, having a tip diameter of only 1 or 2 micrometers, is abutted against the outside of a cell membrane. Then suction is applied inside the pipette to pull the membrane against the tip of the pipette. This creates a seal where the edges of the pipette touch the cell membrane. The result is a minute membrane "patch" at the tip of the pipette through which electrical current flow can be recorded.

Alternatively, as shown to the right in Figure 4-6*B*, the small cell membrane patch at the end of the pipette can be torn away from the cell. The pipette with its sealed patch is then inserted into a free solution. This allows the concentrations of ions both inside the micropipette and in the outside solution to be altered as desired. Also, the voltage between the two sides of the membrane can be set at will—that is, "clamped" to a given voltage.

It has been possible to make such patches small enough so that only a single channel protein is found in the membrane patch being studied. By varying the concentrations of different ions, as well as the voltage across the membrane, one can determine the transport characteristics of the single channel and also its gating properties.

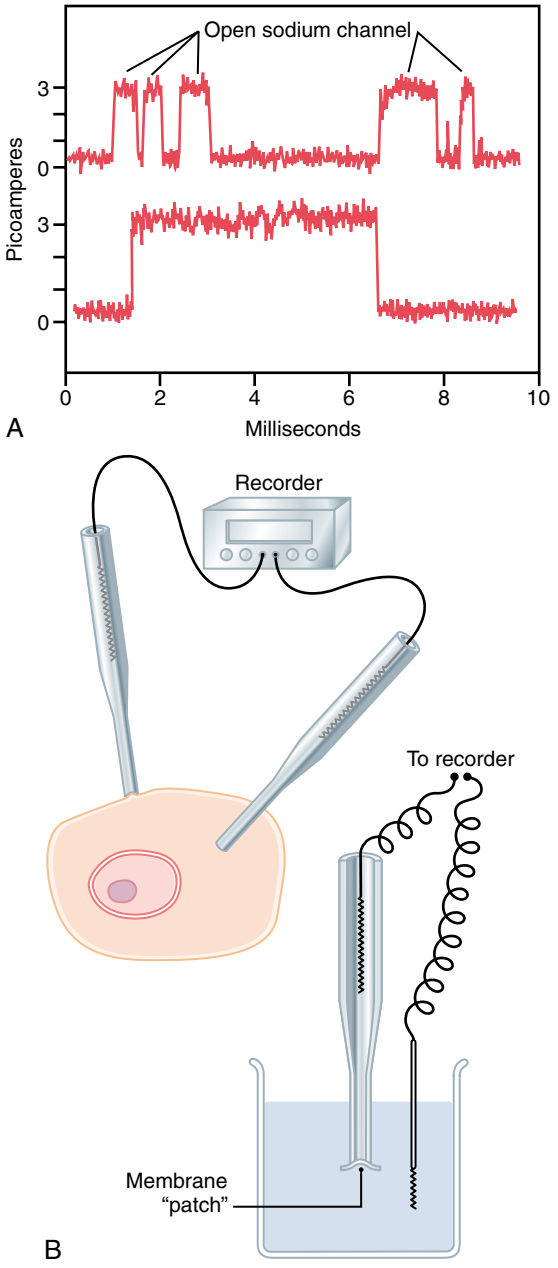


Figure 4-6 *A,* Record of current flow through a single voltage-gated sodium channel, demonstrating the "all or none" principle for opening and closing of the channel. *B,* The "patch-clamp" method for recording current flow through a single protein channel. To the left, recording is performed from a "patch" of a living cell membrane. To the right, recording is from a membrane patch that has been torn away from the cell.

Facilitated Diffusion

Facilitated diffusion is also called *carrier-mediated diffusion* because a substance transported in this manner diffuses through the membrane using a specific carrier protein to help. That is, the carrier *facilitates* diffusion of the substance to the other side.

Facilitated diffusion differs from simple diffusion in the following important way: Although the rate of simple diffusion through an open channel increases proportionately with the concentration of the diffusing substance,

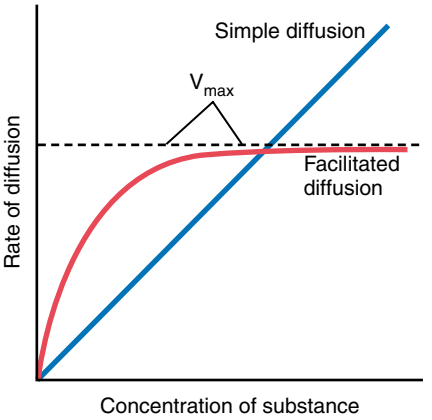


Figure 4-7 Effect of concentration of a substance on rate of diffusion through a membrane by simple diffusion and facilitated diffusion. This shows that facilitated diffusion approaches a maximum rate called the V_{max}.

in facilitated diffusion the rate of diffusion approaches a maximum, called $V_{max,}$ as the concentration of the diffusing substance increases. This difference between simple diffusion and facilitated diffusion is demonstrated in Figure 4-7. The figure shows that as the concentration of the diffusing substance increases, the rate of simple diffusion continues to increase proportionately, but in the case of facilitated diffusion, the rate of diffusion cannot rise greater than the V_{max} level.

What is it that limits the rate of facilitated diffusion? A probable answer is the mechanism illustrated in Figure 4-8. This figure shows a carrier protein with a pore large enough to transport a specific molecule partway through. It also shows a binding "receptor" on the inside of the protein carrier. The molecule to be transported enters the pore and becomes bound. Then, in a fraction of a second, a conformational or chemical change occurs in the carrier protein, so the pore now opens to the opposite side of the membrane. Because the binding force of the receptor is weak, the thermal motion of the attached molecule causes

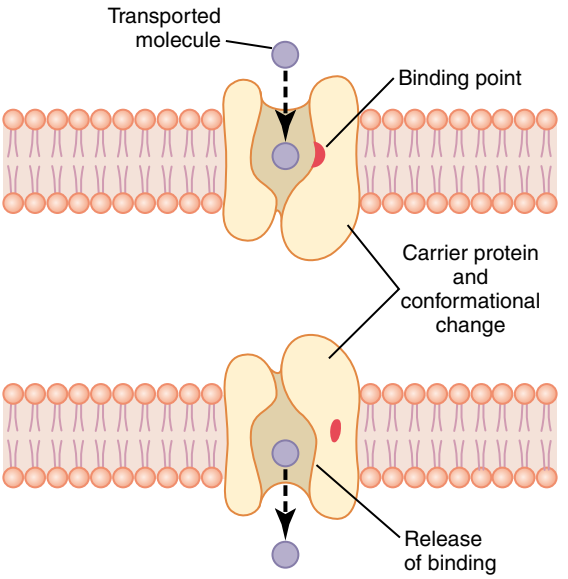


Figure 4-8 Postulated mechanism for facilitated diffusion.

it to break away and to be released on the opposite side of the membrane. The rate at which molecules can be transported by this mechanism can never be greater than the rate at which the carrier protein molecule can undergo change back and forth between its two states. Note specifically, though, that this mechanism allows the transported molecule to move—that is, to "diffuse"—in either direction through the membrane.

Among the most important substances that cross cell membranes by facilitated diffusion are *glucose* and most of the *amino acids*. In the case of glucose, at least five glucose transporter molecules have been discovered in various tissues. Some of these can also transport other monosaccharides that have structures similar to that of glucose, including galactose and fructose. One of these, glucose transporter 4 (GLUT4), is activated by insulin, which can increase the rate of facilitated diffusion of glucose as much as 10-fold to 20-fold in insulin-sensitive tissues. This is the principal mechanism by which insulin controls glucose use in the body, as discussed in Chapter 78.

Factors That Affect Net Rate of Diffusion

By now it is evident that many substances can diffuse through the cell membrane. What is usually important is the *net* rate of diffusion of a substance in the desired direction. This net rate is determined by several factors.

Net Diffusion Rate Is Proportional to the Concentration Difference Across a Membrane. Figure 4-9*A* shows a cell membrane with a substance in high concentration on the outside and low concentration on the inside. The rate at which the substance diffuses *inward* is proportional to the concentration of molecules on the *outside* because this concentration determines how many molecules strike the outside of the membrane each second. Conversely, the rate at which molecules diffuse *outward* is proportional to their concentration *inside* the membrane. Therefore, the rate of net diffusion into the cell is proportional to the concentration on the outside *minus* the concentration on the inside, or:

$$\text{Net diffusion} \propto (C_o - C_i)$$

in which C_o is concentration outside and C_i is concentration inside.

Effect of Membrane Electrical Potential on Diffusion of Ions—The "Nernst Potential." If an electrical potential is applied across the membrane, as shown in Figure 4-9*B*, the electrical charges of the ions cause them to move through the membrane even though no concentration difference exists to cause movement. Thus, in the left panel of Figure 4-9*B*, the concentration of *negative* ions is the same on both sides of the membrane, but a positive charge has been applied to the right side of the membrane and a negative charge to the left, creating an electrical gradient across the membrane. The positive charge attracts the negative ions, whereas the negative charge repels them. Therefore, net diffusion

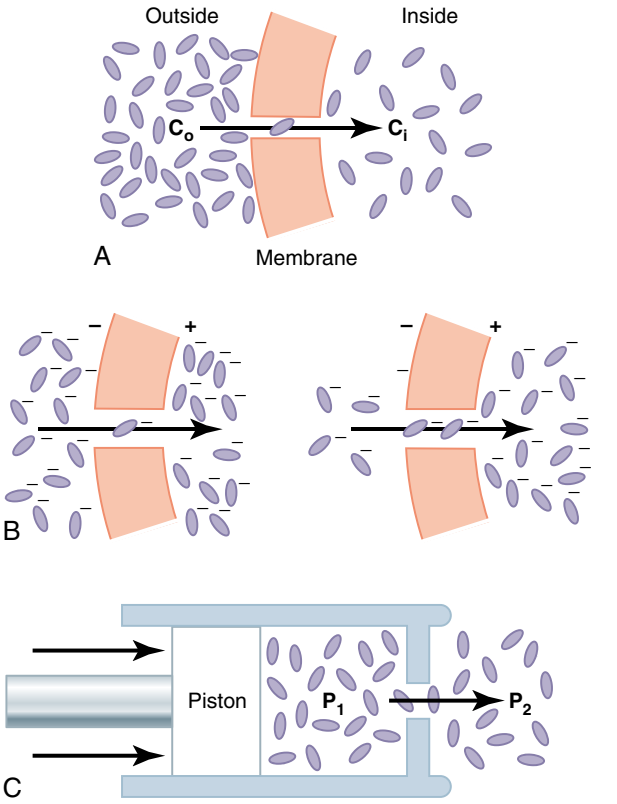


Figure 4-9 Effect of concentration difference (*A*), electrical potential difference affecting negative ions (*B*), and pressure difference (*C*) to cause diffusion of molecules and ions through a cell membrane.

occurs from left to right. After some time, large quantities of negative ions have moved to the right, creating the condition shown in the right panel of Figure 4-9*B*, in which a concentration difference of the ions has developed in the direction opposite to the electrical potential difference. The concentration difference now tends to move the ions to the left, while the electrical difference tends to move them to the right. When the concentration difference rises high enough, the two effects balance each other. At normal body temperature (37°C), the electrical difference that will balance a given concentration difference of *univalent* ions—such as sodium (Na^+) ions—can be determined from the following formula, called the *Nernst equation:*

$$\text{EMF (in millivolts)} = \pm 61 \log \frac{C_1}{C_2}$$

in which EMF is the electromotive force (voltage) between side 1 and side 2 of the membrane, C_1 is the concentration on side 1, and C_2 is the concentration on side 2. This equation is extremely important in understanding the transmission of nerve impulses and is discussed in much greater detail in Chapter 5.

Effect of a Pressure Difference Across the Membrane. At times, considerable pressure difference develops between the two sides of a diffusible membrane.

This occurs, for instance, at the blood capillary membrane in all tissues of the body. The pressure is about 20 mm Hg greater inside the capillary than outside.

Pressure actually means the sum of all the forces of the different molecules striking a unit surface area at a given instant. Therefore, when the pressure is higher on one side of a membrane than on the other, this means that the sum of all the forces of the molecules striking the channels on that side of the membrane is greater than on the other side. In most instances, this is caused by greater numbers of molecules striking the membrane per second on one side than on the other side. The result is that increased amounts of energy are available to cause net movement of molecules from the high-pressure side toward the low-pressure side. This effect is demonstrated in Figure 4-9*C*, which shows a piston developing high pressure on one side of a "pore," thereby causing more molecules to strike the pore on this side and, therefore, more molecules to "diffuse" to the other side.

Osmosis Across Selectively Permeable Membranes—"Net Diffusion" of Water

By far the most abundant substance that diffuses through the cell membrane is water. Enough water ordinarily diffuses in each direction through the red cell membrane per second to equal about *100 times the volume of the cell itself.* Yet normally the amount that diffuses in the two directions is balanced so precisely that zero net movement of water occurs. Therefore, the volume of the cell remains constant. However, under certain conditions, a *concentration difference for water* can develop across a membrane, just as concentration differences for other substances can occur. When this happens, net movement of water does occur across the cell membrane, causing the cell either to swell or shrink, depending on the direction of the water movement. This process of net movement of water caused by a concentration difference of water is called *osmosis.*

To give an example of osmosis, let us assume the conditions shown in Figure 4-10, with pure water on one side of the cell membrane and a solution of sodium chloride on the other side. Water molecules pass through the cell membrane with ease, whereas sodium and chloride ions pass through only with difficulty. Therefore, sodium chloride solution is actually a mixture of permeant water molecules and nonpermeant sodium and chloride ions, and the membrane is said to be *selectively permeable* to water but much less so to sodium and chloride ions. Yet the presence of the sodium and chloride has displaced some of the water molecules on the side of the membrane where these ions are present and, therefore, has reduced the concentration of water molecules to less than that of pure water. As a result, in the example of Figure 4-10, more water molecules strike the channels on the left side, where there is pure water, than on the right side, where the water concentration has been reduced. Thus, net movement of water occurs from left to right—that is, *osmosis* occurs from the pure water into the sodium chloride solution.

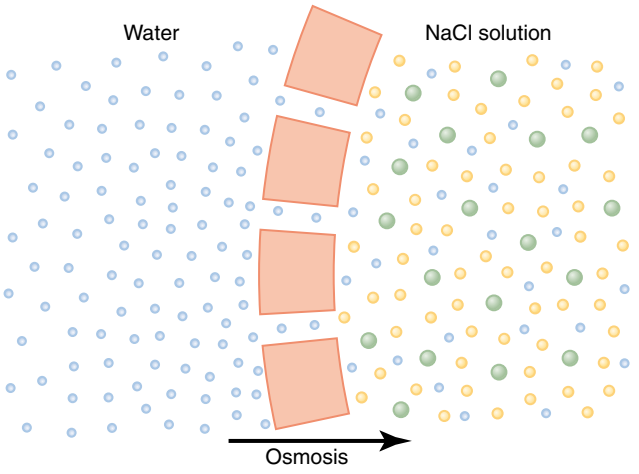


Figure 4-10 Osmosis at a cell membrane when a sodium chloride solution is placed on one side of the membrane and water is placed on the other side.

Osmotic Pressure

If in Figure 4-10 pressure were applied to the sodium chloride solution, osmosis of water into this solution would be slowed, stopped, or even reversed. The exact amount of pressure required to stop osmosis is called the *osmotic pressure* of the sodium chloride solution.

The principle of a pressure difference opposing osmosis is demonstrated in Figure 4-11, which shows a selectively permeable membrane separating two columns of fluid, one containing pure water and the other containing a solution of water and any solute that will not penetrate the membrane. Osmosis of water from chamber B into chamber A causes the levels of the fluid columns to become farther and farther apart, until eventually a pressure difference develops between the two sides of the membrane great enough to oppose the osmotic effect.

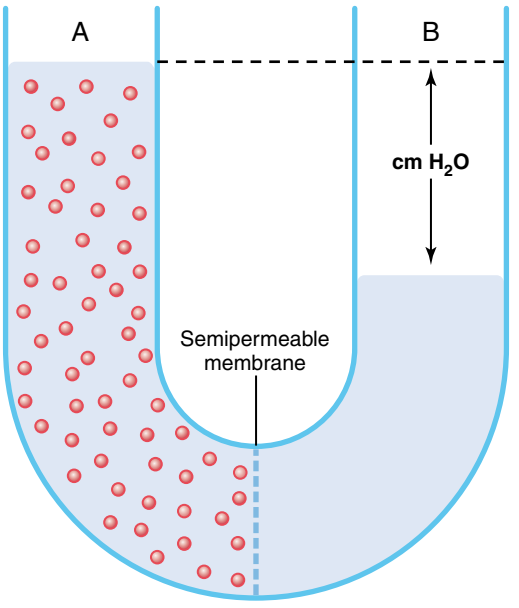


Figure 4-11 Demonstration of osmotic pressure caused by osmosis at a semipermeable membrane.

The pressure difference across the membrane at this point is equal to the osmotic pressure of the solution that contains the nondiffusible solute.

Importance of Number of Osmotic Particles (Molar Concentration) in Determining Osmotic Pressure. The osmotic pressure exerted by particles in a solution, whether they are molecules or ions, is determined by the *number* of particles per unit volume of fluid, *not by the mass* of the particles. The reason for this is that each particle in a solution, regardless of its mass, exerts, on average, the same amount of pressure against the membrane. That is, large particles, which have greater mass (m) than small particles, move at slower velocities (v). The small particles move at higher velocities in such a way that their average kinetic energies (k), determined by the equation

$$k = \frac{mv^2}{2}$$

are the same for each small particle as for each large particle. Consequently, the factor that determines the osmotic pressure of a solution is the concentration of the solution in terms of number of particles (which is the same as its *molar concentration* if it is a nondissociated molecule), not in terms of mass of the solute.

"Osmolality"—The Osmole. To express the concentration of a solution in terms of numbers of particles, the unit called the *osmole* is used in place of grams.

One osmole is 1 gram molecular weight of osmotically active solute. Thus, 180 grams of glucose, which is 1 gram molecular weight of glucose, is equal to 1 osmole of glucose because glucose does not dissociate into ions. If a solute dissociates into two ions, 1 gram molecular weight of the solute will become 2 osmoles because the number of osmotically active particles is now twice as great as is the case for the nondissociated solute. Therefore, when fully dissociated, 1 gram molecular weight of sodium chloride, 58.5 grams, is equal to 2 osmoles.

Thus, a solution that has *1 osmole of solute dissolved in each kilogram of water* is said to have an *osmolality of 1 osmole per kilogram,* and a solution that has 1/1000 osmole dissolved per kilogram has an osmolality of 1 milliosmole per kilogram. The normal osmolality of the extracellular and intracellular fluids is about *300 milliosmoles per kilogram of water.*

Relation of Osmolality to Osmotic Pressure. At normal body temperature, 37°C, a concentration of 1 osmole per liter will cause *19,300 mm Hg* osmotic pressure in the solution. Likewise, *1 milliosmole* per liter concentration is equivalent to *19.3 mm Hg* osmotic pressure. Multiplying this value by the 300 milliosmolar concentration of the body fluids gives a total calculated osmotic pressure of the body fluids of 5790 mm Hg. The measured value for this, however, averages only about 5500 mm Hg. The reason for this difference is that many of the ions in the body fluids, such as sodium and chloride ions, are highly attracted to one another; consequently, they cannot move entirely unrestrained in the fluids and create their full osmotic pressure potential. Therefore, on average, the actual osmotic pressure of the body fluids is about 0.93 times the calculated value.

The Term "Osmolarity." *Osmolarity* is the osmolar concentration expressed as *osmoles per liter of solution* rather than osmoles per kilogram of water. Although, strictly speaking, it is osmoles per kilogram of water (osmolality) that determines osmotic pressure, for dilute solutions such as those in the body, the quantitative differences between osmolarity and osmolality are less than 1 percent. Because it is far more practical to measure osmolarity than osmolality, this is the usual practice in almost all physiological studies.

"Active Transport" of Substances Through Membranes

At times, a large concentration of a substance is required in the intracellular fluid even though the extracellular fluid contains only a small concentration. This is true, for instance, for potassium ions. Conversely, it is important to keep the concentrations of other ions very low inside the cell even though their concentrations in the extracellular fluid are great. This is especially true for sodium ions. Neither of these two effects could occur by simple diffusion because simple diffusion eventually equilibrates concentrations on the two sides of the membrane. Instead, some energy source must cause excess movement of potassium ions to the inside of cells and excess movement of sodium ions to the outside of cells. When a cell membrane moves molecules or ions "uphill" against a concentration gradient (or "uphill" against an electrical or pressure gradient), the process is called *active transport.*

Different substances that are actively transported through at least some cell membranes include sodium ions, potassium ions, calcium ions, iron ions, hydrogen ions, chloride ions, iodide ions, urate ions, several different sugars, and most of the amino acids.

Primary Active Transport and Secondary Active Transport. Active transport is divided into two types according to the source of the energy used to cause the transport: *primary active transport* and *secondary active transport.* In primary active transport, the energy is derived directly from breakdown of adenosine triphosphate (ATP) or of some other high-energy phosphate compound. In secondary active transport, the energy is derived secondarily from energy that has been stored in the form of ionic concentration differences of secondary molecular or ionic substances between the two sides of a cell membrane, created originally by primary active transport. In both instances, transport depends on *carrier proteins* that penetrate through the cell membrane, as is true for facilitated diffusion. However, in active transport, the carrier protein functions differently from the carrier in facilitated diffusion because it is capable of imparting energy to the transported substance to move it against the

electrochemical gradient. Following are some examples of primary active transport and secondary active transport, with more detailed explanations of their principles of function.

Primary Active Transport

Sodium-Potassium Pump

Among the substances that are transported by primary active transport are sodium, potassium, calcium, hydrogen, chloride, and a few other ions.

The active transport mechanism that has been studied in greatest detail is the *sodium-potassium (Na^+-K^+)* pump, a transport process that pumps sodium ions outward through the cell membrane of all cells and at the same time pumps potassium ions from the outside to the inside. This pump is responsible for maintaining the sodium and potassium concentration differences across the cell membrane, as well as for establishing a negative electrical voltage inside the cells. Indeed, Chapter 5 shows that this pump is also the basis of nerve function, transmitting nerve signals throughout the nervous system.

Figure 4-12 shows the basic physical components of the Na^+-K^+ pump. The *carrier protein* is a complex of two separate globular proteins: a larger one called the α subunit, with a molecular weight of about 100,000, and a smaller one called the β subunit, with a molecular weight of about 55,000. Although the function of the smaller protein is not known (except that it might anchor the protein complex in the lipid membrane), the larger protein has three specific features that are important for the functioning of the pump:

1. It has three *receptor sites for binding sodium ions* on the portion of the protein that protrudes to the inside of the cell.
2. It has two *receptor sites for potassium ions* on the outside.
3. The inside portion of this protein near the sodium binding sites has ATPase activity.

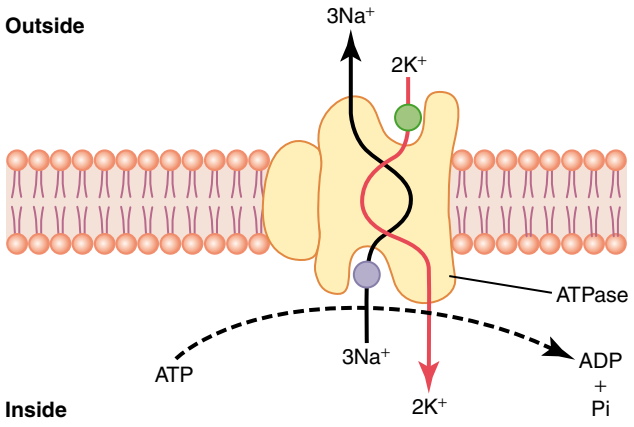


Figure 4-12 Postulated mechanism of the sodium-potassium pump. ADP, adenosine diphosphate; ATP, adenosine triphosphate; Pi, phosphate ion.

When two potassium ions bind on the outside of the carrier protein and three sodium ions bind on the inside, the ATPase function of the protein becomes activated. This then cleaves one molecule of ATP, splitting it to adenosine diphosphate (ADP) and liberating a high-energy phosphate bond of energy. This liberated energy is then believed to cause a chemical and conformational change in the protein carrier molecule, extruding the three sodium ions to the outside and the two potassium ions to the inside.

As with other enzymes, the Na^+-K^+ ATPase pump can run in reverse. If the electrochemical gradients for Na^+ and K^+ are experimentally increased enough so that the energy stored in their gradients is greater than the chemical energy of ATP hydrolysis, these ions will move down their concentration gradients and the Na^+-K^+ pump will synthesize ATP from ADP and phosphate. The phosphorylated form of the Na^+-K^+ pump, therefore, can either donate its phosphate to ADP to produce ATP or use the energy to change its conformation and pump Na^+ out of the cell and K^+ into the cell. The relative concentrations of ATP, ADP, and phosphate, as well as the electrochemical gradients for Na^+ and K^+, determine the direction of the enzyme reaction. For some cells, such as electrically active nerve cells, 60 to 70 percent of the cells' energy requirement may be devoted to pumping Na^+ out of the cell and K^+ into the cell.

The Na^+-K^+ Pump Is Important For Controlling Cell Volume. One of the most important functions of the Na^+-K^+ pump is to control the volume of each cell. Without function of this pump, most cells of the body would swell until they burst. The mechanism for controlling the volume is as follows: Inside the cell are large numbers of proteins and other organic molecules that cannot escape from the cell. Most of these are negatively charged and therefore attract large numbers of potassium, sodium, and other positive ions as well. All these molecules and ions then cause osmosis of water to the interior of the cell. Unless this is checked, the cell will swell indefinitely until it bursts. The normal mechanism for preventing this is the Na^+-K^+ pump. Note again that this device pumps three Na^+ ions to the outside of the cell for every two K^+ ions pumped to the interior. Also, the membrane is far less permeable to sodium ions than to potassium ions, so once the sodium ions are on the outside, they have a strong tendency to stay there. Thus, this represents a net loss of ions out of the cell, which initiates osmosis of water out of the cell as well.

If a cell begins to swell for any reason, this automatically activates the Na^+-K^+ pump, moving still more ions to the exterior and carrying water with them. Therefore, the Na^+-K^+ pump performs a continual surveillance role in maintaining normal cell volume.

Electrogenic Nature of the Na^+-K^+ Pump. The fact that the Na^+-K^+ pump moves three Na^+ ions to the exterior for every two K^+ ions to the interior means that a net of one positive charge is moved from the interior of the cell to the exterior for each cycle of the pump. This creates positivity

outside the cell but leaves a deficit of positive ions inside the cell; that is, it causes negativity on the inside. Therefore, the Na^+-K^+ pump is said to be *electrogenic* because it creates an electrical potential across the cell membrane. As discussed in Chapter 5, this electrical potential is a basic requirement in nerve and muscle fibers for transmitting nerve and muscle signals.

Primary Active Transport of Calcium Ions

Another important primary active transport mechanism is the *calcium pump.* Calcium ions are normally maintained at extremely low concentration in the intracellular cytosol of virtually all cells in the body, at a concentration about 10,000 times less than that in the extracellular fluid. This is achieved mainly by two primary active transport calcium pumps. One is in the cell membrane and pumps calcium to the outside of the cell. The other pumps calcium ions into one or more of the intracellular vesicular organelles of the cell, such as the sarcoplasmic reticulum of muscle cells and the mitochondria in all cells. In each of these instances, the carrier protein penetrates the membrane and functions as an enzyme ATPase, having the same capability to cleave ATP as the ATPase of the sodium carrier protein. The difference is that this protein has a highly specific binding site for calcium instead of for sodium.

Primary Active Transport of Hydrogen Ions

At two places in the body, primary active transport of hydrogen ions is important: (1) in the gastric glands of the stomach and (2) in the late distal tubules and cortical collecting ducts of the kidneys.

In the gastric glands, the deep-lying *parietal cells* have the most potent primary active mechanism for transporting hydrogen ions of any part of the body. This is the basis for secreting hydrochloric acid in the stomach digestive secretions. At the secretory ends of the gastric gland parietal cells, the hydrogen ion concentration is increased as much as a millionfold and then released into the stomach along with chloride ions to form hydrochloric acid.

In the renal tubules are special *intercalated cells* in the late distal tubules and cortical collecting ducts that also transport hydrogen ions by primary active transport. In this case, large amounts of hydrogen ions are secreted from the blood into the urine for the purpose of eliminating excess hydrogen ions from the body fluids. The hydrogen ions can be secreted into the urine against a concentration gradient of about 900-fold.

Energetics of Primary Active Transport

The amount of energy required to transport a substance actively through a membrane is determined by how much the substance is concentrated during transport. Compared with the energy required to concentrate a substance 10-fold, to concentrate it 100-fold requires twice as much energy, and to concentrate it 1000-fold requires three times as much energy. In other words, the energy required is proportional to the *logarithm* of the degree that the substance is concentrated, as expressed by the following formula:

$$\text{Energy (in calories per osmole)} = 1400 \log \frac{C_1}{C_2}$$

Thus, in terms of calories, the amount of energy required to concentrate 1 osmole of a substance 10-fold is about 1400 calories; or to concentrate it 100-fold, 2800 calories. One can see that the energy expenditure for concentrating substances in cells or for removing substances from cells against a concentration gradient can be tremendous. Some cells, such as those lining the renal tubules and many glandular cells, expend as much as 90 percent of their energy for this purpose alone.

Secondary Active Transport—Co-Transport and Counter-Transport

When sodium ions are transported out of cells by primary active transport, a large concentration gradient of sodium ions across the cell membrane usually develops—high concentration outside the cell and low concentration inside. This gradient represents a storehouse of energy because the excess sodium outside the cell membrane is always attempting to diffuse to the interior. Under appropriate conditions, this diffusion energy of sodium can pull other substances along with the sodium through the cell membrane. This phenomenon is called *co-transport;* it is one form of *secondary active transport.*

For sodium to pull another substance along with it, a coupling mechanism is required. This is achieved by means of still another carrier protein in the cell membrane. The carrier in this instance serves as an attachment point for both the sodium ion and the substance to be co-transported. Once they both are attached, the energy gradient of the sodium ion causes both the sodium ion and the other substance to be transported together to the interior of the cell.

In *counter-transport,* sodium ions again attempt to diffuse to the interior of the cell because of their large concentration gradient. However, this time, the substance to be transported is on the inside of the cell and must be transported to the outside. Therefore, the sodium ion binds to the carrier protein where it projects to the exterior surface of the membrane, while the substance to be counter-transported binds to the interior projection of the carrier protein. Once both have bound, a conformational change occurs, and energy released by the sodium ion moving to the interior causes the other substance to move to the exterior.

Co-Transport of Glucose and Amino Acids Along with Sodium Ions

Glucose and many amino acids are transported into most cells against large concentration gradients; the mechanism of this is entirely by co-transport, as shown in Figure 4-13. Note that the transport carrier protein has two binding sites on its exterior side, one for sodium and one

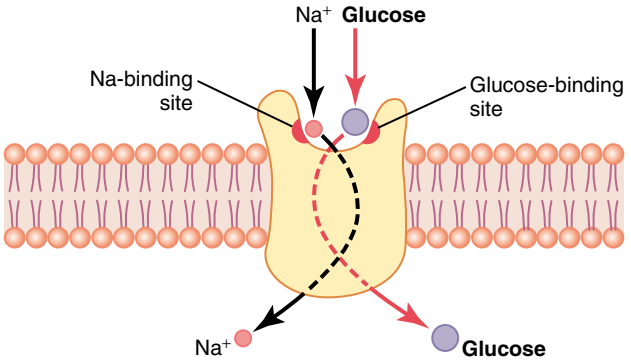


Figure 4-13 Postulated mechanism for sodium co-transport of glucose.

for glucose. Also, the concentration of sodium ions is high on the outside and low inside, which provides energy for the transport. A special property of the transport protein is that a conformational change to allow sodium movement to the interior will not occur until a glucose molecule also attaches. When they both become attached, the conformational change takes place automatically, and the sodium and glucose are transported to the inside of the cell at the same time. Hence, this is a *sodium-glucose co-transport* mechanism. Sodium-glucose co-transporters are especially important mechanisms in transporting glucose across renal and intestinal epithelial cells, as discussed in Chapters 27 and 65.

Sodium co-transport of the amino acids occurs in the same manner as for glucose, except that it uses a different set of transport proteins. Five *amino acid transport proteins* have been identified, each of which is responsible for transporting one subset of amino acids with specific molecular characteristics.

Sodium co-transport of glucose and amino acids occurs especially through the epithelial cells of the intestinal tract and the renal tubules of the kidneys to promote absorption of these substances into the blood, as is discussed in later chapters.

Other important co-transport mechanisms in at least some cells include co-transport of chloride ions, iodine ions, iron ions, and urate ions.

Sodium Counter-Transport of Calcium and Hydrogen Ions

Two especially important counter-transport mechanisms (transport in a direction opposite to the primary ion) are *sodium-calcium counter-transport* and *sodium-hydrogen counter-transport* (Figure 4-14).

Sodium-calcium counter-transport occurs through all or almost all cell membranes, with sodium ions moving to the interior and calcium ions to the exterior, both bound to the same transport protein in a counter-transport mode. This is in addition to primary active transport of calcium that occurs in some cells.

Sodium-hydrogen counter-transport occurs in several tissues. An especially important example is in the *proximal tubules* of the kidneys, where sodium ions move

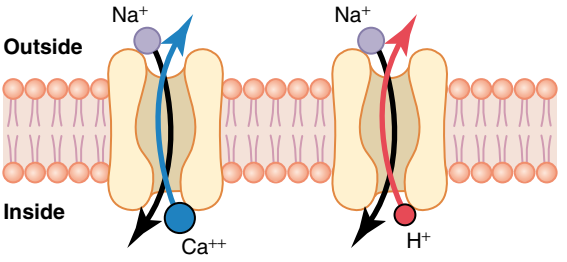


Figure 4-14 Sodium counter-transport of calcium and hydrogen ions.

from the lumen of the tubule to the interior of the tubular cell, while hydrogen ions are counter-transported into the tubule lumen. As a mechanism for concentrating hydrogen ions, counter-transport is not nearly as powerful as the primary active transport of hydrogen ions that occurs in the more distal renal tubules, but it can transport extremely *large numbers of hydrogen ions,* thus making it a key to hydrogen ion control in the body fluids, as discussed in detail in Chapter 30.

Active Transport Through Cellular Sheets

At many places in the body, substances must be transported all the way through a cellular sheet instead of simply through the cell membrane. Transport of this type occurs through the (1) intestinal epithelium, (2) epithelium of the renal tubules, (3) epithelium of all exocrine glands, (4) epithelium of the gallbladder, and (5) membrane of the choroid plexus of the brain and other membranes.

The basic mechanism for transport of a substance through a cellular sheet is (1) *active transport* through the cell membrane *on one side* of the transporting cells in the sheet, and then (2) either *simple diffusion* or *facilitated diffusion* through the membrane *on the opposite side* of the cell.

Figure 4-15 shows a mechanism for transport of sodium ions through the epithelial sheet of the intestines, gallbladder, and renal tubules. This figure shows that the epithelial cells are connected together tightly at the luminal pole by means of junctions called "kisses." The brush border on the luminal surfaces of the cells

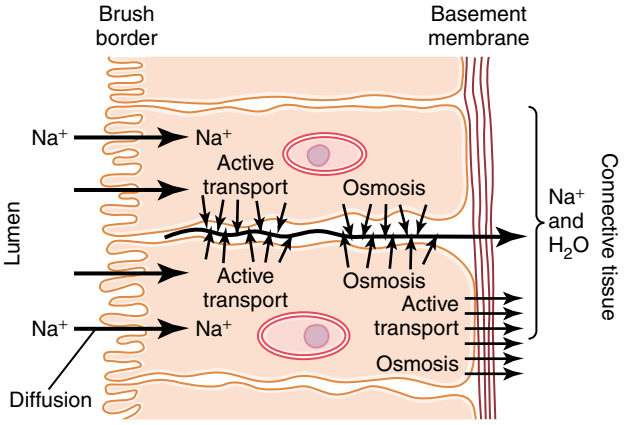


Figure 4-15 Basic mechanism of active transport across a layer of cells.

is permeable to both sodium ions and water. Therefore, sodium and water diffuse readily from the lumen into the interior of the cell. Then, at the basal and lateral membranes of the cells, sodium ions are actively transported into the extracellular fluid of the surrounding connective tissue and blood vessels. This creates a high sodium ion concentration gradient across these membranes, which in turn causes osmosis of water as well. Thus, active transport of sodium ions at the basolateral sides of the epithelial cells results in transport not only of sodium ions but also of water.

These are the mechanisms by which almost all the nutrients, ions, and other substances are absorbed into the blood from the intestine; they are also the way the same substances are reabsorbed from the glomerular filtrate by the renal tubules.

Throughout this text are numerous examples of the different types of transport discussed in this chapter.

Bibliography

Agre P, Kozono D: Aquaporin water channels: molecular mechanisms for human diseases, *FEBS Lett* 555:72, 2003.

Ashcroft FM: From molecule to malady, *Nature* 440:440, 2006.

Benos DJ, Stanton BA: Functional domains within the degenerin/epithelial sodium channel (Deg/ENaC) superfamily of ion channels, *J Physiol* 520:631, 1999.

Benziane B, Chibalin AV: Frontiers: skeletal muscle sodium pump regulation: a translocation paradigm, *Am J Physiol Endocrinol Metab* 295:E553, 2008.

Biel M, Wahl-Schott C, Michalakis S, Zong X: Hyperpolarization-activated cation channels: from genes to function, *Physiol Rev* 89:847, 2009.

Blaustein MP, Zhang J, Chen L, et al: The pump, the exchanger, and endogenous ouabain: signaling mechanisms that link salt retention to hypertension, *Hypertension* 53:291, 2009.

Bröer S: Amino acid transport across mammalian intestinal and renal epithelia, *Physiol Rev* 88:249, 2008.

DeCoursey TE: Voltage-gated proton channels: what's next? *J Physiol* 586:5305, 2008.

DeCoursey TE: Voltage-gated proton channels and other proton transfer pathways, *Physiol Rev* 83:475, 2003.

DiPolo R, Beaugé L: Sodium/calcium exchanger: influence of metabolic regulation on ion carrier interactions, *Physiol Rev* 86:155, 2006.

Drummond HA, Jernigan NL, Grifoni SC: Sensing tension: epithelial sodium channel/acid-sensing ion channel proteins in cardiovascular homeostasis, *Hypertension* 51:1265, 2008.

Gadsby DC: Ion channels versus ion pumps: the principal difference, in principle, *Nat Rev Mol Cell Biol* 10:344, 2009.

Jentsch TJ, Stein V, Weinreich F, Zdebik AA: Molecular structure and physiological function of chloride channels, *Physiol Rev* 82:503, 2002.

Kaupp UB, Seifert R: Cyclic nucleotide-gated ion channels, *Physiol Rev* 82:769, 2002.

King LS, Kozono D, Agre P: From structure to disease: the evolving tale of aquaporin biology, *Nat Rev Mol Cell Biol* 5:687, 2004.

Kleyman TR, Carattino MD, Hughey RP: ENaC at the cutting edge: regulation of epithelial sodium channels by proteases, *J Biol Chem* 284:20447, 2009.

Mazzochi C, Benos DJ, Smith PR: Interaction of epithelial ion channels with the actin-based cytoskeleton, *Am J Physiol Renal Physiol* 291:F1113, 2006.

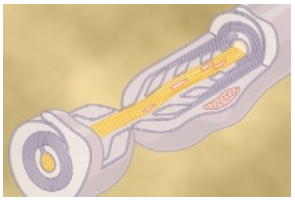
Peres A, Giovannardi S, Bossi E, Fesce R: Electrophysiological insights into the mechanism of ion-coupled cotransporters, *News Physiol Sci* 19:80, 2004.

Russell JM: Sodium-potassium-chloride cotransport, *Physiol Rev* 80:211, 2000.

Shin JM, Munson K, Vagin O, Sachs G: The gastric HK-ATPase: structure, function, and inhibition, *Pflugers Arch* 457:609, 2009.

Tian J, Xie ZJ: The Na-K-ATPase and calcium-signaling microdomains, *Physiology (Bethesda)* 23:205, 2008.

Membrane Potentials and Action Potentials



Electrical potentials exist across the membranes of virtually all cells of the body. In addition, some cells, such as nerve and muscle cells, are capable of generating rapidly changing electrochemical impulses at their membranes, and these impulses are used to transmit signals along the nerve or muscle membranes. In other types of cells, such as glandular cells, macrophages, and ciliated cells, local changes in membrane potentials also activate many of the cells' functions. The present discussion is concerned with membrane potentials generated both at rest and during action by nerve and muscle cells.

Basic Physics of Membrane Potentials

Membrane Potentials Caused by Diffusion

"Diffusion Potential" Caused by an Ion Concentration Difference on the Two Sides of the Membrane. In Figure 5-1*A*, the potassium concentration is great *inside* a nerve fiber membrane but very low *outside* the membrane. Let us assume that the membrane in this instance is permeable to the potassium ions but not to any other ions. Because of the large potassium concentration gradient from inside toward outside, there is a strong tendency for extra numbers of potassium ions to diffuse outward through the membrane. As they do so, they carry positive electrical charges to the outside, thus creating electropositivity outside the membrane and electronegativity inside because of negative anions that remain behind and do not diffuse outward with the potassium. Within a millisecond or so, the potential difference between the inside and outside, called the *diffusion potential,* becomes great enough to block further net potassium diffusion to the exterior, despite the high potassium ion concentration gradient. In the normal mammalian nerve fiber, *the potential difference required is about 94 millivolts, with negativity inside the fiber membrane.*

Figure 5-1*B* shows the same phenomenon as in Figure 5-1*A*, but this time with high concentration of sodium ions *outside* the membrane and low sodium *inside*. These ions are also positively charged. This time, the membrane is highly permeable to the sodium ions but impermeable to all other ions. Diffusion of the positively charged sodium ions to the inside creates a membrane potential of opposite polarity to that in Figure 5-1*A*, with negativity outside and positivity inside. Again, the membrane potential rises high enough within milliseconds to block further net diffusion of sodium ions to the inside; however, this time, in the mammalian nerve fiber, *the potential is about 61 millivolts positive inside the fiber.*

Thus, in both parts of Figure 5-1, we see that a concentration difference of ions across a selectively permeable membrane can, under appropriate conditions, create a membrane potential. Later in this chapter, we show that many of the rapid changes in membrane potentials observed during nerve and muscle impulse transmission result from the occurrence of such rapidly changing diffusion potentials.

Relation of the Diffusion Potential to the Concentration Difference—The Nernst Potential. The diffusion potential level across a membrane that exactly opposes the net diffusion of a particular ion through the membrane is called the *Nernst potential* for that ion, a term that was introduced in Chapter 4.

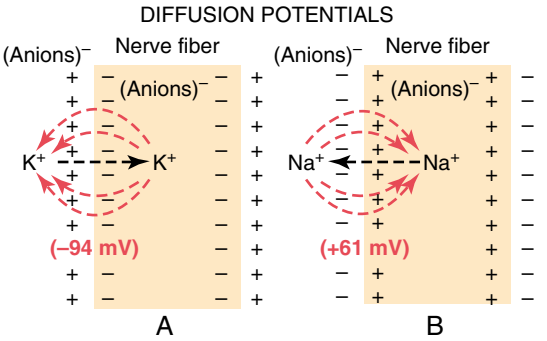


Figure 5-1 *A,* Establishment of a "diffusion" potential across a nerve fiber membrane, caused by diffusion of potassium ions from inside the cell to outside through a membrane that is selectively permeable only to potassium. *B,* Establishment of a "diffusion potential" when the nerve fiber membrane is permeable only to sodium ions. Note that the internal membrane potential is negative when potassium ions diffuse and positive when sodium ions diffuse because of opposite concentration gradients of these two ions.

The magnitude of this Nernst potential is determined by the *ratio* of the concentrations of that specific ion on the two sides of the membrane. The greater this ratio, the greater the tendency for the ion to diffuse in one direction, and therefore the greater the Nernst potential required to prevent additional net diffusion. The following equation, called the *Nernst equation,* can be used to calculate the Nernst potential for any univalent ion at normal body temperature of 98.6°F (37°C):

$$\text{EMF (millivolts)} = \pm 61 \times \log \frac{\text{Concentration inside}}{\text{Concentration outside}}$$

where EMF is electromotive force.

When using this formula, it is usually assumed that the potential in the extracellular fluid outside the membrane remains at zero potential, and the Nernst potential is the potential inside the membrane. Also, the sign of the potential is positive (+) if the ion diffusing from inside to outside is a negative ion, and it is negative (–) if the ion is positive. Thus, when the concentration of positive potassium ions on the inside is 10 times that on the outside, the log of 10 is 1, so the Nernst potential calculates to be –61 millivolts inside the membrane.

Calculation of the Diffusion Potential When the Membrane Is Permeable to Several Different Ions

When a membrane is permeable to several different ions, the diffusion potential that develops depends on three factors: (1) the polarity of the electrical charge of each ion, (2) the permeability of the membrane *(P)* to each ion, and (3) the concentrations *(C)* of the respective ions on the inside *(i)* and outside *(o)* of the membrane. Thus, the following formula, called the *Goldman equation,* or the *Goldman-Hodgkin-Katz equation,* gives the calculated membrane potential on the *inside* of the membrane when two univalent positive ions, sodium (Na^+) and potassium (K^+), and one univalent negative ion, chloride (Cl^-), are involved.

$$\text{EMF (millivolts)} = -61 \times \log \frac{C_{Na_i^+} P_{Na^+} + C_{K_i^+} P_{K^+} + C_{Cl_o^-} P_{Cl^-}}{C_{Na_o^+} P_{Na^+} + C_{K_o^+} P_{K^+} + C_{Cl_i^-} P_{Cl^-}}$$

Let us study the importance and the meaning of this equation. First, sodium, potassium, and chloride ions are the most important ions involved in the development of membrane potentials in nerve and muscle fibers, as well as in the neuronal cells in the nervous system. The concentration gradient of each of these ions across the membrane helps determine the voltage of the membrane potential.

Second, the degree of importance of each of the ions in determining the voltage is proportional to the membrane permeability for that particular ion. That is, if the membrane has zero permeability to both potassium and chloride ions, the membrane potential becomes entirely dominated by the concentration gradient of sodium ions alone, and the resulting potential will be equal to the Nernst potential for sodium. The same holds for each of the other two ions if the membrane should become selectively permeable for either one of them alone.

Third, a positive ion concentration gradient from *inside* the membrane to the *outside* causes electronegativity inside the membrane. The reason for this is that excess positive ions diffuse to the outside when their concentration is higher inside than outside. This carries positive charges to the outside but leaves the nondiffusible negative anions on the inside, thus creating electronegativity on the inside. The opposite effect occurs when there is a gradient for a negative ion. That is, a chloride ion gradient from the *outside to the inside* causes negativity inside the cell because excess negatively charged chloride ions diffuse to the inside, while leaving the nondiffusible positive ions on the outside.

Fourth, as explained later, the permeability of the sodium and potassium channels undergoes rapid changes during transmission of a nerve impulse, whereas the permeability of the chloride channels does not change greatly during this process. Therefore, rapid changes in sodium and potassium permeability are primarily responsible for signal transmission in neurons, which is the subject of most of the remainder of this chapter.

Measuring the Membrane Potential

The method for measuring the membrane potential is simple in theory but often difficult in practice because of the small size of most of the fibers. Figure 5-2 shows a small pipette filled with an electrolyte solution. The pipette is impaled through the cell membrane to the interior of the fiber. Then another electrode, called the "indifferent electrode," is placed in the extracellular fluid, and the potential difference between the inside and outside of the fiber is measured using an appropriate voltmeter. This voltmeter is a highly sophisticated electronic apparatus that is capable of measuring small voltages despite extremely high resistance to electrical flow through the tip of the micropipette, which has a lumen diameter usually less than 1 micrometer and a resistance more than a million ohms. For recording rapid *changes* in the membrane potential during transmission of nerve impulses, the microelectrode is connected to an oscilloscope, as explained later in the chapter.

The lower part of Figure 5-2 shows the electrical potential that is measured at each point in or near the nerve fiber membrane, beginning at the left side of the figure and

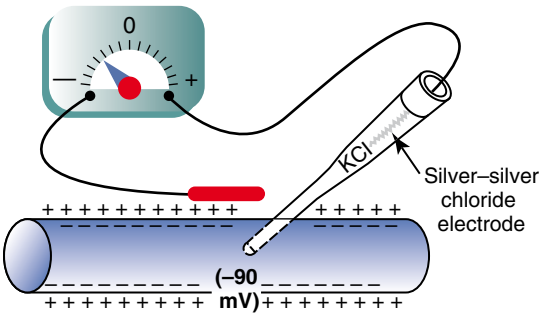


Figure 5-2 Measurement of the membrane potential of the nerve fiber using a microelectrode.

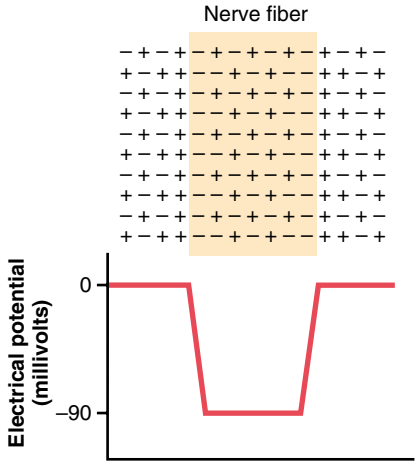


Figure 5-3 Distribution of positively and negatively charged ions in the extracellular fluid surrounding a nerve fiber and in the fluid inside the fiber; note the alignment of negative charges along the inside surface of the membrane and positive charges along the outside surface. The lower panel displays the abrupt changes in membrane potential that occur at the membranes on the two sides of the fiber.

passing to the right. As long as the electrode is outside the nerve membrane, the recorded potential is zero, which is the potential of the extracellular fluid. Then, as the recording electrode passes through the voltage change area at the cell membrane (called the *electrical dipole layer*), the potential decreases abruptly to –90 millivolts. Moving across the center of the fiber, the potential remains at a steady –90-millivolt level but reverses back to zero the instant it passes through the membrane on the opposite side of the fiber.

To create a negative potential inside the membrane, only enough positive ions to develop the electrical dipole layer at the membrane itself must be transported outward. All the remaining ions inside the nerve fiber can be both positive and negative, as shown in the upper panel of Figure 5-3. Therefore, an incredibly small number of ions must be transferred through the membrane to establish the normal "resting potential" of –90 millivolts inside the nerve fiber; this means that only about 1/3,000,000 to 1/100,000,000 of the total positive charges inside the fiber must be transferred. Also, an equally small number of positive ions moving from outside to inside the fiber can reverse the potential from –90 millivolts to as much as +35 millivolts within as little as 1/10,000 of a second. Rapid shifting of ions in this manner causes the nerve signals discussed in subsequent sections of this chapter.

Resting Membrane Potential of Nerves

The resting membrane potential of large nerve fibers when not transmitting nerve signals is about –90 millivolts. That is, the potential *inside the fiber* is 90 millivolts more negative than the potential in the extracellular fluid on the outside of the fiber. In the next few paragraphs, the transport properties of the resting nerve membrane for sodium and potassium and the factors that determine the level of this resting potential are explained.

Active Transport of Sodium and Potassium Ions Through the Membrane—The Sodium-Potassium (Na^+-K^+) Pump. First, let us recall from Chapter 4 that all cell membranes of the body have a powerful Na^+-K^+ pump that continually transports sodium ions to the outside of the cell and potassium ions to the inside, as illustrated on the left-hand side in Figure 5-4. Further, note that this is an *electrogenic pump* because more positive charges are pumped to the outside than to the inside (three Na^+ ions to the outside for each two K^+ ions to the inside), leaving a net deficit of positive ions on the inside; this causes a negative potential inside the cell membrane.

The Na^+-K^+ pump also causes large concentration gradients for sodium and potassium across the resting nerve membrane. These gradients are the following:

Na^+ (outside): 142 mEq/L

Na^+ (inside): 14 mEq/L

K^+ (outside): 4 mEq/L

K^+ (inside): 140 mEq/L

The ratios of these two respective ions from the inside to the outside are

$$Na^+_{inside}/Na^+_{outside} = 0.1$$

$$K^+_{inside}/K^+_{outside} = 35.0$$

Leakage of Potassium Through the Nerve Membrane. The right side of Figure 5-4 shows a channel protein, sometimes called a *"tandem pore domain," potassium channel*, or *potassium (K^+) "leak" channel*, in the nerve membrane through which potassium can leak even in a resting cell. The basic structure of potassium channels was described in Chapter 4 (Figure 4-4). These K^+ leak channels may also leak sodium ions slightly but are far more permeable to potassium than to sodium, normally about 100 times as permeable. As discussed later, this differential in permeability is a key factor in determining the level of the normal resting membrane potential.

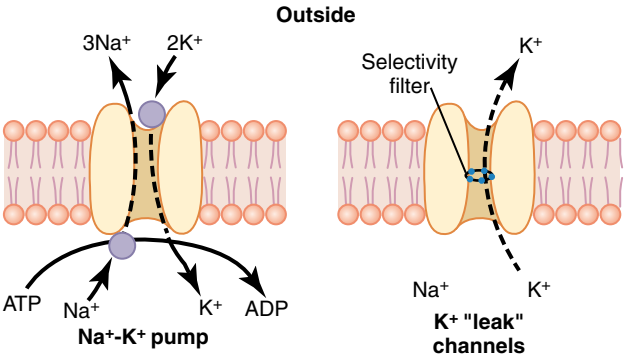


Figure 5-4 Functional characteristics of the Na^+-K^+ pump and of the K^+ "leak" channels. ADP, adenosine diphosphate; ATP, adenosine triphosphate. The K^+ "leak" channels also leak Na^+ ions into the cell slightly, but are much more permeable to K^+.

Origin of the Normal Resting Membrane Potential

Figure 5-5 shows the important factors in the establishment of the normal resting membrane potential of −90 millivolts. They are as follows.

Contribution of the Potassium Diffusion Potential. In Figure 5-5*A*, we make the assumption that the only movement of ions through the membrane is diffusion of potassium ions, as demonstrated by the open channels between the potassium symbols (K^+) inside and outside the membrane. Because of the high ratio of potassium ions inside to outside, 35:1, the Nernst potential corresponding to this ratio is −94 millivolts because the logarithm of 35 is 1.54, and this multiplied by −61 millivolts is −94 millivolts. Therefore, if potassium ions were the only factor causing the resting potential, the resting potential *inside the fiber* would be equal to −94 millivolts, as shown in the figure.

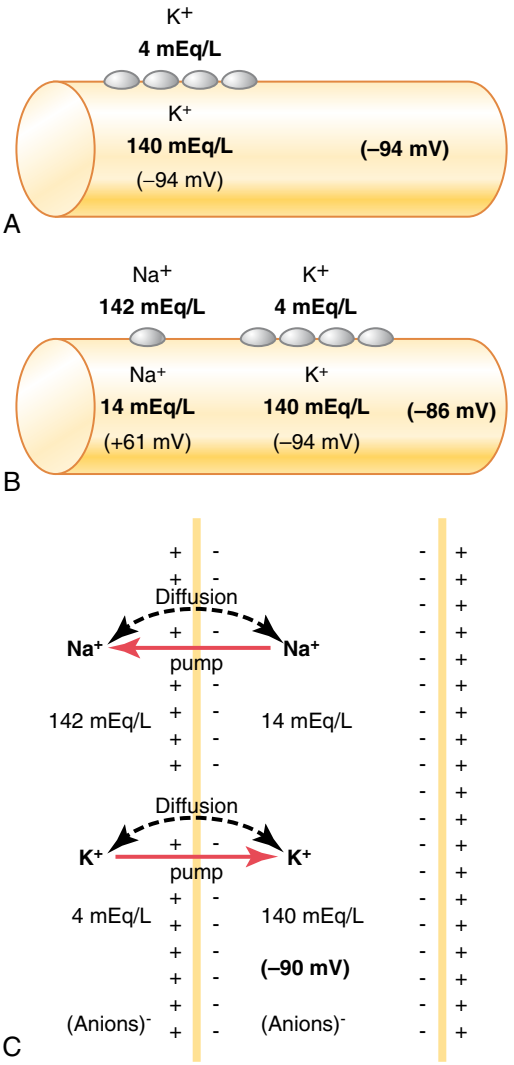


Figure 5-5 Establishment of resting membrane potentials in nerve fibers under three conditions: *A*, when the membrane potential is caused entirely by potassium diffusion alone; *B*, when the membrane potential is caused by diffusion of both sodium and potassium ions; and *C*, when the membrane potential is caused by diffusion of both sodium and potassium ions plus pumping of both these ions by the Na^+-K^+ pump.

Contribution of Sodium Diffusion Through the Nerve Membrane. Figure 5-5*B* shows the addition of slight permeability of the nerve membrane to sodium ions, caused by the minute diffusion of sodium ions through the K^+-Na^+ leak channels. The ratio of sodium ions from inside to outside the membrane is 0.1, and this gives a calculated Nernst potential for the inside of the membrane of +61 millivolts. But also shown in Figure 5-5*B* is the Nernst potential for potassium diffusion of −94 millivolts. How do these interact with each other, and what will be the summated potential? This can be answered by using the Goldman equation described previously. Intuitively, one can see that if the membrane is highly permeable to potassium but only slightly permeable to sodium, it is logical that the diffusion of potassium contributes far more to the membrane potential than does the diffusion of sodium. In the normal nerve fiber, the permeability of the membrane to potassium is about 100 times as great as its permeability to sodium. Using this value in the Goldman equation gives a potential inside the membrane of −86 millivolts, which is near the potassium potential shown in the figure.

Contribution of the Na^+-K^+ Pump. In Figure 5-5*C*, the Na^+-K^+ pump is shown to provide an additional contribution to the resting potential. In this figure, there is continuous pumping of three sodium ions to the outside for each two potassium ions pumped to the inside of the membrane. The fact that more sodium ions are being pumped to the outside than potassium to the inside causes continual loss of positive charges from inside the membrane; this creates an additional degree of negativity (about −4 millivolts additional) on the inside beyond that which can be accounted for by diffusion alone. Therefore, as shown in Figure 5-5*C*, the net membrane potential with all these factors operative at the same time is about −90 millivolts.

In summary, the diffusion potentials alone caused by potassium and sodium diffusion would give a membrane potential of about −86 millivolts, almost all of this being determined by potassium diffusion. Then, an additional −4 millivolts is contributed to the membrane potential by the continuously acting electrogenic Na^+-K^+ pump, giving a net membrane potential of −90 millivolts.

Nerve Action Potential

Nerve signals are transmitted by *action potentials*, which are rapid changes in the membrane potential that spread rapidly along the nerve fiber membrane. Each action potential begins with a sudden change from the normal resting negative membrane potential to a positive potential and then ends with an almost equally rapid change back to the negative potential. To conduct a nerve signal, the action potential moves along the nerve fiber until it comes to the fiber's end.

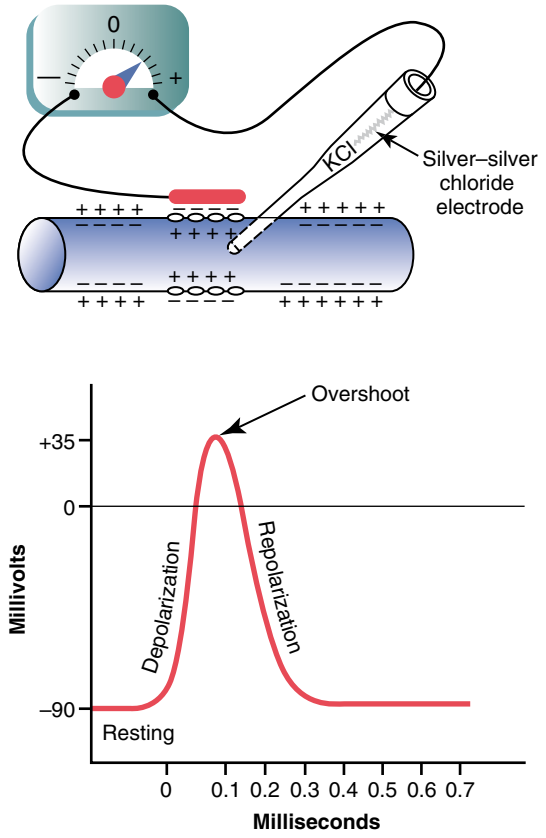


Figure 5-6 Typical action potential recorded by the method shown in the upper panel of the figure.

The upper panel of Figure 5-6 shows the changes that occur at the membrane during the action potential, with transfer of positive charges to the interior of the fiber at its onset and return of positive charges to the exterior at its end. The lower panel shows graphically the successive changes in membrane potential over a few 10,000ths of a second, illustrating the explosive onset of the action potential and the almost equally rapid recovery.

The successive stages of the action potential are as follows.

Resting Stage. This is the resting membrane potential before the action potential begins. The membrane is said to be "polarized" during this stage because of the –90 millivolts negative membrane potential that is present.

Depolarization Stage. At this time, the membrane suddenly becomes permeable to sodium ions, allowing tremendous numbers of positively charged sodium ions to diffuse to the interior of the axon. The normal "polarized" state of –90 millivolts is immediately neutralized by the inflowing positively charged sodium ions, with the potential rising rapidly in the positive direction. This is called *depolarization.* In large nerve fibers, the great excess of positive sodium ions moving to the inside causes the membrane potential to actually "overshoot" beyond the zero level and to become somewhat positive. In some smaller fibers, as well as in many central nervous system neurons, the potential merely approaches the zero level and does not overshoot to the positive state.

Repolarization Stage. Within a few 10,000ths of a second after the membrane becomes highly permeable to sodium ions, the sodium channels begin to close and the potassium channels open more than normal. Then, rapid diffusion of potassium ions to the exterior re-establishes the normal negative resting membrane potential. This is called *repolarization* of the membrane.

To explain more fully the factors that cause both depolarization and repolarization, we will describe the special characteristics of two other types of transport channels through the nerve membrane: the voltage-gated sodium and potassium channels.

Voltage-Gated Sodium and Potassium Channels

The necessary actor in causing both depolarization and repolarization of the nerve membrane during the action potential is the *voltage-gated sodium channel.* A *voltage-gated potassium channel* also plays an important role in increasing the rapidity of repolarization of the membrane. *These two voltage-gated channels are in addition to the Na^+-K^+ pump and the K^+ leak channels.*

Voltage-Gated Sodium Channel—Activation and Inactivation of the Channel

The upper panel of Figure 5-7 shows the voltage-gated sodium channel in three separate states. This channel has two *gates*—one near the outside of the channel called the *activation gate,* and another near the inside called the

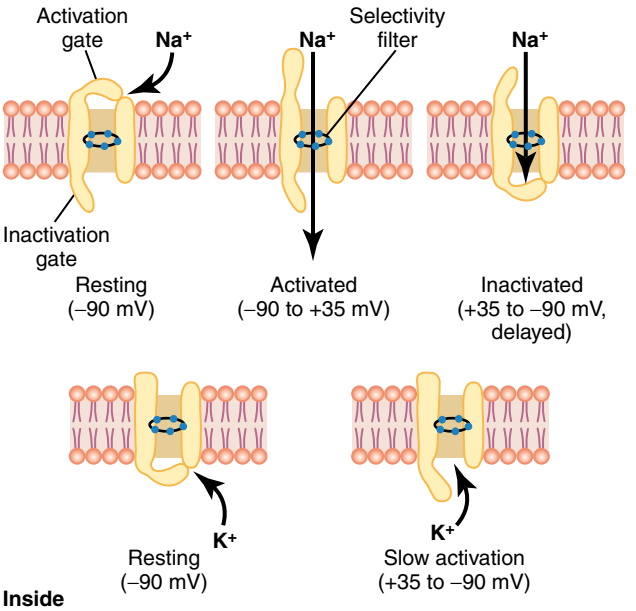


Figure 5-7 Characteristics of the voltage-gated sodium (*top*) and potassium (*bottom*) channels, showing successive activation and inactivation of the sodium channels and delayed activation of the potassium channels when the membrane potential is changed from the normal resting negative value to a positive value.

inactivation gate. The upper left of the figure depicts the state of these two gates in the normal resting membrane when the membrane potential is −90 millivolts. In this state, the activation gate is closed, which prevents any entry of sodium ions to the interior of the fiber through these sodium channels.

Activation of the Sodium Channel. When the membrane potential becomes less negative than during the resting state, rising from −90 millivolts toward zero, it finally reaches a voltage—usually somewhere between −70 and −50 millivolts—that causes a sudden conformational change in the activation gate, flipping it all the way to the open position. This is called the *activated state;* during this state, sodium ions can pour inward through the channel, increasing the sodium permeability of the membrane as much as 500- to 5000-fold.

Inactivation of the Sodium Channel. The upper right panel of Figure 5-7 shows a third state of the sodium channel. The same increase in voltage that opens the activation gate also closes the inactivation gate. The inactivation gate, however, closes a few 10,000ths of a second after the activation gate opens. That is, the conformational change that flips the inactivation gate to the closed state is a slower process than the conformational change that opens the activation gate. Therefore, after the sodium channel has remained open for a few 10,000ths of a second, the inactivation gate closes, and sodium ions no longer can pour to the inside of the membrane. At this point, the membrane potential begins to recover back toward the resting membrane state, which is the repolarization process.

Another important characteristic of the sodium channel inactivation process is that the inactivation gate will not reopen until the membrane potential returns to or near the original resting membrane potential level. Therefore, it is usually not possible for the sodium channels to open again without first repolarizing the nerve fiber.

Voltage-Gated Potassium Channel and Its Activation

The lower panel of Figure 5-7 shows the voltage-gated potassium channel in two states: during the resting state (left) and toward the end of the action potential (right). During the resting state, the gate of the potassium channel is closed and potassium ions are prevented from passing through this channel to the exterior. When the membrane potential rises from −90 millivolts toward zero, this voltage change causes a conformational opening of the gate and allows increased potassium diffusion outward through the channel. However, because of the slight delay in opening of the potassium channels, for the most part, they open just at the same time that the sodium channels are beginning to close because of inactivation. Thus, the decrease in sodium entry to the cell and the simultaneous increase in potassium exit from the cell combine to speed the repolarization process, leading to full recovery of the resting membrane potential within another few 10,000ths of a second.

Research Method for Measuring the Effect of Voltage on Opening and Closing of the Voltage-Gated Channels—The "Voltage Clamp." The original research that led to quantitative understanding of the sodium and potassium channels was so ingenious that it led to Nobel Prizes for the scientists responsible, Hodgkin and Huxley. The essence of these studies is shown in Figures 5-8 and 5-9.

Figure 5-8 shows an experimental apparatus called a *voltage clamp,* which is used to measure flow of ions through the different channels. In using this apparatus, two electrodes are inserted into the nerve fiber. One of these is to measure the voltage of the membrane potential, and the other is to conduct electrical current into or out of the nerve fiber. This apparatus is used in the following way: The investigator decides which voltage he or she wants to establish inside the nerve fiber. The electronic portion of the apparatus is then adjusted to the desired voltage, and this automatically injects either positive or negative electricity through the current electrode at whatever rate is required to hold the voltage, as measured by the voltage electrode, at the level set by the operator. When the membrane potential is suddenly increased by this voltage clamp from −90 millivolts to zero, the voltage-gated sodium and potassium channels open and sodium and potassium ions begin to pour through the channels. To counterbalance

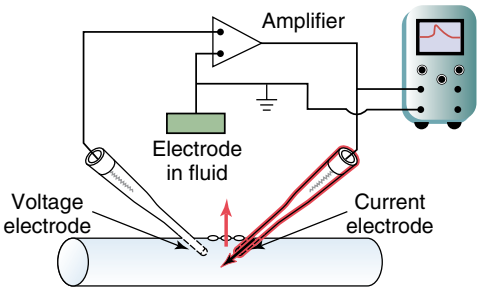


Figure 5-8 "Voltage clamp" method for studying flow of ions through specific channels.

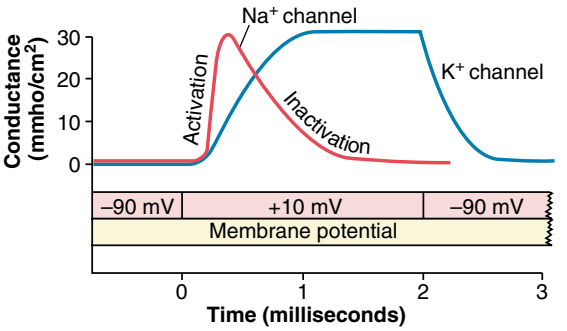


Figure 5-9 Typical changes in conductance of sodium and potassium ion channels when the membrane potential is suddenly increased from the normal resting value of −90 millivolts to a positive value of +10 millivolts for 2 milliseconds. This figure shows that the sodium channels open (activate) and then close (inactivate) before the end of the 2 milliseconds, whereas the potassium channels only open (activate), and the rate of opening is much slower than that of the sodium channels.

the effect of these ion movements on the desired setting of the intracellular voltage, electrical current is injected automatically through the current electrode of the voltage clamp to maintain the intracellular voltage at the required steady zero level. To achieve this, the current injected must be equal to but of opposite polarity to the net current flow through the membrane channels. To measure how much current flow is occurring at each instant, the current electrode is connected to an oscilloscope that records the current flow, as demonstrated on the screen of the oscilloscope in Figure 5-8. Finally, the investigator adjusts the concentrations of the ions to other than normal levels both inside and outside the nerve fiber and repeats the study. This can be done easily when using large nerve fibers removed from some invertebrates, especially the giant squid axon, which in some cases is as large as 1 millimeter in diameter. When sodium is the only permeant ion in the solutions inside and outside the squid axon, the voltage clamp measures current flow only through the sodium channels. When potassium is the only permeant ion, current flow only through the potassium channels is measured.

Another means for studying the flow of ions through an individual type of channel is to block one type of channel at a time. For instance, the sodium channels can be blocked by a toxin called *tetrodotoxin* by applying it to the outside of the cell membrane where the sodium activation gates are located. Conversely, *tetraethylammonium ion* blocks the potassium channels when it is applied to the interior of the nerve fiber.

Figure 5-9 shows typical changes in conductance of the voltage-gated sodium and potassium channels when the membrane potential is suddenly changed by use of the voltage clamp from −90 millivolts to +10 millivolts and then, 2 milliseconds later, back to −90 millivolts. Note the sudden opening of the sodium channels (the activation stage) within a small fraction of a millisecond after the membrane potential is increased to the positive value. However, during the next millisecond or so, the sodium channels automatically close (the inactivation stage).

Note the opening (activation) of the potassium channels. These open slowly and reach their full open state only after the sodium channels have almost completely closed. Further, once the potassium channels open, they remain open for the entire duration of the positive membrane potential and do not close again until after the membrane potential is decreased back to a negative value.

Summary of the Events That Cause the Action Potential

Figure 5-10 shows in summary form the sequential events that occur during and shortly after the action potential. The bottom of the figure shows the changes in membrane conductance for sodium and potassium ions. During the resting state, before the action potential begins, the conductance for potassium ions is 50 to 100 times as great as the conductance for sodium ions. This is caused by much greater leakage of potassium ions than sodium ions through the leak channels. However, at the onset of the action potential, the sodium channels instantaneously become activated and allow up to a 5000-fold increase in

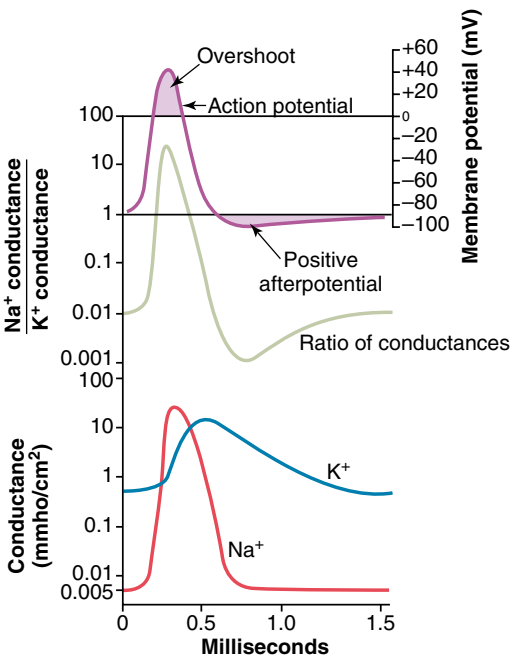


Figure 5-10 Changes in sodium and potassium conductance during the course of the action potential. Sodium conductance increases several thousand-fold during the early stages of the action potential, whereas potassium conductance increases only about 30-fold during the latter stages of the action potential and for a short period thereafter. (These curves were constructed from theory presented in papers by Hodgkin and Huxley but transposed from squid axon to apply to the membrane potentials of large mammalian nerve fibers.)

sodium conductance. Then the inactivation process closes the sodium channels within another fraction of a millisecond. The onset of the action potential also causes voltage gating of the potassium channels, causing them to begin opening more slowly a fraction of a millisecond after the sodium channels open. At the end of the action potential, the return of the membrane potential to the negative state causes the potassium channels to close back to their original status, but again, only after an additional millisecond or more delay.

The middle portion of Figure 5-10 shows the ratio of sodium conductance to potassium conductance at each instant during the action potential, and above this is the action potential itself. During the early portion of the action potential, the ratio of sodium to potassium conductance increases more than 1000-fold. Therefore, far more sodium ions flow to the interior of the fiber than do potassium ions to the exterior. This is what causes the membrane potential to become positive at the action potential onset. Then the sodium channels begin to close and the potassium channels begin to open, so the ratio of conductance shifts far in favor of high potassium conductance but low sodium conductance. This allows very rapid loss of potassium ions to the exterior but virtually zero flow of sodium ions to the interior. Consequently, the action potential quickly returns to its baseline level.

Roles of Other Ions During the Action Potential

Thus far, we have considered only the roles of sodium and potassium ions in the generation of the action potential. At least two other types of ions must be considered: negative anions and calcium ions.

Impermeant Negatively Charged Ions (Anions) Inside the Nerve Axon. Inside the axon are many negatively charged ions that cannot go through the membrane channels. They include the anions of protein molecules and of many organic phosphate compounds, sulfate compounds, and so forth. Because these ions cannot leave the interior of the axon, any deficit of positive ions inside the membrane leaves an excess of these impermeant negative anions. Therefore, these impermeant negative ions are responsible for the negative charge inside the fiber when there is a net deficit of positively charged potassium ions and other positive ions.

Calcium Ions. The membranes of almost all cells of the body have a calcium pump similar to the sodium pump, and calcium serves along with (or instead of) sodium in some cells to cause most of the action potential. Like the sodium pump, the calcium pump transports calcium ions from the interior to the exterior of the cell membrane (or into the endoplasmic reticulum of the cell), creating a calcium ion gradient of about 10,000-fold. This leaves an internal cell concentration of calcium ions of about 10^{-7} molar, in contrast to an external concentration of about 10^{-3} molar.

In addition, there are *voltage-gated calcium channels*. Because calcium ion concentration is more than 10,000 times greater in the extracellular than the intracellular fluid, there is a tremendous diffusion gradient for passive flow of calcium ions into the cells. These channels are slightly permeable to sodium ions and calcium ions, but their permeability to calcium is about 1000-fold greater than to sodium under normal physiological conditions. When they open in response to a stimulus that depolarizes the cell membrane, calcium ions flow to the interior of the cell.

A major function of the voltage-gated calcium ion channels is to contribute to the depolarizing phase on the action potential in some cells. The gating of calcium channels, however, is slow, requiring 10 to 20 times as long for activation as for the sodium channels. For this reason they are often called *slow channels,* in contrast to the sodium channels, which are called *fast channels.* Therefore, the opening of calcium channels provides a more sustained depolarization, whereas the sodium channels play a key role in initiating action potentials.

Calcium channels are numerous in both cardiac muscle and smooth muscle. In fact, in some types of smooth muscle, the fast sodium channels are hardly present; therefore, the action potentials are caused almost entirely by activation of slow calcium channels.

Increased Permeability of the Sodium Channels When There Is a Deficit of Calcium Ions. The concentration of calcium ions in the extracellular fluid also has a profound effect on the voltage level at which the sodium channels become activated. When there is a deficit of calcium ions, the sodium channels become activated (opened) by a small increase of the membrane potential from its normal, very negative level. Therefore, the nerve fiber becomes highly excitable, sometimes discharging repetitively without provocation rather than remaining in the resting state. In fact, the calcium ion concentration needs to fall only 50 percent below normal before spontaneous discharge occurs in some peripheral nerves, often causing *muscle "tetany."* This is sometimes lethal because of tetanic contraction of the respiratory muscles.

The probable way in which calcium ions affect the sodium channels is as follows: These ions appear to bind to the exterior surfaces of the sodium channel protein molecule. The positive charges of these calcium ions in turn alter the electrical state of the sodium channel protein itself, in this way altering the voltage level required to open the sodium gate.

Initiation of the Action Potential

Up to this point, we have explained the changing sodium and potassium permeability of the membrane, as well as the development of the action potential itself, but we have not explained what initiates the action potential.

A Positive-Feedback Cycle Opens the Sodium Channels. First, as long as the membrane of the nerve fiber remains undisturbed, no action potential occurs in the normal nerve. However, if any event causes enough initial rise in the membrane potential from –90 millivolts toward the zero level, the rising voltage itself causes many voltage-gated sodium channels to begin opening. This allows rapid inflow of sodium ions, which causes a further rise in the membrane potential, thus opening still more voltage-gated sodium channels and allowing more streaming of sodium ions to the interior of the fiber. This process is a positive-feedback cycle that, once the feedback is strong enough, continues until all the voltage-gated sodium channels have become activated (opened). Then, within another fraction of a millisecond, the rising membrane potential causes closure of the sodium channels and opening of potassium channels and the action potential soon terminates.

Threshold for Initiation of the Action Potential. An action potential will not occur until the initial rise in membrane potential is great enough to create the positive feedback described in the preceding paragraph. This occurs when the number of Na^+ ions entering the fiber becomes greater than the number of K^+ ions leaving the fiber. A sudden rise in membrane potential of 15 to 30 millivolts is usually required. Therefore, a sudden increase in the membrane potential in a large nerve fiber from –90 millivolts up to about –65 millivolts usually causes the explosive development of an action potential. This level of –65 millivolts is said to be the *threshold* for stimulation.

Propagation of the Action Potential

In the preceding paragraphs, we discussed the action potential as it occurs at one spot on the membrane. However, an action potential elicited at any one point on an excitable membrane usually excites adjacent portions of the membrane, resulting in propagation of the action

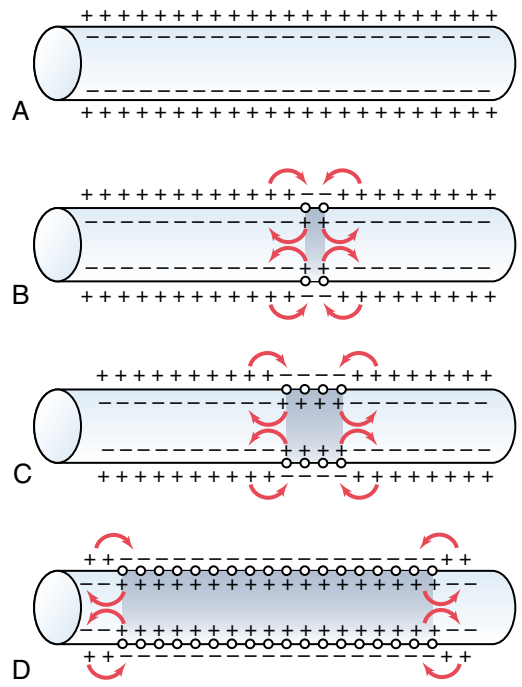


Figure 5-11 Propagation of action potentials in both directions along a conductive fiber.

potential along the membrane. This mechanism is demonstrated in Figure 5-11. Figure 5-11*A* shows a normal resting nerve fiber, and Figure 5-11*B* shows a nerve fiber that has been excited in its midportion—that is, the midportion suddenly develops increased permeability to sodium. The arrows show a "local circuit" of current flow from the depolarized areas of the membrane to the adjacent resting membrane areas. That is, positive electrical charges are carried by the inward-diffusing sodium ions through the depolarized membrane and then for several millimeters in both directions along the core of the axon. These positive charges increase the voltage for a distance of 1 to 3 millimeters inside the large myelinated fiber to above the threshold voltage value for initiating an action potential. Therefore, the sodium channels in these new areas immediately open, as shown in Figure 5-11*C* and *D,* and the explosive action potential spreads. These newly depolarized areas produce still more local circuits of current flow farther along the membrane, causing progressively more and more depolarization. Thus, the depolarization process travels along the entire length of the fiber. This transmission of the depolarization process along a nerve or muscle fiber is called a *nerve* or *muscle impulse.*

Direction of Propagation. As demonstrated in Figure 5-11, an excitable membrane has no single direction of propagation, but the action potential travels in all directions away from the stimulus—even along all branches of a nerve fiber—until the entire membrane has become depolarized.

All-or-Nothing Principle. Once an action potential has been elicited at any point on the membrane of a normal fiber, the depolarization process travels over the entire membrane if conditions are right, or it does not travel at all if conditions are not right. This is called the *all-or-nothing principle,* and it applies to all normal excitable tissues. Occasionally, the action potential reaches a point on the membrane at which it does not generate sufficient voltage to stimulate the next area of the membrane. When this occurs, the spread of depolarization stops. Therefore, for continued propagation of an impulse to occur, the ratio of action potential to threshold for excitation must at all times be greater than 1. This "greater than 1" requirement is called the *safety factor* for propagation.

Re-establishing Sodium and Potassium Ionic Gradients After Action Potentials Are Completed—Importance of Energy Metabolism

The transmission of each action potential along a nerve fiber reduces slightly the concentration differences of sodium and potassium inside and outside the membrane because sodium ions diffuse to the inside during depolarization and potassium ions diffuse to the outside during repolarization. For a single action potential, this effect is so minute that it cannot be measured. Indeed, 100,000 to 50 million impulses can be transmitted by large nerve fibers before the concentration differences reach the point that action potential conduction ceases. Even so, with time, it becomes necessary to re-establish the sodium and potassium membrane concentration differences. This is achieved by action of the Na^+-K^+ pump in the same way as described previously in the chapter for the original establishment of the resting potential. That is, sodium ions that have diffused to the interior of the cell during the action potentials and potassium ions that have diffused to the exterior must be returned to their original state by the Na^+-K^+ pump. Because this pump requires energy for operation, this "recharging" of the nerve fiber is an active metabolic process, using energy derived from the adenosine triphosphate (ATP) energy system of the cell. Figure 5-12 shows that the nerve fiber produces excess heat during recharging, which is a measure of energy expenditure when the nerve impulse frequency increases.

A special feature of the Na^+-K^+ ATPase pump is that its degree of activity is strongly stimulated when excess sodium ions accumulate inside the cell membrane. In fact, the pumping activity increases approximately in proportion to the third power of this intracellular sodium concentration. That is, as the internal sodium concentration rises from 10 to 20 mEq/L, the activity of the pump does not merely double but increases about eightfold. Therefore, it is easy to understand how the "recharging" process of the nerve fiber can be set rapidly into motion whenever the concentration differences of sodium and potassium ions across the membrane begin to "run down."

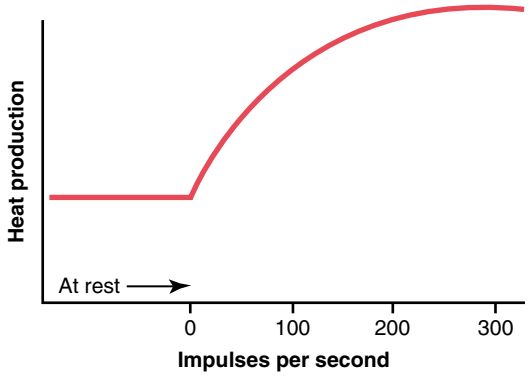


Figure 5-12 Heat production in a nerve fiber at rest and at progressively increasing rates of stimulation.

Plateau in Some Action Potentials

In some instances, the excited membrane does not repolarize immediately after depolarization; instead, the potential remains on a plateau near the peak of the spike potential for many milliseconds, and only then does repolarization begin. Such a plateau is shown in Figure 5-13; one can readily see that the plateau greatly prolongs the period of depolarization. This type of action potential occurs in heart muscle fibers, where the plateau lasts for as long as 0.2 to 0.3 second and causes contraction of heart muscle to last for this same long period.

The cause of the plateau is a combination of several factors. First, in heart muscle, two types of channels enter into the depolarization process: (1) the usual voltage-activated sodium channels, called *fast channels,* and (2) voltage-activated calcium-sodium channels, which are slow to open and therefore are called *slow channels.* Opening of fast channels causes the spike portion of the action potential, whereas the prolonged opening of the slow calcium-sodium channels mainly allows calcium ions to enter the fiber, which is largely responsible for the plateau portion of the action potential as well.

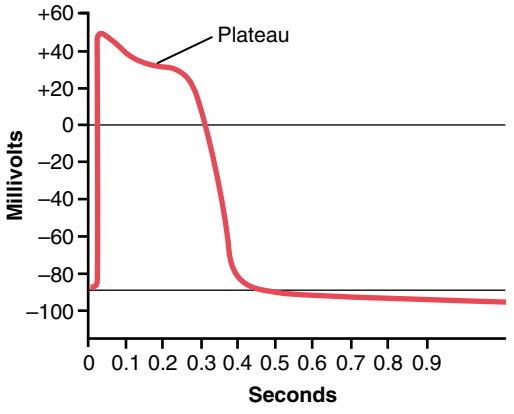


Figure 5-13 Action potential (in millivolts) from a Purkinje fiber of the heart, showing a "plateau."

A second factor that may be partly responsible for the plateau is that the voltage-gated potassium channels are slower than usual to open, often not opening much until the end of the plateau. This delays the return of the membrane potential toward its normal negative value of –80 to –90 millivolts.

Rhythmicity of Some Excitable Tissues—Repetitive Discharge

Repetitive self-induced discharges occur normally in the heart, in most smooth muscle, and in many of the neurons of the central nervous system. These rhythmical discharges cause (1) the rhythmical beat of the heart, (2) rhythmical peristalsis of the intestines, and (3) such neuronal events as the rhythmical control of breathing.

Also, almost all other excitable tissues can discharge repetitively if the threshold for stimulation of the tissue cells is reduced low enough. For instance, even large nerve fibers and skeletal muscle fibers, which normally are highly stable, discharge repetitively when they are placed in a solution that contains the drug veratrine or when the calcium ion concentration falls below a critical value, both of which increase sodium permeability of the membrane.

Re-excitation Process Necessary for Spontaneous Rhythmicity. For spontaneous rhythmicity to occur, the membrane even in its natural state must be permeable enough to sodium ions (or to calcium and sodium ions through the slow calcium-sodium channels) to allow automatic membrane depolarization. Thus, Figure 5-14 shows that the "resting" membrane potential in the rhythmical control center of the heart is only –60 to –70 millivolts. This is not enough negative voltage to keep the sodium and calcium channels totally closed. Therefore, the following sequence occurs: (1) some sodium and calcium ions flow inward; (2) this increases the membrane voltage in the positive direction, which further increases membrane permeability; (3) still more ions flow inward;

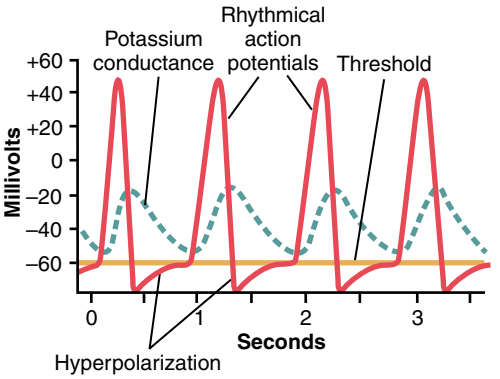


Figure 5-14 Rhythmical action potentials (in millivolts) similar to those recorded in the rhythmical control center of the heart. Note their relationship to potassium conductance and to the state of hyperpolarization.

and (4) the permeability increases more, and so on, until an action potential is generated. Then, at the end of the action potential, the membrane repolarizes. After another delay of milliseconds or seconds, spontaneous excitability causes depolarization again and a new action potential occurs spontaneously. This cycle continues over and over and causes self-induced rhythmical excitation of the excitable tissue.

Why does the membrane of the heart control center not depolarize immediately after it has become repolarized, rather than delaying for nearly a second before the onset of the next action potential? The answer can be found by observing the curve labeled "potassium conductance" in Figure 5-14. This shows that toward the end of each action potential, and continuing for a short period thereafter, the membrane becomes more permeable to potassium ions. The increased outflow of potassium ions carries tremendous numbers of positive charges to the outside of the membrane, leaving inside the fiber considerably more negativity than would otherwise occur. This continues for nearly a second after the preceding action potential is over, thus drawing the membrane potential nearer to the potassium Nernst potential. This is a state called *hyperpolarization,* also shown in Figure 5-14. As long as this state exists, self-re-excitation will not occur. But the increased potassium conductance (and the state of hyperpolarization) gradually disappears, as shown after each action potential is completed in the figure, thereby allowing the membrane potential again to increase up to the *threshold* for excitation. Then, suddenly, a new action potential results and the process occurs again and again.

Special Characteristics of Signal Transmission in Nerve Trunks

Myelinated and Unmyelinated Nerve Fibers. Figure 5-15 shows a cross section of a typical small nerve, revealing many large nerve fibers that constitute most of the cross-sectional area. However, a more careful look reveals many more small fibers lying between the large ones. The large fibers are *myelinated,* and the small ones are *unmyelinated.* The average nerve trunk contains about twice as many unmyelinated fibers as myelinated fibers.

Figure 5-16 shows a typical myelinated fiber. The central core of the fiber is the *axon,* and the membrane of the axon is the membrane that actually conducts the action potential. The axon is filled in its center with *axoplasm,* which is a viscid intracellular fluid. Surrounding the axon is a *myelin sheath* that is often much thicker than the axon itself. About once every 1 to 3 millimeters along the length of the myelin sheath is a *node of Ranvier.*

The myelin sheath is deposited around the axon by Schwann cells in the following manner: The membrane of a Schwann cell first envelops the axon. Then the Schwann cell rotates around the axon many times, laying down multiple layers of Schwann cell membrane containing the lipid substance *sphingomyelin.* This substance is an excellent electrical insulator that decreases ion flow through the membrane about 5000-fold. At the juncture between each two successive Schwann cells along the axon, a small uninsulated area only 2 to 3 *micro*meters in length remains where ions still can flow with ease through the axon membrane between the extracellular fluid and the intracellular fluid inside the axon. This area is called the *node of Ranvier.*

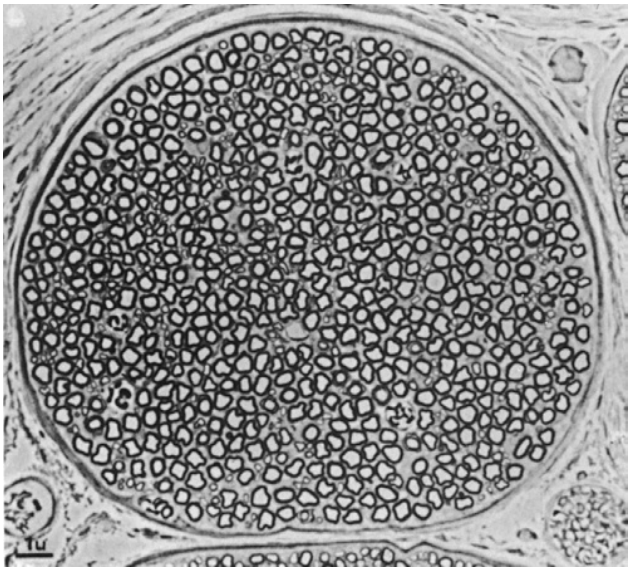


Figure 5-15 Cross section of a small nerve trunk containing both myelinated and unmyelinated fibers.

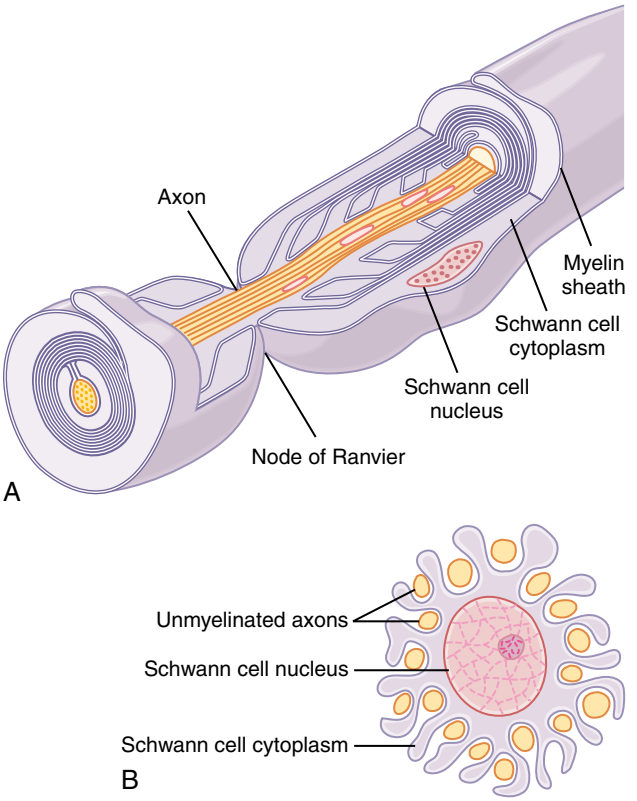


Figure 5-16 Function of the Schwann cell to insulate nerve fibers. *A,* Wrapping of a Schwann cell membrane around a large axon to form the myelin sheath of the myelinated nerve fiber. *B,* Partial wrapping of the membrane and cytoplasm of a Schwann cell around multiple unmyelinated nerve fibers (shown in cross section). (*A,* Modified from Leeson TS, Leeson R: Histology. Philadelphia: WB Saunders, 1979.)

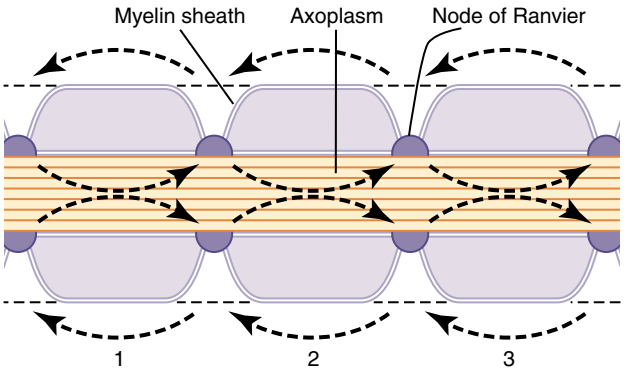


Figure 5-17 Saltatory conduction along a myelinated axon. Flow of electrical current from node to node is illustrated by the arrows.

"Saltatory" Conduction in Myelinated Fibers from Node to Node. Even though almost no ions can flow through the thick myelin sheaths of myelinated nerves, they can flow with ease through the nodes of Ranvier. Therefore, action potentials occur *only at the nodes.* Yet the action potentials are conducted from node to node, as shown in Figure 5-17; this is called *saltatory conduction.* That is, electrical current flows through the surrounding extracellular fluid outside the myelin sheath, as well as through the axoplasm inside the axon from node to node, exciting successive nodes one after another. Thus, the nerve impulse jumps along the fiber, which is the origin of the term "saltatory."

Saltatory conduction is of value for two reasons. First, by causing the depolarization process to jump long intervals along the axis of the nerve fiber, this mechanism increases the velocity of nerve transmission in myelinated fibers as much as 5- to 50-fold. Second, saltatory conduction conserves energy for the axon because only the nodes depolarize, allowing perhaps 100 times less loss of ions than would otherwise be necessary, and therefore requiring little metabolism for re-establishing the sodium and potassium concentration differences across the membrane after a series of nerve impulses.

Still another feature of saltatory conduction in large myelinated fibers is the following: The excellent insulation afforded by the myelin membrane and the 50-fold decrease in membrane capacitance allow repolarization to occur with little transfer of ions.

Velocity of Conduction in Nerve Fibers. The velocity of action potential conduction in nerve fibers varies from as little as 0.25 m/sec in small unmyelinated fibers to as great as 100 m/sec (the length of a football field in 1 second) in large myelinated fibers.

Excitation—The Process of Eliciting the Action Potential

Basically, any factor that causes sodium ions to begin to diffuse inward through the membrane in sufficient numbers can set off automatic regenerative opening of the sodium channels. This can result from *mechanical* disturbance of the membrane, *chemical* effects on the membrane, or passage of *electricity* through the membrane. All these are used at different points in the body to elicit nerve or muscle action potentials: mechanical pressure to excite sensory nerve endings in the skin, chemical neurotransmitters to transmit signals from one neuron to the next in the brain, and electrical current to transmit signals between successive muscle cells in the heart and intestine. For the purpose of understanding the excitation process, let us begin by discussing the principles of electrical stimulation.

Excitation of a Nerve Fiber by a Negatively Charged Metal Electrode. The usual means for exciting a nerve or muscle in the experimental laboratory is to apply electricity to the nerve or muscle surface through two small electrodes, one of which is negatively charged and the other positively charged. When this is done, the excitable membrane becomes stimulated at the negative electrode.

The cause of this effect is the following: Remember that the action potential is initiated by the opening of voltage-gated sodium channels. Further, these channels are opened by a decrease in the normal resting electrical voltage across the membrane. That is, negative current from the electrode decreases the voltage on the outside of the membrane to a negative value nearer to the voltage of the negative potential inside the fiber. This decreases the electrical voltage across the membrane and allows the sodium channels to open, resulting in an action potential. Conversely, at the positive electrode, the injection of positive charges on the outside of the nerve membrane heightens the voltage difference across the membrane rather than lessening it. This causes a state of hyperpolarization, which actually decreases the excitability of the fiber rather than causing an action potential.

Threshold for Excitation, and "Acute Local Potentials." A weak negative electrical stimulus may not be able to excite a fiber. However, when the voltage of the stimulus is increased, there comes a point at which excitation does take place. Figure 5-18 shows the effects of successively applied stimuli of progressing strength. A weak stimulus at point A causes the membrane potential to change from −90 to −85 millivolts, but this is not a sufficient change for the automatic regenerative processes of the action potential to develop. At point B, the stimulus is greater, but again, the intensity is still not enough. The stimulus does, however, disturb the membrane potential locally for as long as 1 millisecond or more after both of these weak stimuli. These local potential changes are called *acute local potentials,* and when they fail to elicit an action potential, they are called *acute subthreshold potentials.*

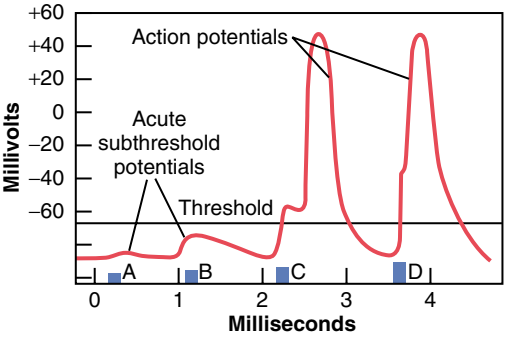


Figure 5-18 Effect of stimuli of increasing voltages to elicit an action potential. Note development of "acute subthreshold potentials" when the stimuli are below the threshold value required for eliciting an action potential.

At point C in Figure 5-18, the stimulus is even stronger. Now the local potential has barely reached the level required to elicit an action potential, called the *threshold level,* but this occurs only after a short "latent period." At point D, the stimulus is still stronger, the acute local potential is also stronger, and the action potential occurs after less of a latent period.

Thus, this figure shows that even a weak stimulus causes a local potential change at the membrane, but the intensity of the local potential must rise to a threshold level before the action potential is set off.

"Refractory Period" After an Action Potential, During Which a New Stimulus Cannot Be Elicited

A new action potential cannot occur in an excitable fiber as long as the membrane is still depolarized from the preceding action potential. The reason for this is that shortly after the action potential is initiated, the sodium channels (or calcium channels, or both) become inactivated and no amount of excitatory signal applied to these channels at this point will open the inactivation gates. The only condition that will allow them to reopen is for the membrane potential to return to or near the original resting membrane potential level. Then, within another small fraction of a second, the inactivation gates of the channels open and a new action potential can be initiated.

The period during which a second action potential cannot be elicited, even with a strong stimulus, is called the *absolute refractory period.* This period for large myelinated nerve fibers is about 1/2500 second. Therefore, one can readily calculate that such a fiber can transmit a maximum of about 2500 impulses per second.

Inhibition of Excitability—"Stabilizers" and Local Anesthetics

In contrast to the factors that increase nerve excitability, still others, called *membrane-stabilizing factors, can decrease excitability.* For instance, *a high extracellular fluid calcium ion concentration* decreases membrane permeability to sodium ions and simultaneously reduces excitability. Therefore, calcium ions are said to be a "stabilizer."

Local Anesthetics. Among the most important stabilizers are the many substances used clinically as local anesthetics, including *procaine* and *tetracaine.* Most of these act directly on the activation gates of the sodium channels, making it much more difficult for these gates to open, thereby reducing membrane excitability. When excitability has been reduced so low that the ratio of *action potential strength to excitability threshold* (called the "safety factor") is reduced below 1.0, nerve impulses fail to pass along the anesthetized nerves.

Recording Membrane Potentials and Action Potentials

Cathode Ray Oscilloscope. Earlier in this chapter, we noted that the membrane potential changes extremely rapidly during the course of an action potential. Indeed, most of the action potential complex of large nerve fibers takes place in less than 1/1000 second. In some figures of this chapter, an electrical meter has been shown recording these potential changes. However, it must be understood that any meter capable of recording most action potentials must be capable of responding extremely rapidly. For practical purposes, the only common type of meter that is capable of responding accurately to the rapid membrane potential changes is the cathode ray oscilloscope.

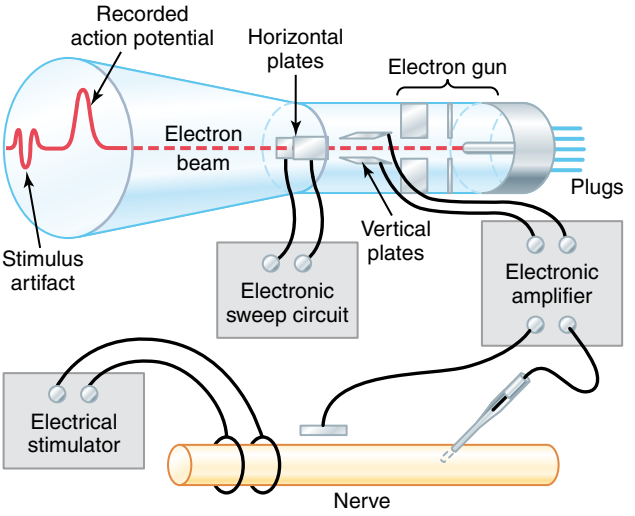


Figure 5-19 Cathode ray oscilloscope for recording transient action potentials.

Figure 5-19 shows the basic components of a cathode ray oscilloscope. The cathode ray tube itself is composed basically of an *electron gun* and a *fluorescent screen* against which electrons are fired. Where the electrons hit the screen surface, the fluorescent material glows. If the electron beam is moved across the screen, the spot of glowing light also moves and draws a fluorescent line on the screen.

In addition to the electron gun and fluorescent surface, the cathode ray tube is provided with two sets of electrically charged plates—one set positioned on the two sides of the electron beam, and the other set positioned above and below. Appropriate electronic control circuits change the voltage on these plates so that the electron beam can be bent up or down in response to electrical signals coming from recording electrodes on nerves. The beam of electrons also is swept horizontally across the screen at a constant time rate by an internal electronic circuit of the oscilloscope. This gives the record shown on the face of the cathode ray tube in the figure, giving a time base horizontally and voltage changes from the nerve electrodes shown vertically. Note at the left end of the record a small *stimulus artifact* caused by the electrical stimulus used to elicit the nerve action potential. Then further to the right is the recorded action potential itself.

Bibliography

Alberts B, Johnson A, Lewis J, et al: *Molecular Biology of the Cell*, ed 3, New York, 2008, Garland Science.

Biel M, Wahl-Schott C, Michalakis S, Zong X: Hyperpolarization-activated cation channels: from genes to function, *Physiol Rev* 89:847, 2009.

Blaesse P, Airaksinen MS, Rivera C, Kaila K: Cation-chloride cotransporters and neuronal function, *Neuron* 61:820, 2009.

Dai S, Hall DD, Hell JW: Supramolecular assemblies and localized regulation of voltage-gated ion channels, *Physiol Rev* 89:411, 2009.

Hodgkin AL, Huxley AF: Quantitative description of membrane current and its application to conduction and excitation in nerve, *J Physiol (Lond)* 117:500, 1952.

Kandel ER, Schwartz JH, Jessell TM: *Principles of Neural Science*, ed 4, New York, 2000, McGraw-Hill.

Kleber AG, Rudy Y: Basic mechanisms of cardiac impulse propagation and associated arrhythmias, *Physiol Rev* 84:431, 2004.

Luján R, Maylie J, Adelman JP: New sites of action for GIRK and SK channels, *Nat Rev Neurosci* 10:475, 2009.

Mangoni ME, Nargeot J: Genesis and regulation of the heart automaticity, *Physiol Rev* 88:919, 2008.

Perez-Reyes E: Molecular physiology of low-voltage-activated T-type calcium channels, *Physiol Rev* 83:117, 2003.

Poliak S, Peles E: The local differentiation of myelinated axons at nodes of Ranvier, *Nat Rev Neurosci* 12:968, 2003.

Schafer DP, Rasband MN: Glial regulation of the axonal membrane at nodes of Ranvier, *Curr Opin Neurobiol* 16:508, 2006.

Vacher H, Mohapatra DP, Trimmer JS: Localization and targeting of voltage-dependent ion channels in mammalian central neurons, *Physiol Rev* 88:1407, 2008.