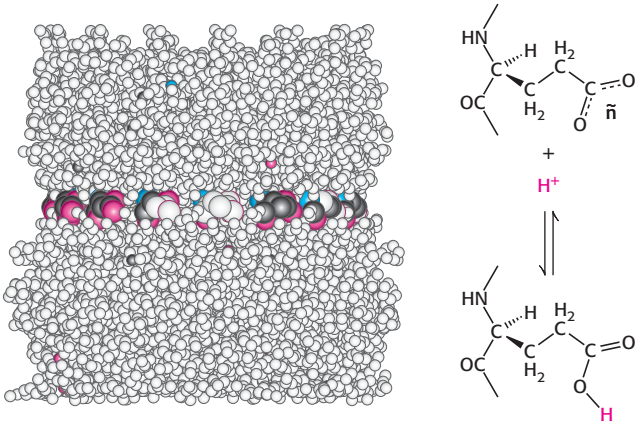
CHAPTER 1

Biochemistry: An Evolving Science



Chemistry in action. Human activities require energy. The interconversion of different forms of energy requires large biochemical machines comprising many thousands of atoms such as the complex shown above. Yet, the functions of these elaborate assemblies depend on simple chemical processes such as the protonation and deprotonation of the carboxylic acid groups shown on the right. The photograph is of Nobel Prize winners Peter Agre, M.D., and Carol Greider, Ph.D., who used biochemical techniques to study the structure and function of proteins. [Courtesy of Johns Hopkins Medicine.]

Biochemistry is the study of the chemistry of life processes. Since the discovery that biological molecules such as urea could be synthesized from nonliving components in 1828, scientists have explored the chemistry of life with great intensity. Through these investigations, many of the most fundamental mysteries of how living things function at a biochemical level have now been solved. However, much remains to be investigated. As is often the case, each discovery raises at least as many new questions as it answers. Furthermore, we are now in an age of unprecedented opportunity for the application of our tremendous knowledge of biochemistry to problems in medicine, dentistry, agriculture, forensics, anthropology, environmental sciences, and many other fields. We begin our journey into biochemistry with one of the most startling discoveries of the past century: namely, the great unity of all living things at the biochemical level.

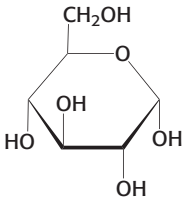
OUTLINE

1.1 Biochemical Unity Underlies Biological Diversity

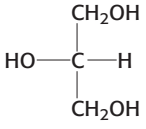
The biological world is magnificently diverse. The animal kingdom is rich with species ranging from nearly microscopic insects to elephants and whales. The plant kingdom includes species as small and relatively simple

as algae and as large and complex as giant sequoias. This diversity extends further when we descend into the microscopic world. Single-celled organisms such as protozoa, yeast, and bacteria are present with great diversity in water, in soil, and on or within larger organisms. Some organisms can survive and even thrive in seemingly hostile environments such as hot springs and glaciers.

The development of the microscope revealed a key unifying feature that underlies this diversity. Large organisms are built up of *cells,* resembling, to some extent, single-celled microscopic organisms. The construction of animals, plants, and microorganisms from cells suggested that these diverse organisms might have more in common than is apparent from their outward appearance. With the development of biochemistry, this suggestion has been tremendously supported and expanded. At the biochemical level, all organisms have many common features (Figure 1.1).

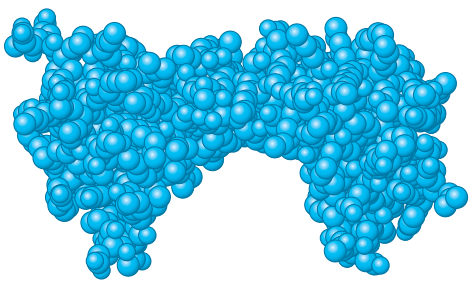


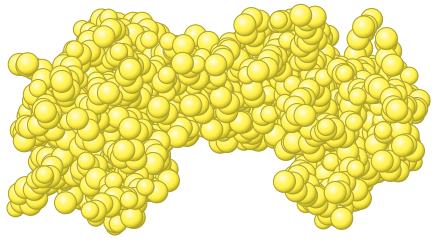
Glucose

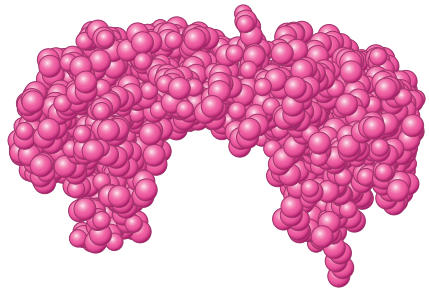


Glycerol

As mentioned earlier, biochemistry is the study of the chemistry of life processes. These processes entail the interplay of two different classes of molecules: large molecules such as proteins and nucleic acids, referred to as *biological macromolecules,* and low-molecular-weight molecules such as glucose and glycerol, referred to as *metabolites,* that are chemically transformed in biological processes. Members of both these classes of molecules are common, with minor variations, to all living things. For example, *deoxyribonucleic acid* (DNA) stores genetic information in all cellular organisms. *Proteins,* the macromolecules that are key participants in most biological processes, are built from the same set of 20 building blocks in all organisms. Furthermore, proteins that play similar roles in different organisms often have very similar three-dimensional structures (see Figure 1.1).


Sulfolobus acidicaldarius


Arabidopsis thaliana


Homo sapiens

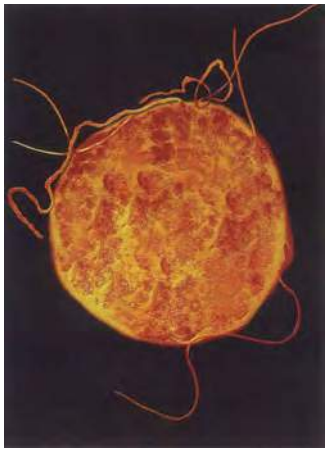


Figure 1.1 Biological diversity and similarity. The shape of a key molecule in gene regulation (the TATA-box-binding protein) is similar in three very different organisms that are separated from one another by billions of years of evolution. [(Left) Dr. T. J. Beveridge/Visuals Unlimited; (middle) Holt Studios/Photo Researchers; (right) Time Life Pictures/Getty Images.]

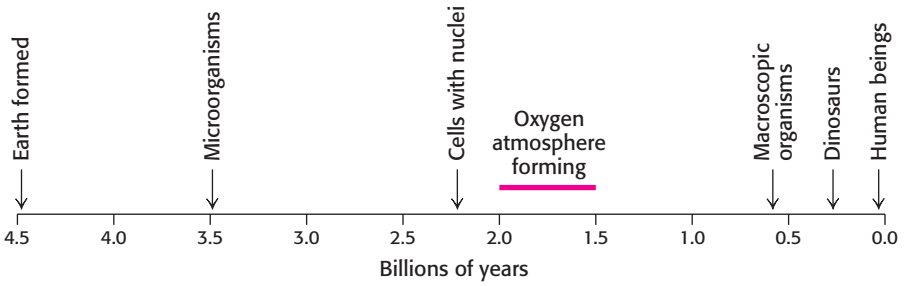


Figure 1.2 A possible time line for biochemical evolution. Selected key events are indicated. Note that life on Earth began approximately 3.5 billion years ago, whereas human beings emerged quite recently.

Key metabolic processes also are common to many organisms. For example, the set of chemical transformations that converts glucose and oxygen into carbon dioxide and water is essentially identical in simple bacteria such as *Escherichia coli (E. coli)* and human beings. Even processes that appear to be quite distinct often have common features at the biochemical level. Remarkably, the biochemical processes by which plants capture light energy and convert it into more-useful forms are strikingly similar to steps used in animals to capture energy released from the breakdown of glucose.

These observations overwhelmingly suggest that all living things on Earth have a common ancestor and that modern organisms have evolved from this ancestor into their present forms. Geological and biochemical findings support a time line for this evolutionary path (Figure 1.2). On the basis of their biochemical characteristics, the diverse organisms of the modern world can be divided into three fundamental groups called *domains: Eukarya* (eukaryotes), *Bacteria,* and *Archaea*. Domain Eukarya comprises all multicellular organisms, including human beings as well as many microscopic unicellular organisms such as yeast. The defining characteristic of *eukaryotes* is the presence of a well-defined nucleus within each cell. Unicellular organisms such as bacteria, which lack a nucleus, are referred to as *prokaryotes*. The prokaryotes were reclassified as two separate domains in response to Carl Woese's discovery in 1977 that certain bacteria-like organisms are biochemically quite distinct from other previously characterized bacterial species. These organisms, now recognized as having diverged from bacteria early in evolution, are the *archaea*. Evolutionary paths from a common ancestor to modern organisms can be deduced on the basis of biochemical information. One such path is shown in Figure 1.3.

Much of this book will explore the chemical reactions and the associated biological macromolecules and metabolites that are found in biological processes common to all organisms. The unity of life at the biochemical level makes this approach possible. At the same time, different organisms have specific needs, depending on the particular biological niche in which they evolved and live. By comparing and contrasting details of particular biochemical pathways in different organisms, we can learn how biological challenges are solved at the biochemical level. In most cases, these challenges are addressed by the adaptation of existing macromolecules to new roles rather than by the evolution of entirely new ones.

Biochemistry has been greatly enriched by our ability to examine the three-dimensional structures of biological

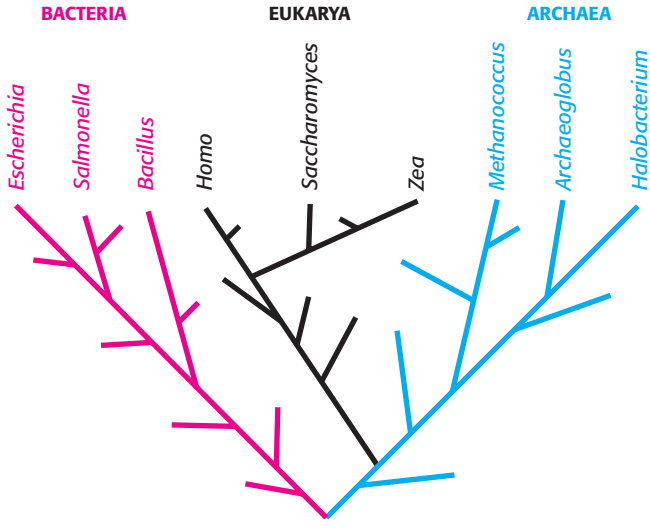


Figure 1.3 The tree of life. A possible evolutionary path from a common ancestor approximately 3.5 billion years ago at the bottom of the tree to organisms found in the modern world at the top.

macromolecules in great detail. Some of these structures are simple and elegant, whereas others are incredibly complicated but, in any case, these structures provide an essential framework for understanding function. We begin our exploration of the interplay between structure and function with the genetic material, DNA.

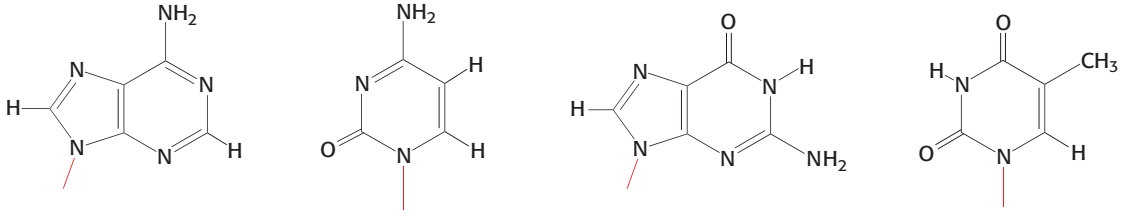
1.2 DNA Illustrates the Interplay Between Form and Function

A fundamental biochemical feature common to all cellular organisms is the use of DNA for the storage of genetic information. The discovery that DNA plays this central role was first made in studies of bacteria in the 1940s. This discovery was followed by the elucidation of the three-dimensional structure of DNA in 1953, an event that set the stage for many of the advances in biochemistry and many other fields, extending to the present.

The structure of DNA powerfully illustrates a basic principle common to all biological macromolecules: the intimate relation between structure and function. The remarkable properties of this chemical substance allow it to function as a very efficient and robust vehicle for storing information. We start with an examination of the covalent structure of DNA and its extension into three dimensions.

DNA is constructed from four building blocks

DNA is a *linear polymer* made up of four different types of monomers. It has a fixed backbone from which protrude variable substituents (Figure 1.4). The backbone is built of repeating sugar–phosphate units. The sugars are molecules of *deoxyribose* from which DNA receives its name. Each sugar is connected to two phosphate groups through different linkages. Moreover, each sugar is oriented in the same way, and so each DNA strand has directionality, with one end distinguishable from the other. Joined to each deoxyribose is one of four possible bases: adenine (A), cytosine (C), guanine (G), and thymine (T).



These bases are connected to the sugar components in the DNA backbone through the bonds shown in black in Figure 1.4. All four bases are planar but differ significantly in other respects. Thus, each monomer of DNA consists of a sugar–phosphate unit and one of four bases attached to the sugar. These bases can be arranged in any order along a strand of DNA.

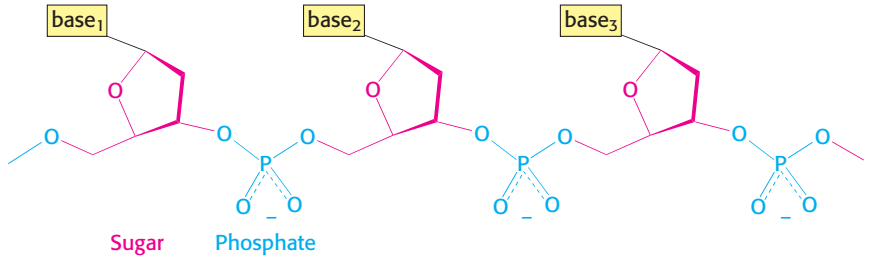


Figure 1.4 Covalent structure of DNA. Each unit of the polymeric structure is composed of a sugar (deoxyribose), a phosphate, and a variable base that protrudes from the sugar–phosphate backbone.

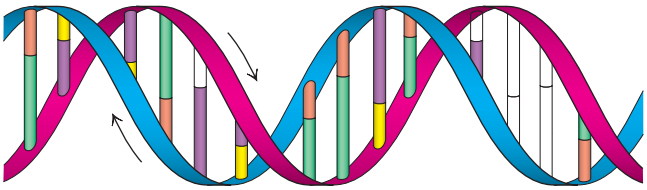


Figure 1.5 The double helix. The double-helical structure of DNA proposed by Watson and Crick. The sugar–phosphate backbones of the two chains are shown in red and blue, and the bases are shown in green, purple, orange, and yellow. The two strands are antiparallel, running in opposite directions with respect to the axis of the double helix, as indicated by the arrows.

Two single strands of DNA combine to form a double helix

Most DNA molecules consist of not one but two strands (Figure 1.5). In 1953, James Watson and Francis Crick deduced the arrangement of these strands and proposed a three-dimensional structure for DNA molecules. This structure is a *double helix* composed of two intertwined strands arranged such that the sugar–phosphate backbone lies on the outside and the bases on the inside. The key to this structure is that the bases form *specific base pairs* (bp) held together by *hydrogen bonds* (Section 1.3): adenine pairs with thymine (A–T) and guanine pairs with cytosine (G–C), as shown in Figure 1.6. Hydrogen bonds are much weaker than *covalent bonds* such as the carbon–carbon or carbon–nitrogen bonds that define the structures of the bases themselves. Such weak bonds are crucial to biochemical systems; they are weak enough to be reversibly broken in biochemical processes, yet they are strong enough, when many form simultaneously, to help stabilize specific structures such as the double helix.

DNA structure explains heredity and the storage of information

The structure proposed by Watson and Crick has two properties of central importance to the role of DNA as the hereditary material. First, the structure is compatible with any sequence of bases. The base pairs have essentially the same shape (see Figure 1.6) and thus fit equally well into the center of the double-helical structure of any sequence. Without any constraints, the sequence of bases along a DNA strand can act as an efficient means of storing information. Indeed, the sequence of bases along DNA strands is how genetic information is stored. The DNA sequence determines the sequences of the ribonucleic acid (RNA) and protein molecules that carry out most of the activities within cells.

Second, because of base-pairing, the sequence of bases along one strand completely determines the sequence along the other strand. As Watson and Crick so coyly wrote: "It has not escaped our notice that the specific pairing

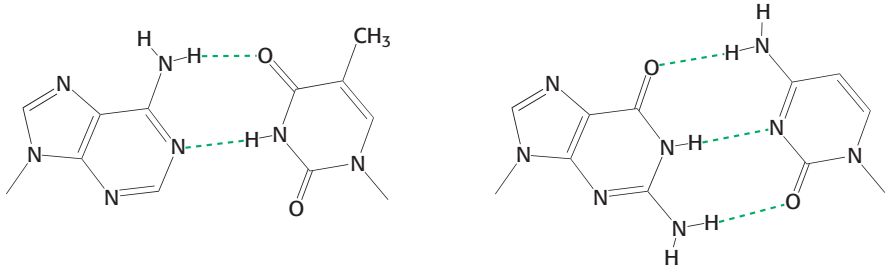


Figure 1.6 Watson–Crick base pairs. Adenine pairs with thymine (A–T), and guanine with cytosine (G–C). The dashed green lines represent hydrogen bonds.

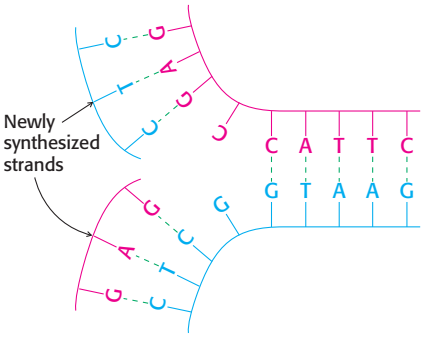


Figure 1.7 DNA replication. If a DNA molecule is separated into two strands, each strand can act as the template for the generation of its partner strand.

we have postulated immediately suggests a possible copying mechanism for the genetic material." Thus, if the DNA double helix is separated into two single strands, each strand can act as a template for the generation of its partner strand through specific base-pair formation (Figure 1.7). The three-dimensional structure of DNA beautifully illustrates the close connection between molecular form and function.

1.3 Concepts from Chemistry Explain the Properties of Biological Molecules

We have seen how a chemical insight, into the hydrogen-bonding capabilities of the bases of DNA, led to a deep understanding of a fundamental biological process. To lay the groundwork for the rest of the book, we begin our study of biochemistry by examining selected concepts from chemistry and showing how these concepts apply to biological systems. The concepts include the types of chemical bonds; the structure of water, the solvent in which most biochemical processes take place; the First and Second Laws of Thermodynamics; and the principles of acid–base chemistry. We will use these concepts to examine an archetypical biochemical process—namely, the formation of a DNA double helix from its two component strands. The process is but one of many examples that could have been chosen to illustrate these topics. Keep in mind that, although the specific discussion is about DNA and double-helix formation, the concepts considered are quite general and will apply to many other classes of molecules and processes that will be discussed in the remainder of the book.

The double helix can form from its component strands

The discovery that DNA from natural sources exists in a double-helical form with Watson–Crick base pairs suggested, but did not prove, that such double helices would form spontaneously outside biological systems. Suppose that two short strands of DNA were chemically synthesized to have complementary sequences so that they could, in principle, form a double helix with Watson–Crick base pairs. Two such sequences are CGATTAAT and ATTAATCG. The structures of these molecules in solution can be examined by a variety of techniques. In isolation, each sequence exists almost exclusively as a single-stranded molecule. However, when the two sequences are mixed, a double helix with Watson–Crick base pairs does form (Figure 1.8). This reaction proceeds nearly to completion.

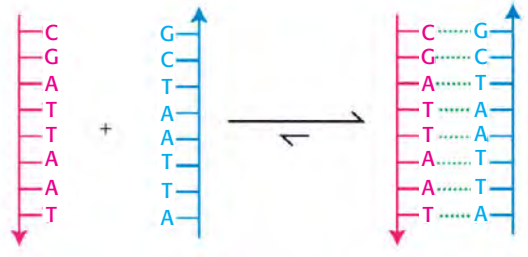


Figure 1.8 Formation of a double helix. When two DNA strands with appropriate, complementary sequences are mixed, they spontaneously assemble to form a double helix.

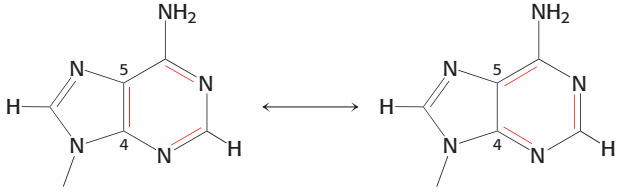
What forces cause the two strands of DNA to bind to each other? To analyze this binding reaction, we must consider several factors: the types of interactions and bonds in biochemical systems and the energetic favorability of the reaction. We must also consider the influence of the solution conditions—in particular, the consequences of acid–base reactions.

Covalent and noncovalent bonds are important for the structure and stability of biological molecules

Atoms interact with one another through chemical bonds. These bonds include the covalent bonds that define the structure of molecules as well as a variety of noncovalent bonds that are of great importance to biochemistry.

Covalent bonds. The strongest bonds are covalent bonds, such as the bonds that hold the atoms together within the individual bases shown on page 4. A covalent bond is formed by the sharing of a pair of electrons between adjacent atoms. A typical carbon–carbon (C—C) covalent bond has a bond length of 1.54 Å and bond energy of 355 kJ mol^{-1} (85 kcal mol^{-1}). Because covalent bonds are so strong, considerable energy must be expended to break them. More than one electron pair can be shared between two atoms to form a multiple covalent bond. For example, three of the bases in Figure 1.6 include carbon–oxygen (C═O) double bonds. These bonds are even stronger than C—C single bonds, with energies near 730 kJ mol^{-1} (175 kcal mol^{-1}) and are somewhat shorter.

For some molecules, more than one pattern of covalent bonding can be written. For example, adenine can be written in two equivalent ways called *resonance structures.*



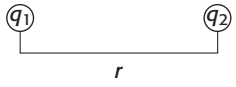
These adenine structures depict alternative arrangements of single and double bonds that are possible within the same structural framework. Resonance structures are shown connected by a double-headed arrow. Adenine's true structure is a composite of its two resonance structures. The composite structure is manifested in the bond lengths such as that for the bond joining carbon atoms C-4 and C-5. The observed bond length of 1.40 Å is between that expected for a C—C single bond (1.54 Å) and a C═C double bond (1.34 Å). A molecule that can be written as several resonance structures of approximately equal energies has greater stability than does a molecule without multiple resonance structures.

Noncovalent bonds. Noncovalent bonds are weaker than covalent bonds but are crucial for biochemical processes such as the formation of a double helix. Four fundamental noncovalent bond types are *electrostatic interactions, hydrogen bonds, van der Waals interactions,* and *hydrophobic interactions*. They differ in geometry, strength, and specificity. Furthermore, these bonds are affected in vastly different ways by the presence of water. Let us consider the characteristics of each type:

1. *Electrostatic Interactions*. A charged group on one molecule can attract an oppositely charged group on another molecule. The energy of an electrostatic interaction is given by *Coulomb's law:*

$$E = kq_1q_2/Dr$$

where E is the energy, q_1 and q_2 are the charges on the two atoms (in units of the electronic charge), r is the distance between the two atoms (in angstroms), D is the dielectric constant (which accounts for the effects of the intervening



Distance and energy units

InterZatomic distances and bond lengths are usually measured in angstrom (Å) units:

$$1\ \text{Å} = 10^{-10}\ \text{m} = 10^{-8}\ \text{cm} = 0.1\ \text{nm}$$

Several energy units are in common use. One joule (J) is the amount of energy required to move 1 meter against a force of 1 newton. A kilojoule (kJ) is 1000 joules. One calorie is the amount of energy required to raise the temperature of 1 gram of water 1 degree Celsius. A kilocalorie (kcal) is 1000 calories. One joule is equal to 0.239 cal.

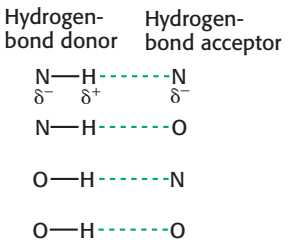
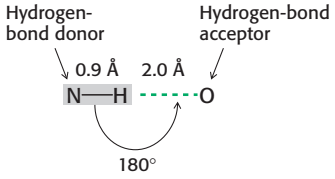


Figure 1.9 Hydrogen bonds. Hydrogen bonds are depicted by dashed green lines. The positions of the partial charges (δ^+ and δ^-) are shown.



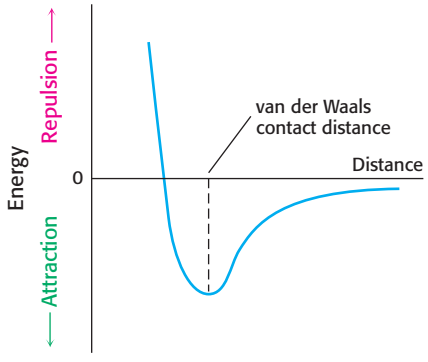
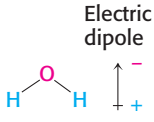


Figure 1.10 Energy of a van der Waals interaction as two atoms approach each other. The energy is most favorable at the van der Waals contact distance. Owing to electron–electron repulsion, the energy rises rapidly as the distance between the atoms becomes shorter than the contact distance.



medium), and k is a proportionality constant ($k = 1389$, for energies in units of kilojoules per mole, or 332 for energies in kilocalories per mole).

By convention, an attractive interaction has a negative energy. The electrostatic interaction between two ions bearing single opposite charges separated by 3 Å in water (which has a dielectric constant of 80) has an energy of 5.8 kJ mol^{-1} (−1.4 kcal mol^{-1}). Note how important the dielectric constant of the medium is. For the same ions separated by 3 Å in a nonpolar solvent such as hexane (which has a dielectric constant of 2), the energy of this interaction is −232 kJ mol^{-1} (−55 kcal mol^{-1}).

2. *Hydrogen Bonds.* These interactions are fundamentally electrostatic interactions. Hydrogen bonds are responsible for specific base-pair formation in the DNA double helix. The hydrogen atom in a hydrogen bond is partially shared by two electronegative atoms such as nitrogen or oxygen. The *hydrogen-bond donor* is the group that includes both the atom to which the hydrogen atom is more tightly linked and the hydrogen atom itself, whereas the *hydrogen-bond acceptor* is the atom less tightly linked to the hydrogen atom (Figure 1.9). The electronegative atom to which the hydrogen atom is covalently bonded pulls electron density away from the hydrogen atom, which thus develops a partial positive charge (δ^+). Thus, the hydrogen atom can interact with an atom having a partial negative charge (δ^-) through an electrostatic interaction.

Hydrogen bonds are much weaker than covalent bonds. They have energies ranging from 4 to 20 kJ mol^{-1} (from 1 to 5 kcal mol^{-1}). Hydrogen bonds are also somewhat longer than covalent bonds; their bond lengths (measured from the hydrogen atom) range from 1.5 Å to 2.6 Å; hence, a distance ranging from 2.4 Å to 3.5 Å separates the two nonhydrogen atoms in a hydrogen bond. The strongest hydrogen bonds have a tendency to be approximately straight, such that the hydrogen-bond donor, the hydrogen atom, and the hydrogen-bond acceptor lie along a straight line. Hydrogen-bonding interactions are responsible for many of the properties of water that make it such a special solvent, as will be described shortly.

3. *van der Waals Interactions.* The basis of a van der Waals interaction is that the distribution of electronic charge around an atom fluctuates with time. At any instant, the charge distribution is not perfectly symmetric. This transient asymmetry in the electronic charge about an atom acts through electrostatic interactions to induce a complementary asymmetry in the electron distribution within its neighboring atoms. The atom and its neighbors then attract one another. This attraction increases as two atoms come closer to each other, until they are separated by the van der Waals *contact distance* (Figure 1.10). At distances shorter than the van der Waals contact distance, very strong repulsive forces become dominant because the outer electron clouds of the two atoms overlap.

Energies associated with van der Waals interactions are quite small; typical interactions contribute from 2 to 4 kJ mol^{-1} (from 0.5 to 1 kcal mol^{-1}) per atom pair. When the surfaces of two large molecules come together, however, a large number of atoms are in van der Waals contact, and the net effect, summed over many atom pairs, can be substantial.

Properties of water. Water is the solvent in which most biochemical reactions take place, and its properties are essential to the formation of macromolecular structures and the progress of chemical reactions. Two properties of water are especially relevant:

1. *Water is a polar molecule.* The water molecule is bent, not linear, and so the distribution of charge is asymmetric. The oxygen nucleus draws elec-

trons away from the two hydrogen nuclei, which leaves the region around each hydrogen atom with a net positive charge. The water molecule is thus an electrically polar structure.

2. *Water is highly cohesive*. Water molecules interact strongly with one another through hydrogen bonds. These interactions are apparent in the structure of ice (Figure 1.11). Networks of hydrogen bonds hold the structure together; similar interactions link molecules in liquid water and account for the cohesion of liquid water, although, in the liquid state, approximately one-fourth of the hydrogen bonds present in ice are broken. The polar nature of water is responsible for its high dielectric constant of 80. Molecules in aqueous solution interact with water molecules through the formation of hydrogen bonds and through ionic interactions. These interactions make water a versatile solvent, able to readily dissolve many species, especially polar and charged compounds that can participate in these interactions.

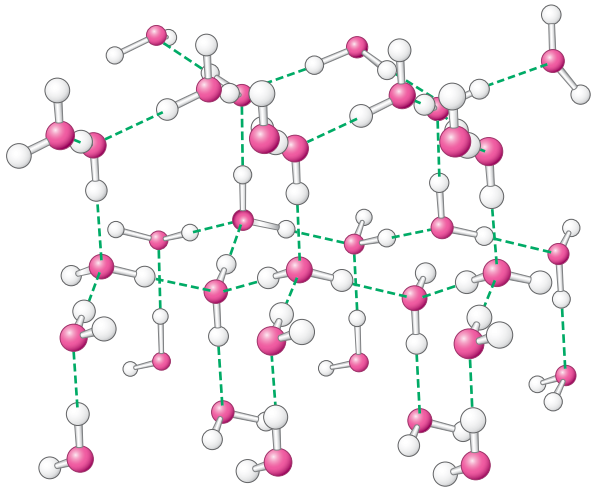


Figure 1.11 Structure of ice. Hydrogen bonds (shown as dashed green lines) are formed between water molecules to produce a highly ordered and open structure.

The hydrophobic effect. A final fundamental interaction called the *hydrophobic effect* is a manifestation of the properties of water. Some molecules (termed *nonpolar molecules*) cannot participate in hydrogen bonding or ionic interactions. The interactions of nonpolar molecules with water molecules are not as favorable as are interactions between the water molecules themselves. The water molecules in contact with these nonpolar molecules form "cages" around them, becoming more well ordered than water molecules free in solution. However, when two such nonpolar molecules come together, some of the water molecules are released, allowing them to interact freely with bulk water (Figure 1.12). The release of water from such cages is favorable for reasons to be considered shortly. The result is that nonpolar

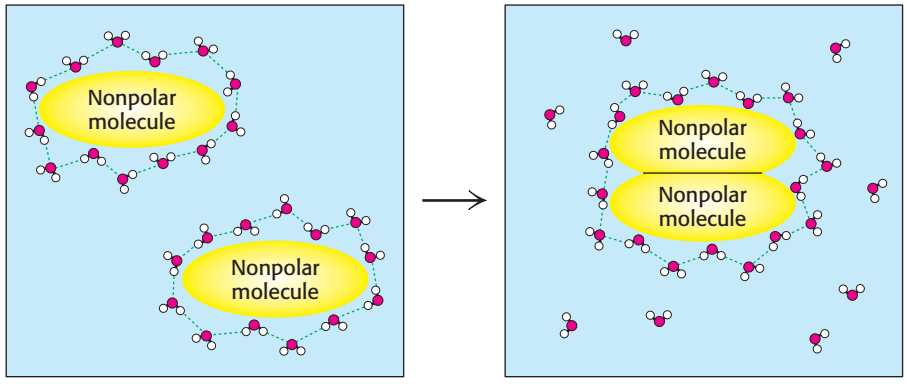


Figure 1.12 The hydrophobic effect. The aggregation of nonpolar groups in water leads to the release of water molecules, initially interacting with the nonpolar surface, into bulk water. The release of water molecules into solution makes the aggregation of nonpolar groups favorable.

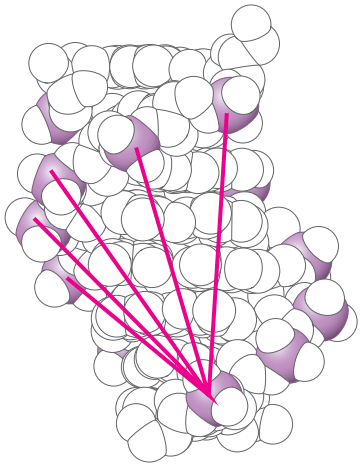


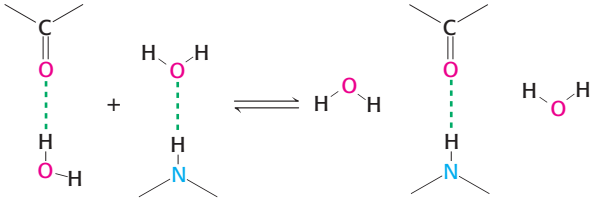
Figure 1.13 Electrostatic interactions in DNA. Each unit within the double helix includes a phosphate group (the phosphorus atom being shown in purple) that bears a negative charge. The unfavorable interactions of one phosphate with several others are shown by red lines. These repulsive interactions oppose the formation of a double helix.

molecules show an increased tendency to associate with one another in water compared with other, less polar and less self-associating, solvents. This tendency is called the hydrophobic effect and the associated interactions are called *hydrophobic interactions*.

The double helix is an expression of the rules of chemistry

Let us now see how these four noncovalent interactions work together in driving the association of two strands of DNA to form a double helix. First, each phosphate group in a DNA strand carries a negative charge. These negatively charged groups interact unfavorably with one another over distances. Thus, unfavorable electrostatic interactions take place when two strands of DNA come together. These phosphate groups are far apart in the double helix with distances greater than 10 Å, but many such interactions take place (Figure 1.13). Thus, electrostatic interactions oppose the formation of the double helix. The strength of these repulsive electrostatic interactions is diminished by the high dielectric constant of water and the presence of ionic species such as Na^+ or Mg^{2+} ions in solution. These positively charged species interact with the phosphate groups and partly neutralize their negative charges.

Second, as already noted, hydrogen bonds are important in determining the formation of specific base pairs in the double helix. However, in single-stranded DNA, the hydrogen-bond donors and acceptors are exposed to solution and can form hydrogen bonds with water molecules.



When two single strands come together, these hydrogen bonds with water are broken and new hydrogen bonds between the bases are formed. Because the number of hydrogen bonds broken is the same as the number formed, these hydrogen bonds do not contribute substantially to driving the overall process of double-helix formation. However, they contribute greatly to the specificity of binding. Suppose two bases that cannot form Watson–Crick base pairs are brought together. Hydrogen bonds with water must be broken as the bases come into contact. Because the bases are not complementary in structure, not all of these bonds can be simultaneously replaced by hydrogen bonds between the bases. Thus, the formation of a double helix between noncomplementary sequences is disfavored.

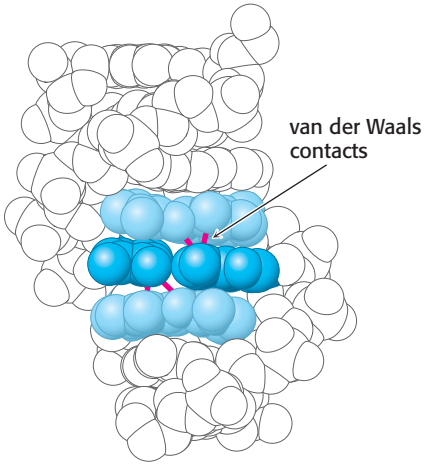


Figure 1.14 Base stacking. In the DNA double helix, adjacent base pairs are stacked nearly on top of one another, and so many atoms in each base pair are separated by their van der Waals contact distance. The central base pair is shown in dark blue and the two adjacent base pairs in light blue. Several van der Waals contacts are shown in red.

Third, within a double helix, the base pairs are parallel and stacked nearly on top of one another. The typical separation between the planes of adjacent base pairs is 3.4 Å, and the distances between the most closely approaching atoms are approximately 3.6 Å. This separation distance corresponds nicely to the van der Waals contact distance (Figure 1.14). Bases tend to stack even in single-stranded DNA molecules. However, the base stacking and associated van der Waals interactions are nearly optimal in a double-helical structure.

Fourth, the hydrophobic effect also contributes to the favorability of base stacking. More-complete base stacking moves the nonpolar surfaces of the bases out of water into contact with each other.

The principles of double-helix formation between two strands of DNA apply to many other biochemical processes. Many weak interactions contribute to the overall energetics of the process, some favorably and some

unfavorably. Furthermore, surface complementarity is a key feature: when complementary surfaces meet, hydrogen-bond donors align with hydrogen-bond acceptors and nonpolar surfaces come together to maximize van der Waals interactions and minimize nonpolar surface area exposed to the aqueous environment. The properties of water play a major role in determining the importance of these interactions.

The laws of thermodynamics govern the behavior of biochemical systems

We can look at the formation of the double helix from a different perspective by examining the laws of thermodynamics. These laws are general principles that apply to all physical (and biological) processes. They are of great importance because they determine the conditions under which specific processes can or cannot take place. We will consider these laws from a general perspective first and then apply the principles that we have developed to the formation of the double helix.

The laws of thermodynamics distinguish between a system and its surroundings. A *system* refers to the matter within a defined region of space. The matter in the rest of the universe is called the *surroundings. The First Law of Thermodynamics states that the total energy of a system and its surroundings is constant*. In other words, the energy content of the universe is constant; energy can be neither created nor destroyed. Energy can take different forms, however. Heat, for example, is one form of energy. Heat is a manifestation of the *kinetic energy* associated with the random motion of molecules. Alternatively, energy can be present as *potential energy*—energy that will be released on the occurrence of some process. Consider, for example, a ball held at the top of a tower. The ball has considerable potential energy because, when it is released, the ball will develop kinetic energy associated with its motion as it falls. Within chemical systems, potential energy is related to the likelihood that atoms can react with one another. For instance, a mixture of gasoline and oxygen has a large potential energy because these molecules may react to form carbon dioxide and water and release energy as heat. The First Law requires that any energy released in the formation of chemical bonds must be used to break other bonds, released as heat, or stored in some other form.

Another important thermodynamic concept is that of *entropy,* a measure of the degree of randomness or disorder in a system. *The Second Law of Thermodynamics states that the total entropy of a system plus that of its surroundings always increases*. For example, the release of water from nonpolar surfaces responsible for the hydrophobic effect is favorable because water molecules free in solution are more disordered than they are when they are associated with nonpolar surfaces. At first glance, the Second Law appears to contradict much common experience, particularly about biological systems. Many biological processes, such as the generation of a leaf from carbon dioxide gas and other nutrients, clearly increase the level of order and hence decrease entropy. Entropy may be decreased locally in the formation of such ordered structures only if the entropy of other parts of the universe is increased by an equal or greater amount. The local decrease in entropy is often accomplished by a release of heat, which increases the entropy of the surroundings.

We can analyze this process in quantitative terms. First, consider the system. The entropy (S) of the system may change in the course of a chemical reaction by an amount ΔS_{system}. If heat flows from the system to its surroundings, then the heat content, often referred to as the *enthalpy* (H), of the system will be reduced by an amount ΔH_{system}. To apply the Second Law, we must determine the change in entropy of the surroundings. If heat flows from the system to the surroundings, then the entropy of the

surroundings will increase. The precise change in the entropy of the surroundings depends on the temperature; the change in entropy is greater when heat is added to relatively cold surroundings than when heat is added to surroundings at high temperatures that are already in a high degree of disorder. To be even more specific, the change in the entropy of the surroundings will be proportional to the amount of heat transferred from the system and inversely proportional to the temperature (T) of the surroundings. In biological systems, T [in kelvins (K), absolute temperature] is usually assumed to be constant. Thus, a change in the entropy of the surroundings is given by

$$\Delta S_{\text{surroundings}} = -\Delta H_{\text{system}}/T \quad (1)$$

The total entropy change is given by the expression

$$\Delta S_{\text{total}} = \Delta S_{\text{system}} + \Delta S_{\text{surroundings}} \quad (2)$$

Substituting equation 1 into equation 2 yields

$$\Delta S_{\text{total}} = S_{\text{system}} - \Delta H_{\text{system}}/T \quad (3)$$

Multiplying by $-T$ gives

$$-T\Delta S_{\text{total}} = \Delta H_{\text{system}} - T\Delta S_{\text{system}} \quad (4)$$

The function $-T\Delta S$ has units of energy and is referred to as *free energy* or *Gibbs free energy,* after Josiah Willard Gibbs, who developed this function in 1878:

$$\Delta G = \Delta H_{\text{system}} - T\Delta S_{\text{system}} \quad (5)$$

The free-energy change, ΔG, will be used throughout this book to describe the energetics of biochemical reactions. The Gibbs free energy is essentially an accounting tool that keeps track of both the entropy of the system (directly) and the entropy of the surroundings (in the form of heat released from the system).

Recall that the Second Law of Thermodynamics states that, for a process to take place, the entropy of the universe must increase. Examination of equation 3 shows that the total entropy will increase if and only if

$$\Delta S_{\text{system}} > \Delta H_{\text{system}}/T \quad (6)$$

Rearranging gives $T\Delta S_{\text{system}} > \Delta H$ or, in other words, entropy will increase if and only if

$$\Delta G = \Delta H_{\text{system}} - T\Delta S_{\text{system}} < 0 \quad (7)$$

Thus, the free-energy change must be negative for a process to take place spontaneously. *There is negative free-energy change when and only when the overall entropy of the universe is increased.* Again, the free energy represents a single term that takes into account both the entropy of the system and the entropy of the surroundings.

Heat is released in the formation of the double helix

Let us see how the principles of thermodynamics apply to the formation of the double helix (Figure 1.15). Suppose solutions containing each of the two single strands are mixed. Before the double helix forms, each of the single strands is free to translate and rotate in solution, whereas each matched pair of strands in the double helix must move together. Furthermore, the free single strands exist in more conformations than possible when bound together in a double helix. Thus, the formation of a double helix from two single strands appears to result in an increase in order for the system, that is, a decrease in the entropy of the system.

On the basis of this analysis, we expect that the double helix cannot form without violating the Second Law of Thermodynamics unless heat is released to increase the entropy of the surroundings. Experimentally, we can measure the heat released by allowing the solutions containing the two single strands to come together within a water bath, which here corresponds to the surroundings. We then determine how much heat must be absorbed by the water bath or released from it to maintain it at a constant temperature. This experiment reveals that a substantial amount of heat is released—namely, approximately 250 kJ mol^{-1} (60 kcal mol^{-1}). This experimental result reveals that the change in enthalpy for the process is quite large, -250 kJ mol^{-1}, consistent with our expectation that significant heat would have to be released to the surroundings for the process not to violate the Second Law. We see in quantitative terms how order within a system can be increased by releasing sufficient heat to the surroundings to ensure that the entropy of the universe increases. We will encounter this general theme again and again throughout this book.

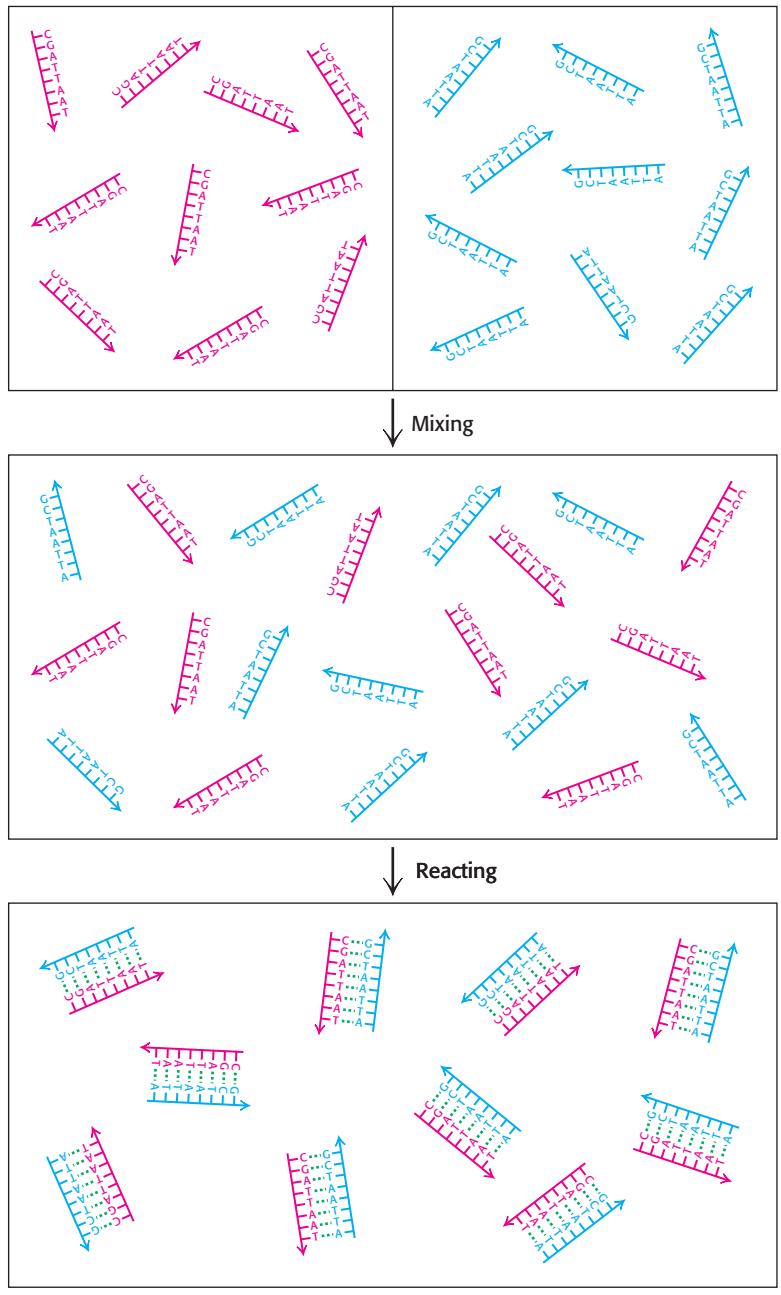


Figure 1.15 Double-helix formation and entropy. When solutions containing DNA strands with complementary sequences are mixed, the strands react to form double helices. This process results in a loss of entropy from the system, indicating that heat must be released to the surroundings to prevent a violation of the Second Law of Thermodynamics.

Acid–base reactions are central in many biochemical processes

Throughout our consideration of the formation of the double helix, we have dealt only with the noncovalent bonds that are formed or broken in this process. Many biochemical processes entail the formation and cleavage of covalent bonds. A particularly important class of reactions prominent in biochemistry is *acid–base reactions*.

In acid and base reactions, hydrogen ions are added to molecules or removed from them. Throughout the book, we will encounter many processes in which the addition or removal of hydrogen atoms is crucial, such as the metabolic processes by which carbohydrates are consumed to release energy for other uses. Thus, a thorough understanding of the basic principles of these reactions is essential.

A hydrogen ion, often written as H^+, corresponds to a proton. In fact, hydrogen ions exist in solution bound to water molecules, thus forming what are known as *hydronium ions*, H_3O^+. For simplicity, we will continue to write H^+, but we should keep in mind that H^+ is shorthand for the actual species present.

The concentration of hydrogen ions in solution is expressed as the pH. Specifically, the *pH* of a solution is defined as

$$\text{pH} = -\log[H^+]$$

where $[H^+]$ is in units of molarity. Thus, pH 7.0 refers to a solution for which $-\log[H^+] = 7.0$, and so $\log[H^+] = -7.0$ and $[H^+] = 10^{\log[H+]} = 10^{-7.0} = 1.0 \times 10^{-7}$ M.

The pH also indirectly expresses the concentration of hydroxide ions, $[OH^-]$, in solution. To see how, we must realize that water molecules dissociate to form H^+ and OH^- ions in an equilibrium process.

$$H_2O \rightleftharpoons H^+ + OH^-$$

The equilibrium constant (K) for the dissociation of water is defined as

$$K = [H^+][OH^-]/[H_2O]$$

and has a value of $K = 1.8 \times 10^{-16}$. Note that an equilibrium constant does not formally have units. Nonetheless, the value of the equilibrium constant given assumes that particular units are used for concentration; in this case and in most others, units of molarity (M) are assumed.

The concentration of water, $[H_2O]$, in pure water is 55.5 M, and this concentration is constant under most conditions. Thus, we can define a new constant, K_W:

$$K_W = K[H_2O] = [H^+][OH^-]$$

$$K[H_2O] = 1.8 \times 10^{-16} \times 55.5$$

$$= 1.0 \times 10^{-14}$$

Because $K_W = [H^+][OH^-] = 1.0 \times 10^{-14}$, we can calculate

$$[OH^-] = 10^{-14}/[H^+] \quad \text{and} \quad [H^+] = 10^{-14}/[OH^-]$$

With these relations in hand, we can easily calculate the concentration of hydroxide ions in an aqueous solution, given the pH. For example, at pH = 7.0, we know that $[H^+] = 10^{-7}$ M and so $[OH^-] = 10^{-14}/10^{-7} = 10^{-7}$ M. In acidic solutions, the concentration of hydrogen ions is higher than 10^{-7} and, hence, the pH is below 7. For example, in 0.1 M HCl, $[H^+] = 10^{-1}$ M and so pH = 1.0 and $[OH^-] = 10^{-14}/10^{-1} = 10^{-13}$ M.

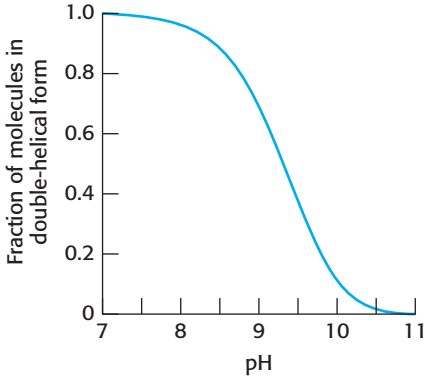
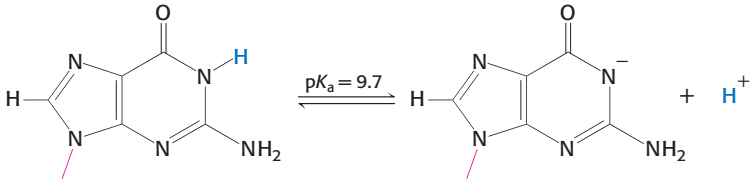


Figure 1.16 DNA denaturation by the addition of a base. The addition of a base to a solution of double-helical DNA initially at pH 7 causes the double helix to separate into single strands. The process is half complete at slightly above pH 9.

Acid–base reactions can disrupt the double helix

The reaction that we have been considering between two strands of DNA to form a double helix takes place readily at pH 7.0. Suppose that we take the solution containing the double-helical DNA and treat it with a solution of concentrated base (i.e., with a high concentration of OH^-). As the base is added, we monitor the pH and the fraction of DNA in double-helical form (Figure 1.16). When the first additions of base are made, the pH rises, but the concentration of the double-helical DNA does not change significantly. However, as the pH approaches 9, the DNA double helix begins to dissociate into its component single strands. As the pH continues to rise from 9 to 10, this dissociation becomes essentially complete. Why do the two strands dissociate? The hydroxide ions can react with bases in DNA base pairs to remove certain protons. The most susceptible proton is the one bound to the N-1 nitrogen atom in a guanine base.



Guanine (G)

Proton dissociation for a substance HA has an equilibrium constant defined by the expression

$$K_a = [H^+][A^-]/[HA]$$

The susceptibility of a proton to removal by reaction with a base is described by its *pK_a value:*

$$pK_a = -\log(K_a)$$

When the pH is equal to the pK_a, we have

$$pH = pK_a$$

and so

$$-\log[H^+] = -\log([H^+][A^-]/[HA])$$

and

$$[H^+] = [H^+][A^-]/[HA]$$

Dividing by $[H^+]$ reveals that

$$1 = [A^-]/[HA]$$

and so

$$[A^-] = [HA]$$

Thus, when the pH equals the pK_a, the concentration of the deprotonated form of the group or molecule is equal to the concentration of the protonated form; the deprotonation process is halfway to completion.

The pK_a for the proton on N-1 of guanine is typically 9.7. When the pH approaches this value, the proton on N-1 is lost (see Figure 1.16). Because this proton participates in an important hydrogen bond, its loss substantially destabilizes the DNA double helix. The DNA double helix is also destabilized by *low* pH. Below pH 5, some of the hydrogen bond *acceptors* that participate in base-pairing become protonated. In their protonated forms, these bases can no longer form hydrogen bonds and the double helix separates. Thus, acid–base reactions that remove or donate protons at specific positions on the DNA bases can disrupt the double helix.

Buffers regulate pH in organisms and in the laboratory

These observations about DNA reveal that a significant change in pH can disrupt molecular structure. The same is true for many other biological macromolecules; changes in pH can protonate or deprotonate key groups, potentially disrupting structures and initiating harmful reactions. Thus, systems have evolved to mitigate changes in pH in biological systems. Solutions that resist such changes are called *buffers*. Specifically, when acid is added to an unbuffered aqueous solution, the pH drops in proportion to the amount of acid added. In contrast, when acid is added to a buffered solution, the pH drops more gradually. Buffers also mitigate the pH increase caused by the addition of base and changes in pH caused by dilution.

Compare the result of adding a 1 M solution of the strong acid HCl drop by drop to pure water with adding it to a solution containing 100 mM of the buffer sodium acetate ($Na^+CH_3COO^-$; Figure 1.17). The process of gradually adding known amounts of reagent to a solution with which the reagent reacts while monitoring the results is called a *titration*. For pure water, the pH drops from 7 to close to 2 on the addition of the first few drops of acid. However, for the sodium acetate solution, the pH first falls rapidly from its initial value near 10, then changes more gradually until the pH

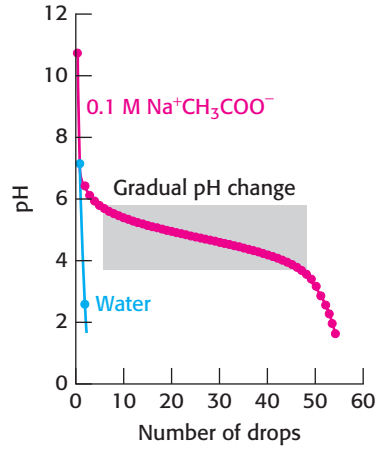


Figure 1.17 Buffer action. The addition of a strong acid, 1 M HCl, to pure water results in an immediate drop in pH to near 2. In contrast, the addition of the acid to a 0.1 M sodium acetate (Na^+ CH_3COO^-) solution results in a much more gradual change in pH until the pH drops below 3.5.

reaches 3.5, and then falls more rapidly again. Why does the pH decrease gradually in the middle of the titration? The answer is that, when hydrogen ions are added to this solution, they react with acetate ions to form acetic acid. This reaction consumes some of the added hydrogen ions so that the pH does not drop. Hydrogen ions continue reacting with acetate ions until essentially all of the acetate ion is converted into acetic acid. After this point, added protons remain free in solution and the pH begins to fall sharply again.

We can analyze the effect of the buffer in quantitative terms. The equilibrium constant for the deprotonation of an acid is

$$K_a = [H^+][A^-]/[HA]$$

Taking logarithms of both sides yields

$$\log(K_a) = \log([H^+]) + \log([A^-]/[HA])$$

Recalling the definitions of pK_a and pH and rearranging gives

$$pH = pK_a + \log([A^-]/[HA])$$

This expression is referred to as the *Henderson–Hasselbalch equation*.

We can apply the equation to our titration of sodium acetate. The pK_a of acetic acid is 4.75. We can calculate the ratio of the concentration of acetate ion to the concentration of acetic acid as a function of pH by using the Henderson–Hasselbalch equation, slightly rearranged.

$$[\text{Acetate ion}]/[\text{acetic acid}] = [A^-]/[HA] = 10^{pH-pKa}$$

At pH 9, this ratio is $10^{9-4.75} = 10^{4.25} = 17{,}800$; very little acetic acid has been formed. At pH 4.75 (when the pH equals the pK_a), the ratio is $10^{4.75-4.75} = 10^0 = 1$. At pH 3, the ratio is $10^{3-4.75} = 10^{-1.25} = 0.02$; almost all of the acetate ion has been converted into acetic acid. We can follow the conversion of acetate ion into acetic acid over the entire titration (Figure 1.18). The graph shows that the region of relatively constant pH corresponds precisely to the region in which acetate ion is being protonated to form acetic acid.

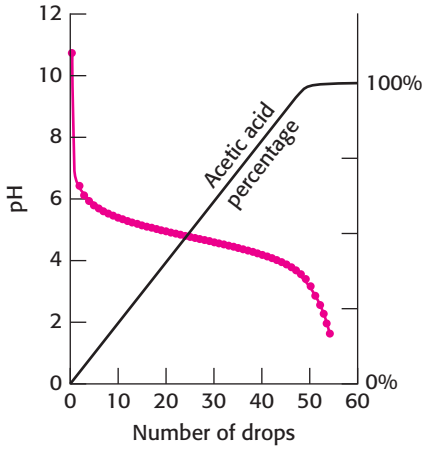


Figure 1.18 Buffer protonation. When acid is added to sodium acetate, the added hydrogen ions are used to convert acetate ion into acetic acid. Because the proton concentration does not increase significantly, the pH remains relatively constant until all of the acetate has been converted into acetic acid.

From this discussion, we see that a buffer functions best close to the pK_a value of its acid component. Physiological pH is typically about 7.4. An important buffer in biological systems is based on phosphoric acid (H_3PO_4). The acid can be deprotonated in three steps to form a phosphate ion.

$$H_3PO_4 \underset{pK_a = 2.12}{\overset{H^+}{\rightleftharpoons}} H_2PO_4^- \qquad H_2PO_4^- \underset{pK_a = 7.21}{\overset{H^+}{\rightleftharpoons}} HPO_4^{2-} \qquad HPO_4^{2-} \underset{pK_a = 12.67}{\overset{H^+}{\rightleftharpoons}} PO_4^{3-}$$

At about pH 7.4, inorganic phosphate exists primarily as a nearly equal mixture of $H_2PO_4^-$ and HPO_4^{2-}. Thus, phosphate solutions function as effective buffers near pH 7.4. The concentration of inorganic phosphate in blood is typically approximately 1 mM, providing a useful buffer against processes that produce either acid or base. We can examine this utility in quantitative terms with the use of the Henderson–Hasselbalch equation. What concentration of acid must be added to change the pH of 1 mM phosphate buffer from 7.4 to 7.3? Without buffer, this change in $[H^+]$ corresponds to a change of $10^{-7.3} - 10^{-7.4}$ M $= (5.0 \times 10^{-8} - 4.0 \times 10^{-8})$ M $= 1.0 \times 10^{-8}$ M. Let us now consider what happens to the buffer components. At pH 7.4,

$$[HPO_4^{2-}]/[H_2PO_4^-] = 10^{7.4-7.21} = 10^{0.19} = 1.55$$

The total concentration of phosphate is 1 mM, $[HPO_4^{2-}] + [H_2PO_4^-]$. Thus,

$$[HPO_4^{2-}] = (1.55/2.55) \times 1\text{ mM} = 0.608\text{ mM}$$

and

$$[H_2PO_4^-] = (1/2.55) \times 1\,\text{mM} = 0.392\,\text{mM}$$

At pH 7.3,

$$[HPO_4^{2-}]/[H_2PO_4^-] = 10^{7.3-7.21} = 10^{0.09} = 1.23$$

and so

$$[HPO_4^{2-}] = (1.23/2.23) = 0.552\,\text{mM}$$

and

$$[H_2PO_4^-] = (1/2.23) = 0.448\,\text{mM}$$

Thus, $(0.608 - 0.552) = 0.056$ mM HPO_4^{2-} is converted into $H_2PO_4^-$, consuming $0.056\,\text{mM} = 5.6 \times 10^{-5}$ M $[H^+]$. Thus, the buffer increases the amount of acid required to produce a drop in pH from 7.4 to 7.3 by a factor of $5.6 \times 10^{-5}/1.0 \times 10^{-8} = 5600$ compared with pure water.

1.4 The Genomic Revolution Is Transforming Biochemistry and Medicine

Watson and Crick's discovery of the structure of DNA suggested the hypothesis that hereditary information is stored as a sequence of bases along long strands of DNA. This remarkable insight provided an entirely new way of thinking about biology. However, at the time that it was made, Watson and Crick's discovery was full of potential but the practical consequences were unclear. Tremendously fundamental questions remained to be addressed. Is the hypothesis correct? How is the sequence information read and translated into action? What are the sequences of naturally occurring DNA molecules and how can such sequences be experimentally determined? Through advances in biochemistry and related sciences, we now have essentially complete answers to these questions. Indeed, in the past decade or so, scientists have determined the complete genome sequences of hundreds of different organisms, including simple microorganisms, plants, animals of varying degrees of complexity, and human beings. Comparisons of these genome sequences with the use of methods introduced in Chapter 6 have been sources of insight into many aspects of biochemistry. Because of these achievements, biochemistry has been transformed. In addition to its experimental and clinical aspects, biochemistry has now become an *information science*.

The sequencing of the human genome is a landmark in human history

The sequencing of the human genome was a daunting task because it contains approximately 3 billion (3×10^9) base pairs. For example, the sequence

ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTC
AAACAGACACCATGGTGCATCTGACTCCTG**A**GGAGAAGT
CTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGA . . .

is a part of one of the genes that encodes hemoglobin, the oxygen carrier in our blood. This gene is found on the end of chromosome 9 among our 24 distinct chromosomes. If we were to include the complete sequence of our entire genome, this chapter would run to more than 500,000 pages. The sequencing of our genome is truly a landmark in human history. This sequence contains a vast amount of information, some of which we can now

extract and interpret, but much of which we are only beginning to understand. For example, some human diseases have been linked to particular variations in genomic sequence. Sickle-cell anemia, discussed in detail in Chapter 7, is caused by a single base change of an A (noted in boldface type in the preceding sequence) to a T. We will encounter many other examples of diseases that have been linked to specific DNA sequence changes.

In addition to the implications for understanding human health and disease, the genome sequence is a source of deep insight into other aspects of human biology and culture. For example, by comparing the sequences of different individual persons and populations, we can learn a great deal about human history. On the basis of such analysis, a compelling case can be made that the human species originated in Africa, and the occurrence and even the timing of important migrations of groups of human beings can be demonstrated. Finally, comparisons of the human genome with the genomes of other organisms are confirming the tremendous unity that exists at the level of biochemistry and are revealing key steps that have been taken in the course of evolution from relatively simple, single-celled organisms to complex, multicellular organisms such as human beings. For example, many genes that are key to the function of the human brain and nervous system have evolutionary and functional relatives that can be recognized in the genomes of bacteria. Because many studies that are possible in model organisms are difficult or unethical to conduct in human beings, these discoveries have many practical implications. *Comparative genomics* has become a powerful science, linking evolution and biochemistry.

Genome sequences encode proteins and patterns of expression

The structure of DNA revealed how information is stored in the base sequence along a DNA strand. But what information is stored and how is this information expressed? The most fundamental role of DNA is to encode the sequences of proteins. Like DNA, proteins are linear polymers. However, proteins differ from DNA in two important ways. First, proteins are built from 20 building blocks, called *amino acids,* rather than just four, as in DNA. The chemical complexity provided by this variety of building blocks enables proteins to perform a wide range of functions. Second, proteins spontaneously fold up into elaborate three-dimensional structures, determined only by their amino acid sequences (Figure 1.19). We have explored in depth how solutions containing two appropriate strands of DNA come together to form a solution of double-helical molecules. A similar spontaneous folding process gives proteins their three-dimensional structure. A balance of hydrogen bonding, van der Waals interactions, and hydrophobic interactions overcome the entropy lost in going from an unfolded ensemble of proteins to a homogenous set of well-folded molecules. Proteins and protein folding will be discussed extensively in Chapter 2.

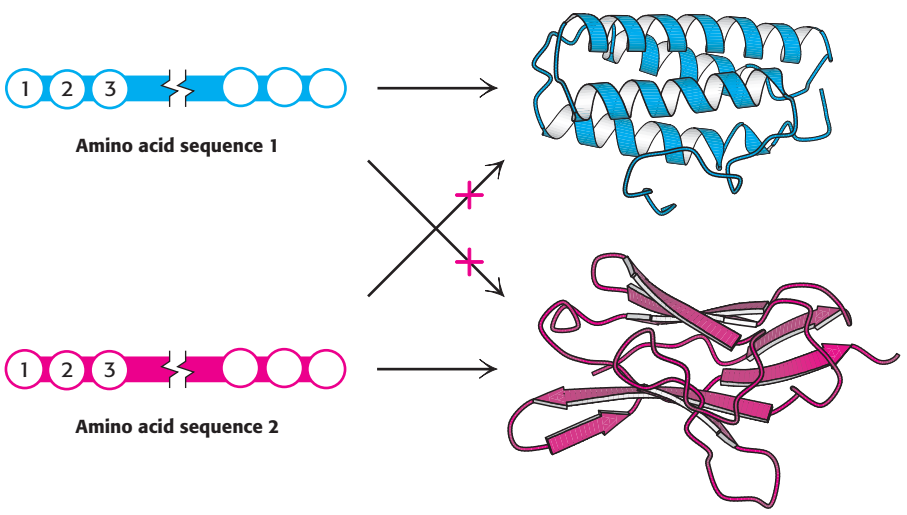


Figure 1.19 Protein folding. Proteins are linear polymers of amino acids that fold into elaborate structures. The sequence of amino acids determines the three-dimensional structure. Thus amino acid sequence 1 gives rise only to a protein with the shape depicted in blue, *not* the shape depicted in red.

The fundamental unit of hereditary information, the *gene,* is becoming increasingly difficult to precisely define as our knowledge of the complexities of genetics and genomics increases. The genes that are simplest to define encode the sequences of proteins. For these protein-encoding genes, a block of DNA bases encodes the amino acid sequence of a specific protein molecule. A set of three bases along the DNA strand, called a *codon,* determines the identity of one amino acid within the protein sequence. The relation that links the DNA sequence to the encoded protein sequence is called the *genetic code*. One of the biggest surprises from the sequencing of the human genome is the small number of protein-encoding genes. Before the genome-sequencing project began, the consensus view was that the human genome would include approximately 100,000 protein-encoding genes. The current analysis suggests that the actual number is between 20,000 and 25,000. We shall use an estimate of 23,000 throughout this book. However, additional mechanisms allow many genes to encode more than one protein. For example, the genetic information in some genes is translated in more than one way to produce a set of proteins that differ from one another in parts of their amino acid sequences. In other cases, proteins are modified after they have been synthesized through the addition of accessory chemical groups. Through these indirect mechanisms, much more complexity is encoded in our genomes than would be expected from the number of protein-encoding genes alone.

On the basis of current knowledge, the protein-encoding regions account for only about 3% of the human genome. What is the function of the rest of the DNA? Some of it contains information that regulates the expression of specific genes (i.e., the production of specific proteins) in particular cell types and physiological conditions. Essentially every cell contains the same DNA genome, yet cell types differ considerably in the proteins that they produce. For example, hemoglobin is expressed only in precursors of red blood cells, even though the genes for hemoglobin are present in essentially every cell. Specific sets of genes are expressed in response to hormones, even though these genes are not expressed in the same cell in the absence of the hormones. The control regions that regulate such differences account for only a small amount of the remainder of our genomes. The truth is that we do not yet understand all of the function of much of the remainder of the DNA. Some of it appears to be "junk," stretches of DNA that were inserted at some stage of evolution and have remained. In some cases, this DNA may, in fact, serve important functions. In others, it may serve no function but, because it does not cause significant harm, it has remained.

Individuality depends on the interplay between genes and environment

With the exception of monozygotic ("identical") twins, each person has a unique sequence of DNA base pairs. How different are we from one another at the genomic level? An examination of variation across the genome reveals that, on average, each pair of individual people has a different base in one position per 200 bases; that is, the difference is approximately 0.5%. This person-to-person variation is quite substantial compared with differences in populations. The average difference between two people within one ethnic group is greater than the difference between the averages of two different ethnic groups.

The significance of much of this genetic variation is not understood. As noted earlier, variation in a single base within the genome can lead to a disease such as sickle-cell anemia. Scientists have now identified the genetic variations associated with hundreds of diseases for which the cause can be traced

to a single gene. For other diseases and traits, we know that variation in many different genes contributes in significant and often complex ways. Many of the most prevalent human ailments such as heart disease are linked to variations in many genes. Furthermore, in most cases, the presence of a particular variation or set of variations does not inevitably result in the onset of a disease but, instead, leads to a *predisposition* to the development of the disease.

In addition to these genetic differences, *epigenetic factors* are important. These factors are associated with the genome but not simply represented in the sequence of DNA. For example, the consequences of some of this genetic variation depend, often dramatically, on whether the unusual gene sequence is inherited from the mother or from the father. This phenomenon, known as *genetic imprinting,* depends on the covalent modification of DNA, particularly the addition of methyl groups to particular bases. Epigenetics is a very active field of study and many novel discoveries can be expected.

Although our genetic makeup and associated epigenetic characteristics are important factors that contribute to disease susceptibility and to other traits, factors in a person's environment also are significant. What are these environmental factors? Perhaps the most obvious are chemicals that we eat or are exposed to in some other way. The adage "you are what you eat" has considerable validity; it applies both to substances that we ingest in significant quantities and to those that we ingest in only trace amounts. Throughout our study of biochemistry, we will encounter *vitamins* and *trace elements* and their derivatives that play crucial roles in many processes. In many cases, the roles of these chemicals were first revealed through investigation of *deficiency disorders* observed in people who do not take in a sufficient quantity of a particular vitamin or trace element. Despite the fact that the most important vitamins and trace elements have been known for some time, new roles for these essential dietary factors continue to be discovered.

A healthful diet requires a balance of major food groups (Figure 1.20). In addition to providing vitamins and trace elements, food provides calories in the form of substances that can be broken down to release energy to drive other biochemical processes. Proteins, fats, and carbohydrates provide the building blocks used to construct the molecules of life. Finally, it is possible to get too much of a good thing. Human beings evolved under circumstances in which food, particularly rich foods such as meat, was scarce. With the development of agriculture and modern economies, rich foods are now plentiful in parts of the world. Some of the most prevalent diseases in the so-called developed world, such as heart disease and diabetes, can be attributed to the large quantities of fats and carbohydrates that are present in modern diets. We are now developing a deeper understanding of the biochemical consequences of these diets and the interplay between diet and genetic factors.

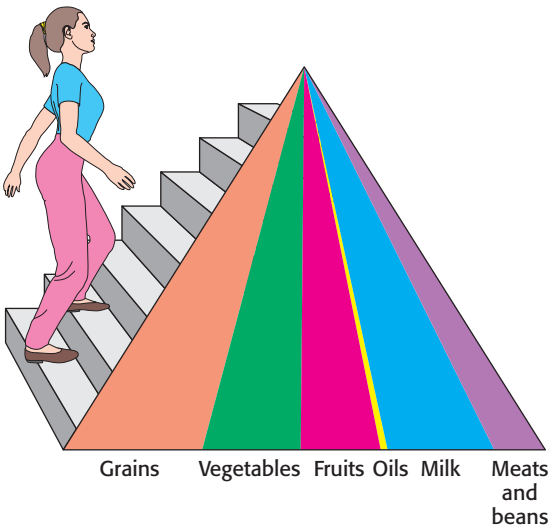


Figure 1.20 Food pyramid. A healthful diet includes a balance of food groups to supply an appropriate number of calories and an appropriate mixture of biochemical building blocks. [Courtesy of the U. S. Department of Agriculture.]

Chemicals are only one important class of environmental factors. The behaviors in which we engage also have biochemical consequences. Through physical activity, we consume the calories that we take in, ensuring an appropriate balance between food intake and energy expenditure. Activities ranging from exercise to emotional responses such as fear and love may activate specific biochemical pathways, leading to changes in levels of gene expression, the release of hormones, and other consequences. For example, recent discoveries reveal that high stress levels are associated with the shortening of telomeres, structures at the ends of chromosomes. Furthermore, the interplay between biochemistry and behavior is bidirectional. Just as our biochemistry is affected by our behavior, so, too, our behavior is affected, although certainly

not completely determined, by our genetic makeup and other aspects of our biochemistry. Genetic factors associated with a range of behavioral characteristics have been at least tentatively identified.

Just as vitamin deficiencies and genetic diseases revealed fundamental principles of biochemistry and biology, investigations of variations in behavior and their linkage to genetic and biochemical factors are potential sources of great insight into mechanisms within the brain. For example, studies of drug addiction have revealed neural circuits and biochemical pathways that greatly influence aspects of behavior. Unraveling the interplay between biology and behavior is one of the great challenges in modern science, and biochemistry is providing some of the most important concepts and tools for this endeavor.

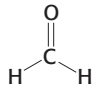
APPENDIX: Visualizing Molecular Structures I: Small Molecules

The authors of a biochemistry textbook face the problem of trying to present three-dimensional molecules in the two dimensions available on the printed page. The interplay between the three-dimensional structures of biomolecules and their biological functions will be discussed extensively throughout this book. Toward this end, we will frequently use representations that, although of necessity are rendered in two dimensions, emphasize the three-dimensional structures of molecules.

Stereochemical Renderings

Most of the chemical formulas in this book are drawn to depict the geometric arrangement of atoms, crucial to chemical bonding and reactivity, as accurately as possible. For example, the carbon atom of methane is tetrahedral, with H–C–H angles of 109.5 degrees, whereas the carbon atom in formaldehyde has bond angles of 120 degrees.

Methane

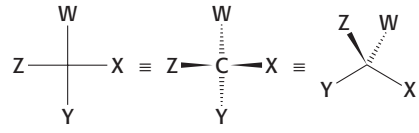


Formaldehyde

To illustrate the correct *stereochemistry* about tetrahedral carbon atoms, wedges will be used to depict the direction of a bond into or out of the plane of the page. A solid wedge with the broad end away from the carbon atom denotes a bond coming toward the viewer out of the plane. A dashed wedge, with its broad end at the carbon atom, represents a bond going away from the viewer behind the plane of the page. The remaining two bonds are depicted as straight lines.

Fischer Projections

Although representative of the actual structure of a compound, stereochemical structures are often difficult to draw quickly. An alternative, less-representative method of depicting structures with tetrahedral carbon centers relies on the use of *Fischer projections*.



Fischer projection

Stereochemical rendering

In a Fischer projection, the bonds to the central carbon are represented by horizontal and vertical lines from the substituent atoms to the carbon atom, which is assumed to be at the center of the cross. By convention, the horizontal bonds are assumed to project out of the page toward the viewer, whereas the vertical bonds are assumed to project behind the page away from the viewer.

Molecular Models for Small Molecules

For depicting the molecular architecture of small molecules in more detail, two types of models will often be used: space filling and ball and stick. These models show structures at the atomic level.

1. *Space-Filling Models*. The space-filling models are the most realistic. The size and position of an atom in a space-filling model are determined by its bonding properties and van der Waals radius, or contact distance. A van der Waals radius describes how closely two atoms can approach each other when they are not linked by a covalent bond. The colors of the model are set by convention.

Carbon, black	Hydrogen, white	Nitrogen, blue
Oxygen, red	Sulfur, yellow	Phosphorus, purple

Space-filling models of several simple molecules are shown in Figure 1.21.

2. *Ball-and-Stick Models*. Ball-and-stick models are not as realistic as space-filling models, because the atoms are depicted as spheres of radii smaller than their van der Waals radii. However, the bonding arrangement is easier to see because the bonds are explicitly represented as sticks. In an illustration, the taper of a stick, representing parallax, tells which of a pair of bonded atoms is closer to the reader. A ball-and-stick model reveals a complex structure more clearly than a space-filling model does. Ball-and-stick models of several simple molecules are shown in Figure 1.21.

Molecular models for depicting large molecules will be discussed in the appendix to Chapter 2.

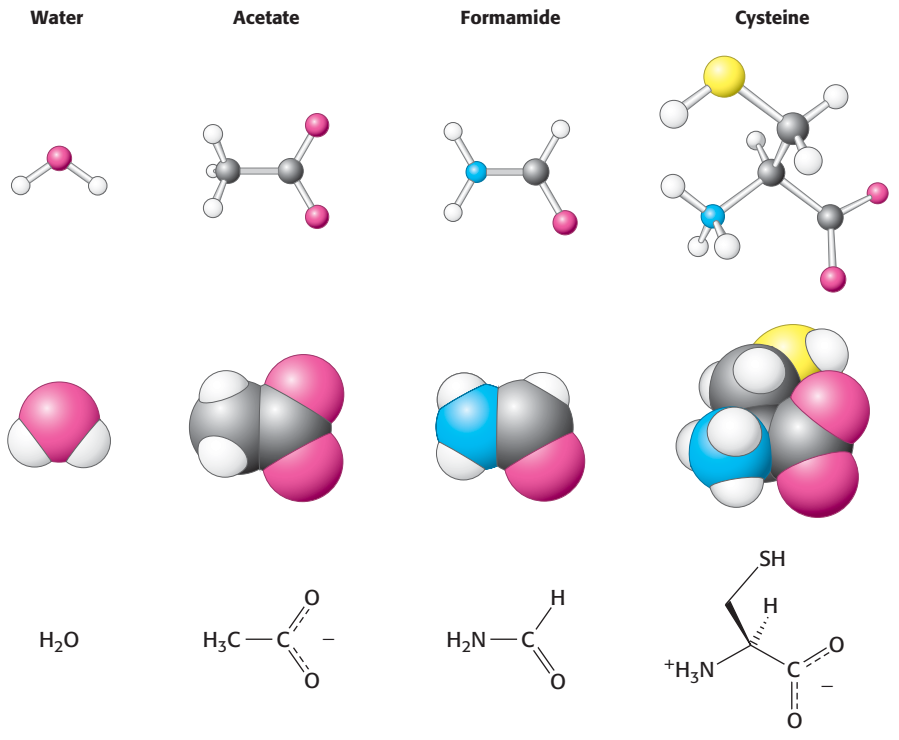


Figure 1.21 Molecular representations. Structural formulas (bottom), ball-and-stick models (middle), and space-filling representations (top) of selected molecules are shown. Black = carbon, red = oxygen, white = hydrogen, yellow = sulfur, blue = nitrogen.

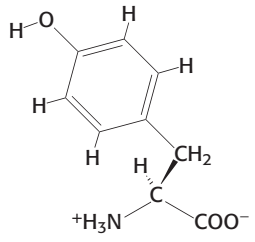
Key Terms

biological macromolecule (p. 2)
metabolite (p. 2)
deoxyribonucleic acid (DNA) (p. 2)
protein (p. 2)
Eukarya (p. 3)
Bacteria (p. 3)
Archaea (p. 3)
eukaryote (p. 3)
prokaryote (p. 3)
double helix (p. 5)
covalent bond (p. 5)
resonance structure (p. 7)
electrostatic interaction (p. 7)
hydrogen bond (p. 8)
van der Waals interaction (p. 8)
hydrophobic effect (p. 9)
hydrophobic interaction (p. 10)
entropy (p. 11)
enthalpy (p. 11)
free energy (Gibbs free energy) (p. 12)
pH (p. 13)
pK_a value (p. 15)
buffer (p. 15)
amino acid (p. 18)
genetic code (p. 19)
predisposition (p. 20)

Problems

1. *Donors and acceptors*. Identify the hydrogen-bond donors and acceptors in each of the four bases on page 4.

2. *Resonance structures*. The structure of an amino acid, tyrosine, is shown here. Draw an alternative resonance structure.



3. *It takes all types*. What types of noncovalent bonds hold together the following solids?

(a) Table salt (NaCl), which contains Na^+ and Cl^- ions.

(b) Graphite (C), which consists of sheets of covalently bonded carbon atoms.

4. *Don't break the law*. Given the following values for the changes in enthalpy (ΔH) and entropy (ΔS), which of the following processes can take place at 298 K without violating the Second Law of Thermodynamics?

(a) $\Delta H = -84$ kJ mol^{-1} (-20 kcal mol^{-1}), $\Delta S = +125$ J mol^{-1} K^{-1} ($+30$ cal mol^{-1}K^{-1})

(b) $\Delta H = -84$ kJ mol^{-1} (-20 kcal mol^{-1}), $\Delta S = -125$ J mol^{-1} K^{-1} (-30 cal mol^{-1} K^{-1})

(c) $\Delta H = +84$ kJ mol^{-1} ($+20$ kcal mol^{-1}), $\Delta S = -125$ J mol^{-1} K^{-1} ($+30$ cal mol^{-1} K^{-1})

(d) $\Delta H = +84$ kJ mol^{-1} ($+20$ kcal mol^{-1}), $\Delta S = -125$ J mol^{-1} K^{-1} (-30 cal mol^{-1} K^{-1})

5. *Double-helix-formation entropy*. For double-helix formation, ΔG can be measured to be -54 kJ mol^{-1} (-13 kcal mol^{-1}) at pH 7.0 in 1 M NaCl at 25°C (298 K). The heat released indicates an enthalpy change of -251 kJ mol^{-1} (-60 kcal mol^{-1}). For this process, calculate the entropy change for the system and the entropy change for the surroundings.

6. *Find the pH*. What are the pH values for the following solutions?

(a) 0.1 M HCl

(b) 0.1 M NaOH

(c) 0.05 M HCl

(d) 0.05 M NaOH

7. *A weak acid*. What is the pH of a 0.1 M solution of acetic acid ($pK_a = 4.75$)?

(Hint: Let x be the concentration of H^+ ions released from acetic acid when it dissociates. The solutions to a quadratic equation of the form $ax^2 + bx + c = 0$ are $x = (-b \pm \sqrt{b^2 - 4ac}) / 2a$.)

8. *Substituent effects*. What is the pH of a 0.1 M solution of chloroacetic acid ($ClCH_2COOH$, $pK_a = 2.86$)?

9. *Basic fact*. What is the pH of a 0.1 M solution of ethylamine, given that the pK_a of ethylammonium ion ($CH_3CH_2NH_3^+$) is 10.70?

10. *Comparison*. A solution is prepared by adding 0.01 M acetic acid and 0.01 M ethylamine to water and adjusting the pH to 7.4. What is the ratio of acetate to acetic acid? What is the ratio of ethylamine to ethylammonium ion?

11. *Concentrate*. Acetic acid is added to water until the pH value reaches 4.0. What is the total concentration of the added acetic acid?

12. *Dilution*. 100 mL of a solution of hydrochloric acid with pH 5.0 is diluted to 1 L. What is the pH of the diluted solution?

13. *Buffer dilution*. 100 mL of a 0.1 mM buffer solution made from acetic acid and sodium acetate with pH 5.0 is diluted to 1 L. What is the pH of the diluted solution?

14. *Find the pK_a*. For an acid HA, the concentrations of HA and A^- are 0.075 and 0.025, respectively, at pH 6.0. What is the pK_a value for HA?

15. *pH indicator*. A dye that is an acid and that appears as different colors in its protonated and deprotonated forms can be used as a pH indicator. Suppose that you have a 0.001 M solution of a dye with a pK_a of 7.2. From the color, the concentration of the protonated form is found to be 0.0002 M. Assume that the remainder of the dye is in the deprotonated form. What is the pH of the solution?

16. *What's the ratio?* An acid with a pK_a of 8.0 is present in a solution with a pH of 6.0. What is the ratio of the protonated to the deprotonated form of the acid?

17. *Phosphate buffer*. What is the ratio of the concentrations of $H_2PO_4^-$ and HPO_4^{2-} at (a) pH 7.0; (b) pH 7.5; (c) pH 8.0?

18. *Buffer capacity*. Two solutions of sodium acetate are prepared, one with a concentration of 0.1 M and the other with a concentration of 0.01 M. Calculate the pH values when the following concentrations of HCl have been added to each of these solutions: 0.0025 M, 0.005 M, 0.01 M, and 0.05 M.

19. *Buffer preparation*. You wish to prepare a buffer consisting of acetic acid and sodium acetate with a total acetic acid plus acetate concentration of 250 mM and a pH of 5.0. What concentrations of acetic acid and sodium acetate should you use? Assuming you wish to make 2 liters of this buffer, how many moles of acetic acid and sodium acetate will you need? How many grams of each will you need (molecular weights: acetic acid 60.05 g mol^{-1}, sodium acetate, 82.03 g mol^{-1})?

20. *An alternative approach*. When you go to prepare the buffer described in Problem 19, you discover that your laboratory is out of sodium acetate, but you do have sodium hydroxide. How much (in moles and grams) acetic acid and sodium hydroxide do you need to make the buffer?

21. *Another alternative*. Your friend from another laboratory was out of acetic acid so he tries to prepare the buffer in Problem 19 by dissolving 41.02 g of sodium acetate in water, carefully adding 180.0 ml of 1 M HCl, and adding more water to reach a total volume of 2 liters. What is the total concentration of acetate plus acetic acid in the solution? Will this solution have pH 5.0? Will it be identical with the desired buffer? If not, how will it differ?

22. *Blood substitute*. As noted in this chapter, blood contains a total concentration of phosphate of approximately 1 mM and typically has a pH of 7.4. You wish to make 100 liters of phosphate buffer with a pH of 7.4 from NaH_2PO_4 (molecular weight, 119.98 g mol^{-1}) and Na_2HPO_4 (molecular weight, 141.96 g mol^{-1}). How much of each (in grams) do you need?

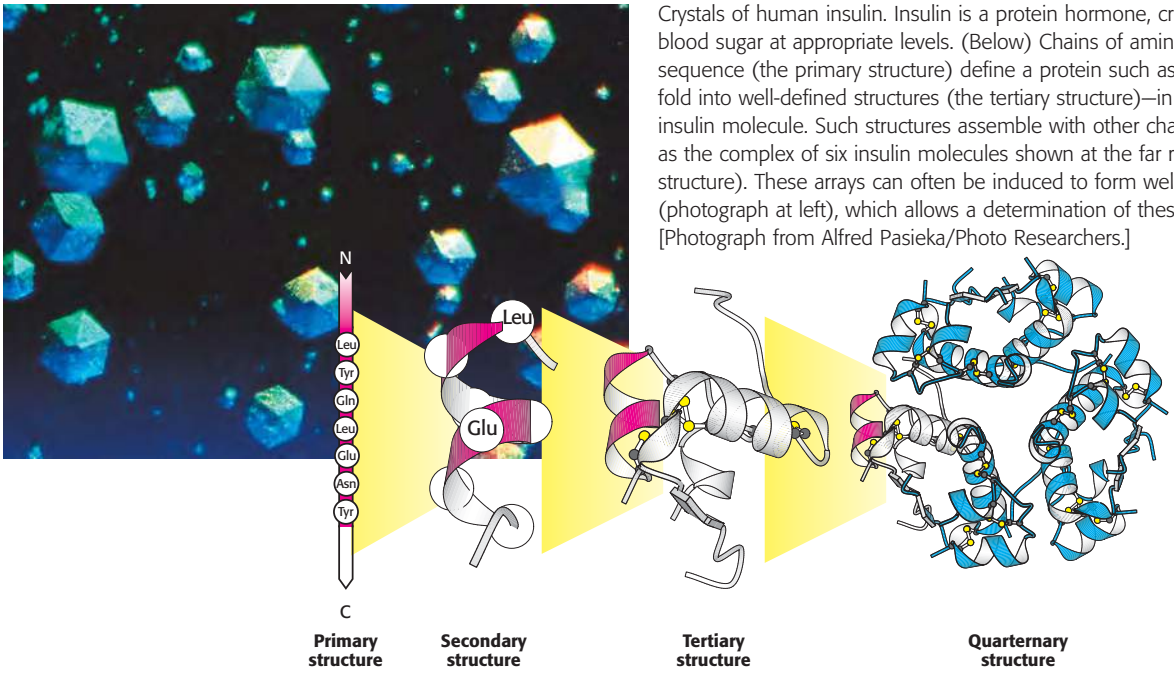
23. *A potential problem*. You wish to make a buffer with pH 7.0. You combine 0.060 grams of acetic acid and 14.59 grams of sodium acetate and add water to yield a total volume of 1 liter. What is the pH? Will this be the useful pH 7.0 buffer you seek?

24. *Charge!* Suppose two phosphate groups in DNA (each with a charge of -1) are separated by 12 Å. What is the energy of the electrostatic interaction between these two phosphates assuming a dielectric constant of 80? Repeat the calculation assuming a dielectric constant of 2.

25. *Viva la différence*. On average, how many base differences are there between two human beings?

CHAPTER 2

Protein Composition and Structure



Crystals of human insulin. Insulin is a protein hormone, crucial for maintaining blood sugar at appropriate levels. (Below) Chains of amino acids in a specific sequence (the primary structure) define a protein such as insulin. These chains fold into well-defined structures (the tertiary structure)—in this case, a single insulin molecule. Such structures assemble with other chains to form arrays such as the complex of six insulin molecules shown at the far right (the quarternary structure). These arrays can often be induced to form well-defined crystals (photograph at left), which allows a determination of these structures in detail. [Photograph from Alfred Pasieka/Photo Researchers.]

Proteins are the most versatile macromolecules in living systems and serve crucial functions in essentially all biological processes. They function as catalysts, transport and store other molecules such as oxygen, provide mechanical support and immune protection, generate movement, transmit nerve impulses, and control growth and differentiation. Indeed, much of this book will focus on understanding what proteins do and how they perform these functions.

Several key properties enable proteins to participate in a wide range of functions.

1. *Proteins are linear polymers built of monomer units called amino acids,* which are linked end to end. The sequence of linked amino acids is called the primary structure. Remarkably, proteins spontaneously fold up into three-dimensional structures that are determined by the sequence of amino acids in the protein polymer. Three-dimensional structure formed by hydrogen bonds between amino acids near one another is called secondary structure, whereas tertiary structure is formed by long-range interactions between amino acids. Protein function depends directly on this three-dimensional structure (Figure 2.1). Thus, *proteins are the embodiment of the transition from the one-dimensional world of sequences to the three-dimensional world of molecules capable of diverse activities*. Many proteins display

OUTLINE

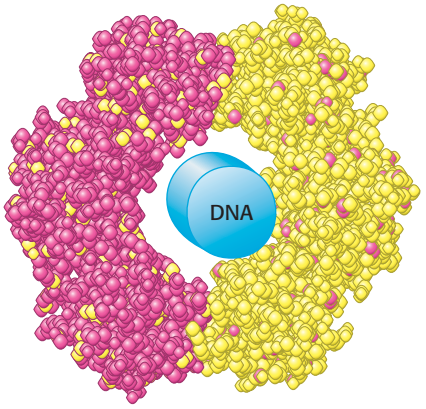


Figure 2.1 Structure dictates function. A protein component of the DNA replication machinery surrounds a section of DNA double helix depicted as a cylinder. The protein, which consists of two identical subunits (shown in red and yellow), acts as a clamp that allows large segments of DNA to be copied without the replication machinery dissociating from the DNA. [Drawn from 2POL.pdb.]

quaternary structure, in which the functional protein is composed of several distinct polypeptide chains.

2. *Proteins contain a wide range of functional groups*. These functional groups include alcohols, thiols, thioethers, carboxylic acids, carboxamides, and a variety of basic groups. Most of these groups are chemically reactive. When combined in various sequences, this array of functional groups accounts for the broad spectrum of protein function. For instance, their reactive properties are essential to the function of *enzymes,* the proteins that catalyze specific chemical reactions in biological systems (see Chapters 8 through 10).

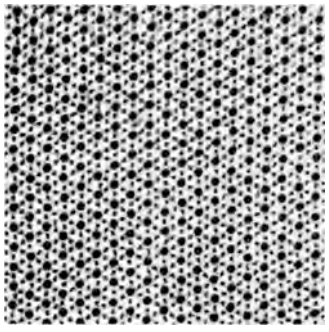


Figure 2.2 A complex protein assembly. An electron micrograph of insect flight tissue in cross section shows a hexagonal array of two kinds of protein filaments. [Courtesy of Dr. Michael Reedy.]

3. *Proteins can interact with one another and with other biological macromolecules to form complex assemblies*. The proteins within these assemblies can act synergistically to generate capabilities that individual proteins may lack (Figure 2.2). Examples of these assemblies include macromolecular machines that replicate DNA, transmit signals within cells, and carry out many other essential processes.

4. *Some proteins are quite rigid, whereas others display a considerable flexibility*. Rigid units can function as structural elements in the cytoskeleton (the internal scaffolding within cells) or in connective tissue. Proteins with some flexibility may act as hinges, springs, or levers that are crucial to protein

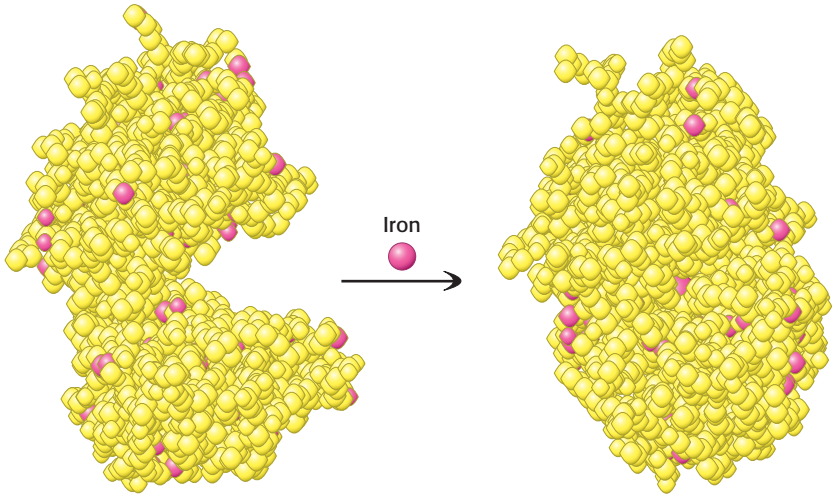


Figure 2.3 Flexibility and function. On binding iron, the protein lactoferrin undergoes a substantial change in conformation that allows other molecules to distinguish between the iron-free and the iron-bound forms. [Drawn from 1 LFH.pdb and 1 LFG.pdb.]

function, to the assembly of proteins with one another and with other molecules into complex units, and to the transmission of information within and between cells (Figure 2.3).

2.1 Proteins Are Built from a Repertoire of 20 Amino Acids

Amino acids are the building blocks of proteins. An *α-amino acid* consists of a central carbon atom, called the *α carbon,* linked to an amino group, a carboxylic acid group, a hydrogen atom, and a distinctive R group. The R group is often referred to as the *side chain.* With four different groups connected to the tetrahedral α-carbon atom, α-amino acids are *chiral:* they may exist in one or the other of two mirror-image forms, called the L isomer and the D isomer (Figure 2.4).

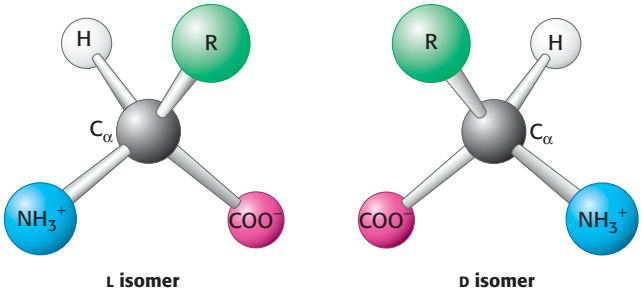


Figure 2.4 The L and D isomers of amino acids. The letter R refers to the side chain. The L and D isomers are mirror images of each other.

Notation for distinguishing stereoisomers

The four different substituents of an asymmetric carbon atom are assigned a priority according to atomic number. The lowest-priority substituent, often hydrogen, is pointed away from the viewer. The configuration about the carbon atom is called *S* (from the Latin *sinister*, "left") if the progression from the highest to the lowest priority is counterclockwise. The configuration is called *R* (from the Latin *rectus*, "right") if the progression is clockwise.

Only L amino acids are constituents of proteins. For almost all amino acids, the L isomer has *S* (rather than *R*) absolute configuration (Figure 2.5). What is the basis for the preference for L amino acids? The answer is not known, but evidence shows that L amino acids are slightly more soluble than is a racemic mixture of D and L amino acids, which tend to form crystals. This small solubility difference could have been amplified over time so that the L isomer became dominant in solution.

Amino acids in solution at neutral pH exist predominantly as *dipolar ions* (also called *zwitterions*). In the dipolar form, the amino group is protonated ($—NH_3^+$) and the carboxyl group is deprotonated ($—COO^-$). The ionization state of an amino acid varies with pH (Figure 2.6). In acid

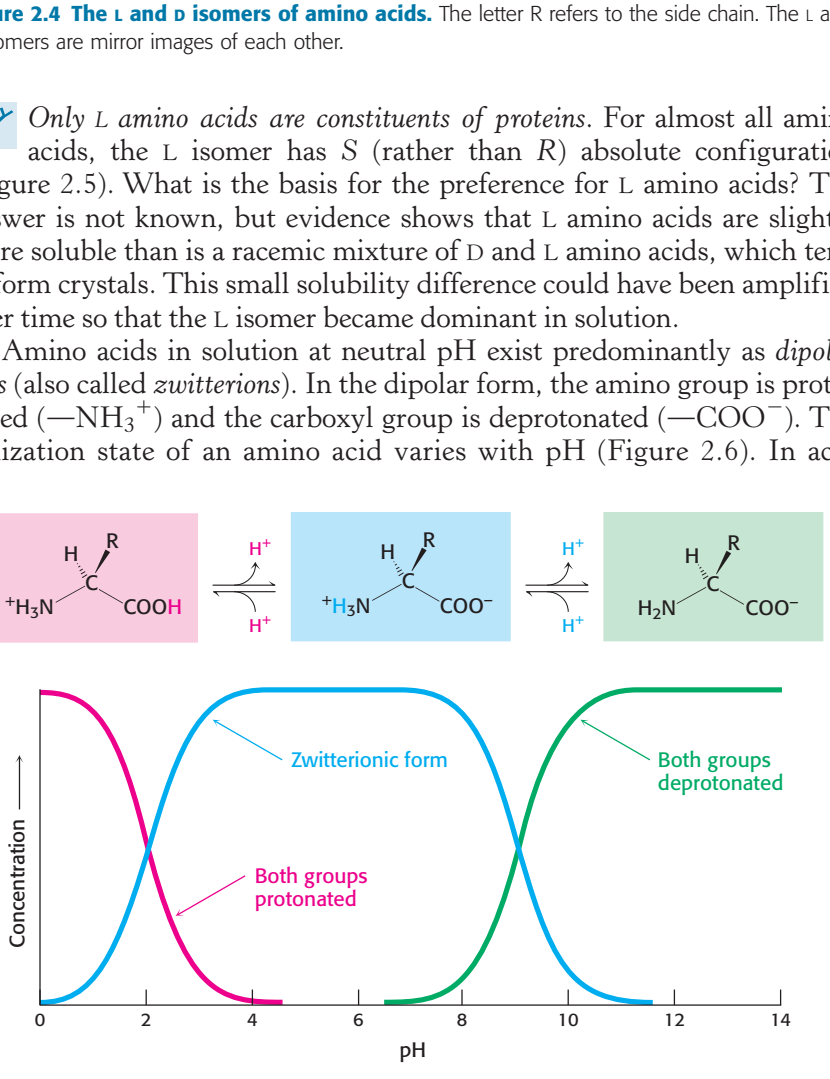


Figure 2.5 Only L amino acids are found in proteins. Almost all L amino acids have an *S* absolute configuration. The counterclockwise direction of the arrow from highest- to lowest-priority substituents indicates that the chiral center is of the *S* configuration.

Figure 2.6 Ionization state as a function of pH. The ionization state of amino acids is altered by a change in pH. The zwitterionic form predominates near physiological pH.

solution (e.g., pH 1), the amino group is protonated ($—NH_3^+$) and the carboxyl group is not dissociated (—COOH). As the pH is raised, the carboxylic acid is the first group to give up a proton, inasmuch as its pK_a is near 2. The dipolar form persists until the pH approaches 9, when the protonated amino group loses a proton.

Twenty kinds of side chains varying in *size, shape, charge, hydrogen-bonding capacity, hydrophobic character,* and *chemical reactivity* are commonly found in proteins. Indeed, all proteins in all species—bacterial, archaeal, and eukaryotic—are constructed from the same set of 20 amino acids with only a few exceptions. This fundamental alphabet for the construction of proteins is several billion years old. The remarkable range of functions mediated by proteins results from the diversity and versatility of these 20 building blocks. Understanding how this alphabet is used to create the intricate three-dimensional structures that enable proteins to carry out so many biological processes is an exciting area of biochemistry and one that we will return to in Section 2.6.

Although there are many ways to classify amino acids, we will assort these molecules into four groups, on the basis of the general chemical characteristics of their R groups:

1. Hydrophobic amino acids with nonpolar R groups
2. Polar amino acids with neutral R groups but the charge is not evenly distributed
3. Positively charged amino acids with R groups that have a positive charge at physiological pH
4. Negatively charged amino acids with R groups that have a negative charge at physiological pH

Hydrophobic amino acids The simplest amino acid is *glycine,* which has a single hydrogen atom as its side chain. With two hydrogen atoms bonded to the α-carbon atom, glycine is unique in being *achiral*. *Alanine,* the next simplest amino acid, has a methyl group ($—CH_3$) as its side chain (Figure 2.7).

Larger hydrocarbon side chains are found in *valine, leucine,* and *isoleucine*. *Methionine* contains a largely aliphatic side chain that includes a *thioether* (—S—) group. The side chain of isoleucine includes an additional chiral center; only the isomer shown in Figure 2.7 is found in proteins. The larger aliphatic side chains are especially hydrophobic; that is, they tend to cluster together rather than contact water. The three-dimensional structures of water-soluble proteins are stabilized by this tendency of hydrophobic groups to come together, which is called *the hydrophobic effect* (Chapter 1). The different sizes and shapes of these hydrocarbon side chains enable them to pack together to form compact structures with little empty space. *Proline* also has an aliphatic side chain, but it differs from other members of the set of 20 in that its side chain is bonded to both the nitrogen and the α-carbon atoms. Proline markedly influences protein architecture because its ring structure makes it more conformationally restricted than the other amino acids.

Two amino acids with relatively simple *aromatic side chains* are part of the fundamental repertoire. *Phenylalanine,* as its name indicates, contains a phenyl ring attached in place of one of the hydrogen atoms of alanine. *Tryptophan* has an indole group joined to a methylene ($—CH_2—$) group; the indole group comprises two fused rings containing an NH group. Phenylalanine is purely hydrophobic, whereas tryptophan is less so because of its NH groups.

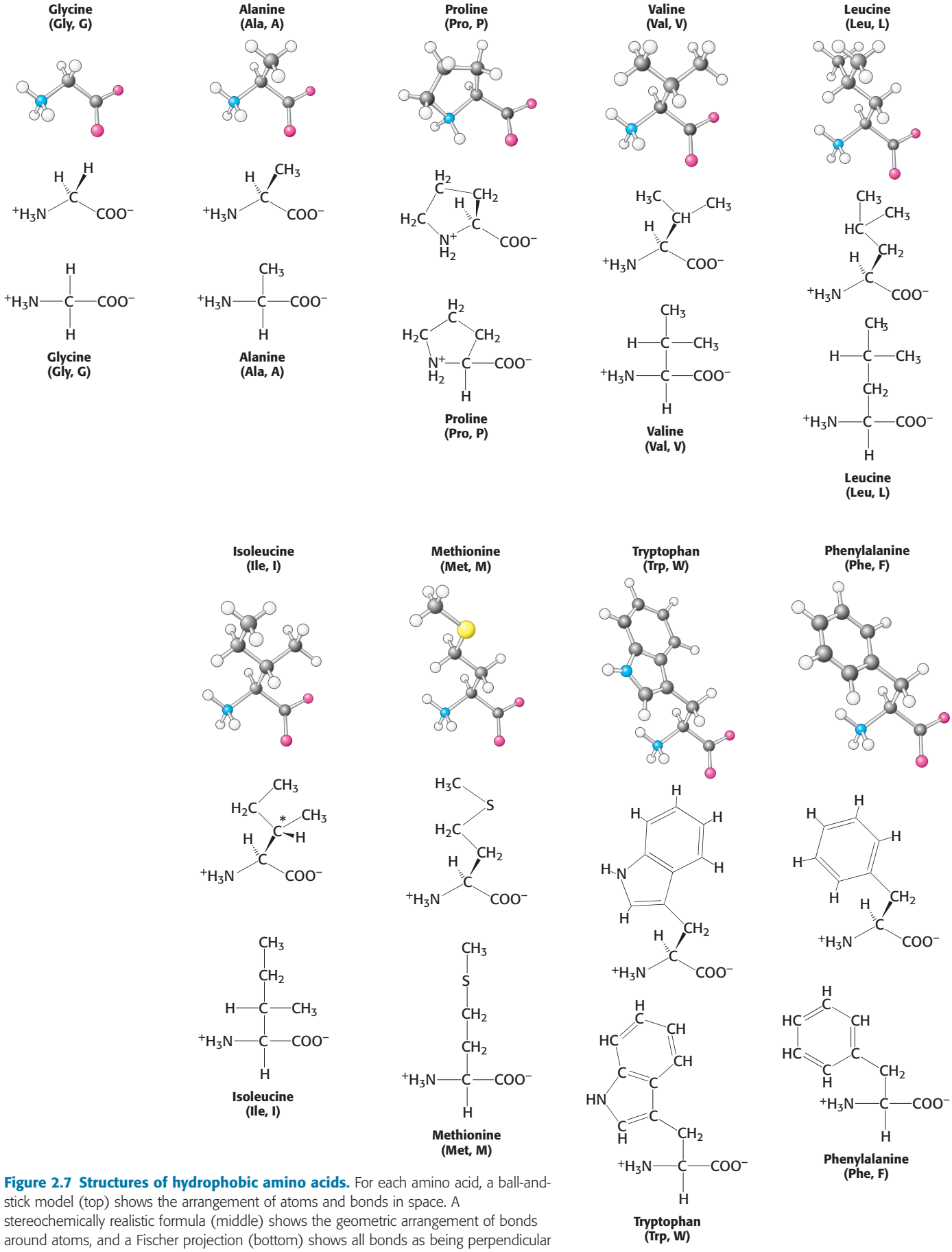


Figure 2.7 Structures of hydrophobic amino acids. For each amino acid, a ball-and-stick model (top) shows the arrangement of atoms and bonds in space. A stereochemically realistic formula (middle) shows the geometric arrangement of bonds around atoms, and a Fischer projection (bottom) shows all bonds as being perpendicular for a simplified representation (see the Appendix to Chapter 1).

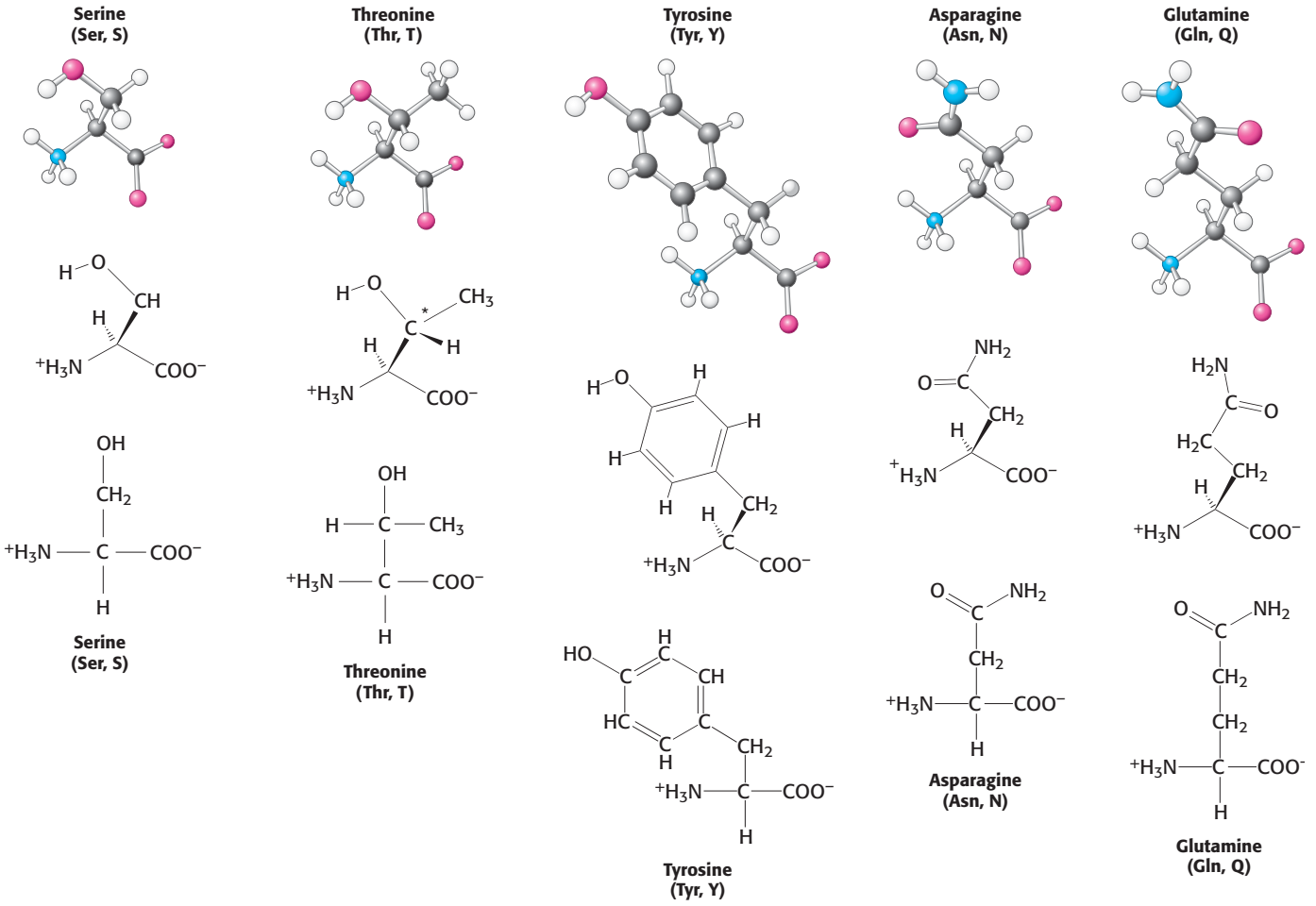


Figure 2.8 Structures of the polar amino acids. The additional chiral center in threonine is indicated by an asterisk.

Polar amino acids Six amino acids are polar but uncharged. Three amino acids, *serine, threonine,* and *tyrosine,* contain *hydroxyl groups* (—OH) attached to a hydrophobic side chain (Figure 2.8). Serine can be thought of as a version of alanine with a hydroxyl group attached, threonine resembles valine with a hydroxyl group in place of one of valine's methyl groups, and tyrosine is a version of phenylalanine with the hydroxyl group replacing a hydrogen atom on the aromatic ring. The hydroxyl group makes these amino acids much more *hydrophilic* (water loving) and *reactive* than their hydrophobic analogs. Threonine, like isoleucine, contains an additional asymmetric center; again, only one isomer is present in proteins.

In addition, the set includes *asparagine* and *glutamine*, uncharged derivatives of the acidic amino acids aspartate and glutamate (see Figure 2.11). Each of these two amino acids contains a terminal *carboxamide* in place of a carboxylic acid. The side chain of glutamine is one methylene group longer than that of asparagine.

Cysteine is structurally similar to serine but contains a *sulfhydryl,* or *thiol* (—SH), group in place of the hydroxyl (—OH) group. The sulfhydryl group is much more reactive. Pairs of sulfhydryl groups may come together to form disulfide bonds, which are particularly important in stabilizing some proteins, as will be discussed shortly.

Positively charged amino acids We turn now to amino acids with complete positive charges that render them highly hydrophilic. *Lysine* and *arginine*

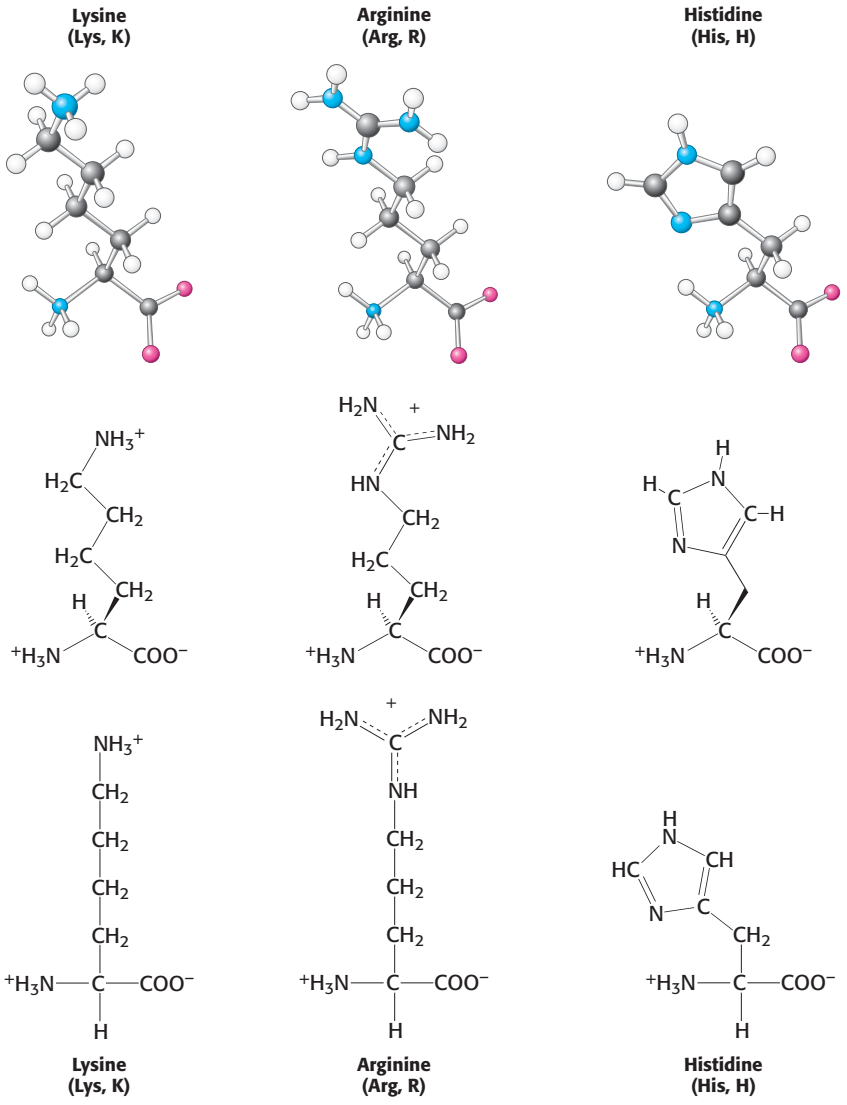
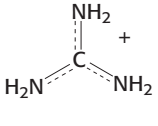
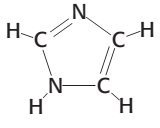


Figure 2.9 Positively charged amino acids lysine, arginine, and histidine.

have long side chains that terminate with groups that are *positively charged* at neutral pH. Lysine is capped by a primary amino group and arginine by a guanidinium group. *Histidine* contains an imidazole group, an aromatic ring that also can be positively charged (Figure 2.9).

With a pK_a value near 6, the imidazole group can be uncharged or positively charged near neutral pH, depending on its local environment (Figure 2.10). Histidine is often found in the active sites of enzymes, where the imidazole ring can bind and release protons in the course of enzymatic reactions.





Negatively charged amino acids This set of amino acids contains two with *acidic side chains: aspartic acid* and *glutamic acid* (Figure 2.11). These amino acids are often called *aspartate* and *glutamate* to emphasize that, at physiological pH, their side chains usually lack a proton that is present in the acid form and hence are negatively charged. Nonetheless, in some proteins, these side chains do accept protons, and this ability is often functionally important.

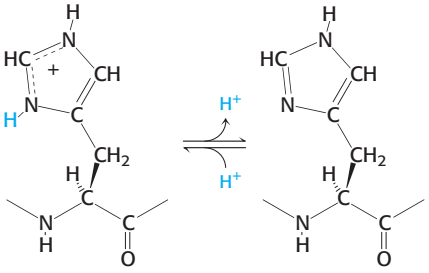


Figure 2.10 Histidine ionization. Histidine can bind or release protons near physiological pH.

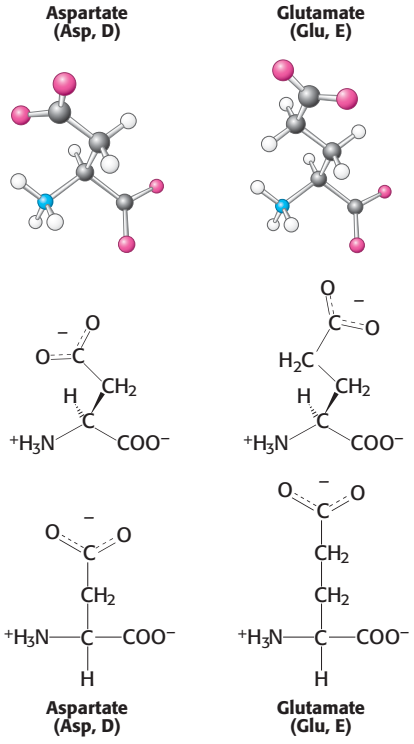
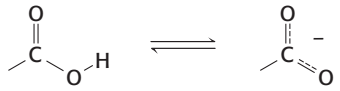
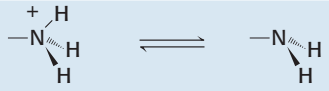
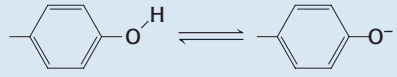
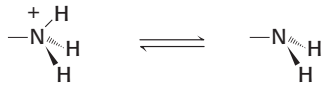
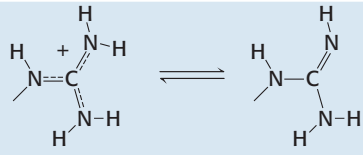


Figure 2.11 Negatively charged amino acids.

Table 2.1 Typical pK_a values of ionizable groups in proteins

Group	Acid ⇌ Base	Typical pK_a*
Terminal α-carboxyl group		3.1
Aspartic acid Glutamic acid		4.1
Histidine		6.0
Terminal α-amino group		8.0
Cysteine		8.3
Tyrosine		10.9
Lysine		10.8
Arginine		12.5

*pK_a values depend on temperature, ionic strength, and the microenvironment of the ionizable group.

Seven of the 20 amino acids have readily ionizable side chains. These 7 amino acids are able to donate or accept protons to facilitate reactions as well as to form ionic bonds. Table 2.1 gives equilibria and typical pK_a values for ionization of the side chains of tyrosine, cysteine, arginine, lysine, histidine, and aspartic and glutamic acids in proteins. Two other groups in proteins—the terminal α-amino group and the terminal α-carboxyl group—can be ionized, and typical pK_a values for these groups also are included in Table 2.1.

Amino acids are often designated by either a three-letter abbreviation or a one-letter symbol (Table 2.2). The abbreviations for amino acids are the first

Table 2.2 Abbreviations for amino acids

Amino acid	Three-letter abbreviation	One-letter abbreviation	Amino acid	Three-letter abbreviation	One-letter abbreviation
Alanine	Ala	A	Methionine	Met	M
Arginine	Arg	R	Phenylalanine	Phe	F
Asparagine	Asn	N	Proline	Pro	P
Aspartic acid	Asp	D	Serine	Ser	S
Cysteine	Cys	C	Threonine	Thr	T
Glutamine	Gln	Q	Tryptophan	Trp	W
Glutamic acid	Glu	E	Tyrosine	Tyr	Y
Glycine	Gly	G	Valine	Val	V
Histidine	His	H	Asparagine or aspartic acid	Asx	B
Isoleucine	Ile	I			
Leucine	Leu	L	Glutamine or glutamic acid	Glx	Z
Lysine	Lys	K			

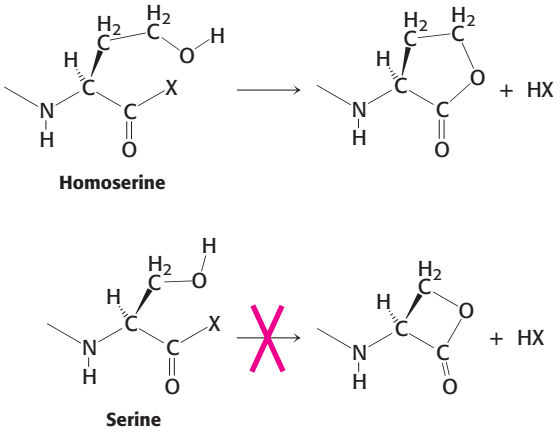


Figure 2.12 Undesirable reactivity in amino acids. Some amino acids are unsuitable for proteins because of undesirable cyclization. Homoserine can cyclize to form a stable, five-membered ring, potentially resulting in peptide-bond cleavage. The cyclization of serine would form a strained, four-membered ring and is thus disfavored. X can be an amino group from a neighboring amino acid or another potential leaving group.

three letters of their names, except for asparagine (Asn), glutamine (Gln), isoleucine (Ile), and tryptophan (Trp). The symbols for many amino acids are the first letters of their names (e.g., G for glycine and L for leucine); the other symbols have been agreed on by convention. These abbreviations and symbols are an integral part of the vocabulary of biochemists.

How did this particular set of amino acids become the building blocks of proteins? First, as a set, they are diverse: their structural and chemical properties span a wide range, endowing proteins with the versatility to assume many functional roles. Second, many of these amino acids were probably available from prebiotic reactions; that is, from reactions that took place before the origin of life. Finally, other possible amino acids may have simply been too reactive. For example, amino acids such as homoserine and homocysteine tend to form five-membered cyclic forms that limit their use in proteins; the alternative amino acids that are found in proteins—serine and cysteine—do not readily cyclize, because the rings in their cyclic forms are too small (Figure 2.12).

2.2 Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains

Proteins are *linear polymers* formed by linking the α-carboxyl group of one amino acid to the α-amino group of another amino acid. This type of linkage is called a *peptide bond* or an *amide bond*. The formation of a dipeptide from two amino acids is accompanied by the loss of a water molecule (Figure 2.13). The equilibrium of this reaction lies on the side of hydrolysis rather than synthesis under most conditions. Hence, the biosynthesis of peptide bonds requires an input of free energy. Nonetheless, peptide bonds are quite *stable kinetically* because the rate of hydrolysis is extremely slow; the lifetime of a peptide bond in aqueous solution in the absence of a catalyst approaches 1000 years.

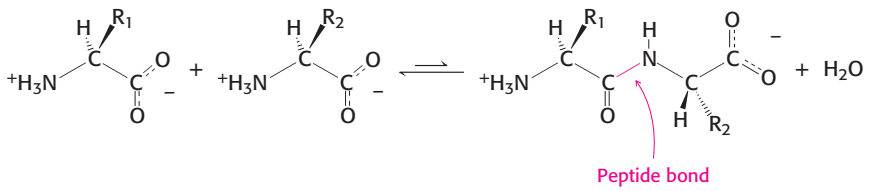


Figure 2.13 Peptide-bond formation. The linking of two amino acids is accompanied by the loss of a molecule of water.

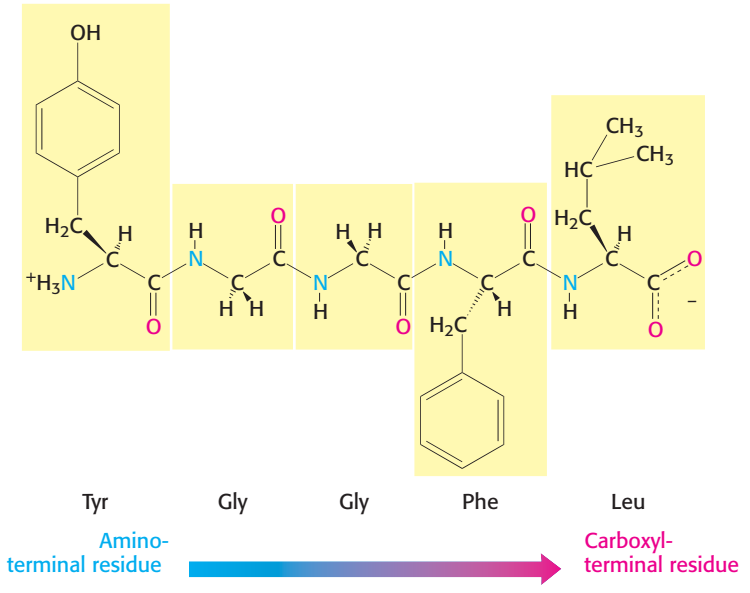


Figure 2.14 Amino acid sequences have direction. This illustration of the pentapeptide Try-Gly-Gly-Phe-Leu (YGGFL) shows the sequence from the amino terminus to the carboxyl terminus. This pentapeptide, Leu-enkephalin, is an opioid peptide that modulates the perception of pain. The reverse pentapeptide, Leu-Phe-Gly-Gly-Tyr (LFGGY), is a different molecule and has no such effects.

A series of amino acids joined by peptide bonds form a *polypeptide chain,* and each amino acid unit in a polypeptide is called a *residue. A polypeptide chain has polarity* because its ends are different: an α-amino group is present at one end and an α-carboxyl group at the other. By convention, *the amino end is taken to be the beginning of a polypeptide chain,* and so the sequence of amino acids in a polypeptide chain is written starting with the amino-terminal residue. Thus, in the pentapeptide Tyr-Gly-Gly-Phe-Leu (YGGFL), tyrosine is the amino-terminal (N-terminal) residue and leucine is the carboxyl-terminal (C-terminal) residue (Figure 2.14). Leu-Phe-Gly-Gly-Tyr (LFGGY) is a different pentapeptide, with different chemical properties.

A polypeptide chain consists of a regularly repeating part, called the *main chain* or *backbone,* and a variable part, comprising the distinctive *side chains* (Figure 2.15). The polypeptide backbone is rich in hydrogen-bonding potential. Each residue contains a carbonyl group (C═O), which is a good hydrogen-bond acceptor, and, with the exception of proline, an NH group, which is a good hydrogen-bond donor. These groups interact with each other and with functional groups from side chains to stabilize particular structures, as will be discussed in Section 2.3.

Most natural polypeptide chains contain between 50 and 2000 amino acid residues and are commonly referred to as *proteins*. The largest protein known is the muscle protein *titin*, which consists of more than 27,000 amino acids. Peptides made of small numbers of amino acids are called *oligopeptides* or simply *peptides*. The mean molecular weight of an amino acid residue is about 110 g mol^{-1}, and so the molecular weights of most proteins are between 5500 and 220,000 g mol^{-1}. We can also refer to the mass of a protein, which is expressed in units of daltons; one *dalton* is equal to one atomic mass unit. A protein with a molecular weight of 50,000 g mol^{-1} has a mass of 50,000 daltons, or 50 kd (kilodaltons).

Dalton
A unit of mass very nearly equal to that of a hydrogen atom. Named after John Dalton (1766–1844), who developed the atomic theory of matter.

Kilodalton (kd)
A unit of mass equal to 1000 daltons

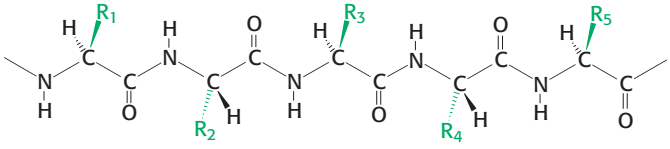


Figure 2.15 Components of a polypeptide chain. A polypeptide chain consists of a constant backbone (shown in black) and variable side chains (shown in green).

In some proteins, the linear polypeptide chain is cross-linked. The most common cross-links are *disulfide bonds,* formed by the oxidation of a pair of cysteine residues (Figure 2.16). The resulting unit of two linked cysteines is called *cystine*. Extracellular proteins often have several disulfide bonds, whereas intracellular proteins usually lack them. Rarely, nondisulfide cross-links derived from other side chains are present in proteins. For example, collagen fibers in connective tissue are strengthened in this way, as are fibrin blood clots.

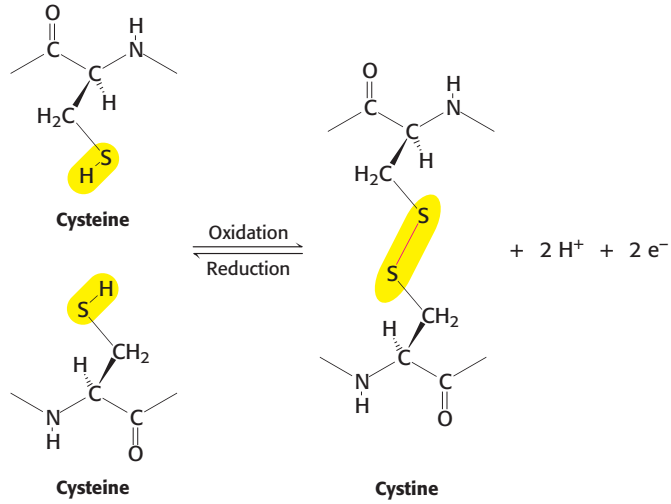


Figure 2.16 Cross-links. The formation of a disulfide bond from two cysteine residues is an oxidation reaction.

Proteins have unique amino acid sequences specified by genes

In 1953, Frederick Sanger determined the amino acid sequence of insulin, a protein hormone (Figure 2.17). *This work is a landmark in biochemistry because it showed for the first time that a protein has a precisely defined amino acid sequence* consisting only of L amino acids linked by peptide bonds. This accomplishment stimulated other scientists to carry out sequence studies of a wide variety of proteins. Currently, the complete amino acid sequences of more than 2,000,000 proteins are known. *The striking fact is that each protein has a unique, precisely defined amino acid sequence*. The amino acid sequence of a protein is referred to as its *primary structure*.

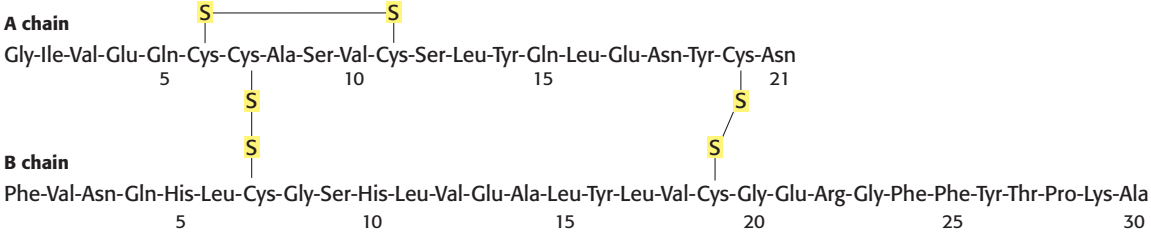


Figure 2.17 Amino acid sequence of bovine insulin.

A series of incisive studies in the late 1950s and early 1960s revealed that the amino acid sequences of proteins are determined by the nucleotide sequences of genes. The sequence of nucleotides in DNA specifies a complementary sequence of nucleotides in RNA, which in turn specifies the amino acid sequence of a protein. In particular, each of the 20 amino acids of the repertoire is encoded by one or more specific sequences of three nucleotides (Section 5.5).

Knowing amino acid sequences is important for several reasons. First, knowledge of the sequence of a protein is usually essential to elucidating its mechanism of action (e.g., the catalytic mechanism of an enzyme). In fact, proteins with novel properties can be generated by varying the sequence of known proteins. Second, amino acid sequences determine the three-dimensional structures of proteins. Amino acid sequence is the link between the genetic message in DNA and the three-dimensional structure that performs a protein's biological function. Analyses of relations between amino acid sequences and three-dimensional structures of proteins are uncovering the rules that govern the folding of polypeptide chains. Third, sequence determination is a component of molecular pathology, a rapidly growing area of medicine. Alterations in amino acid sequence can produce abnormal function and disease. Severe and sometimes fatal diseases, such as sickle-cell anemia (Chapter 7) and cystic

fibrosis, can result from a change in a single amino acid within a protein. Fourth, the sequence of a protein reveals much about its evolutionary history (Chapter 6). Proteins resemble one another in amino acid sequence only if they have a common ancestor. Consequently, molecular events in evolution can be traced from amino acid sequences; molecular paleontology is a flourishing area of research.

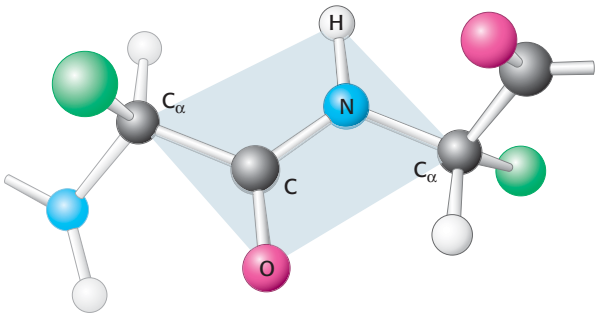
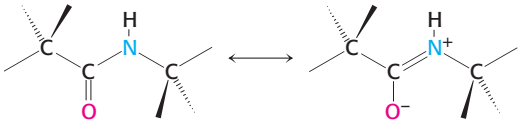


Figure 2.18 Peptide bonds are planar. In a pair of linked amino acids, six atoms (C_α, C, O, N, H, and C_α) lie in a plane. Side chains are shown as green balls.

Polypeptide chains are flexible yet conformationally restricted

Examination of the geometry of the protein backbone reveals several important features. First, *the peptide bond is essentially planar* (Figure 2.18). Thus, for a pair of amino acids linked by a peptide bond, six atoms lie in the same plane: the α-carbon atom and CO group of the first amino acid and the NH group and α-carbon atom of the second amino acid. The nature of the chemical bonding within a peptide accounts for the bond's planarity. The bond resonates between a single bond and a double bond. Because of this *double-bond character,* rotation about this bond is prevented and thus the conformation of the peptide backbone is constrained.



Peptide-bond resonance structures

The double-bond character is also expressed in the length of the bond between the CO and the NH groups. The C—N distance in a peptide bond is typically 1.32 Å, which is between the values expected for a C—N single bond (1.49 Å) and a C═N double bond (1.27 Å), as shown in Figure 2.19. Finally, the peptide bond is uncharged, allowing polymers of amino acids linked by peptide bonds to form tightly packed globular structures.

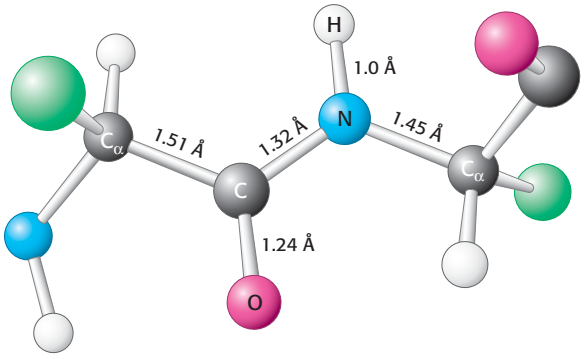


Figure 2.19 Typical bond lengths within a peptide unit. The peptide unit is shown in the trans configuration.

Two configurations are possible for a planar peptide bond. In the *trans* configuration, the two α-carbon atoms are on opposite sides of the peptide bond. In the *cis* configuration, these groups are on the same side of the peptide bond. *Almost all peptide bonds in proteins are trans*. This preference for trans over cis can be explained by the fact that steric clashes between groups attached to the α-carbon atoms hinder the formation of the cis form but do not arise in the trans configuration (Figure 2.20). By far the most common cis peptide bonds are X—Pro linkages. Such bonds show less preference for the trans configuration

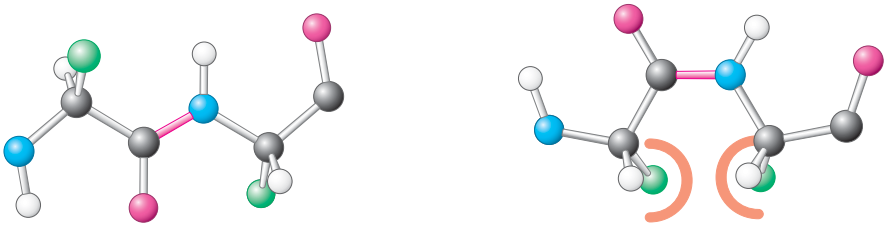


Figure 2.20 Trans and cis peptide bonds. The trans form is strongly favored because of steric clashes that arise in the cis form.

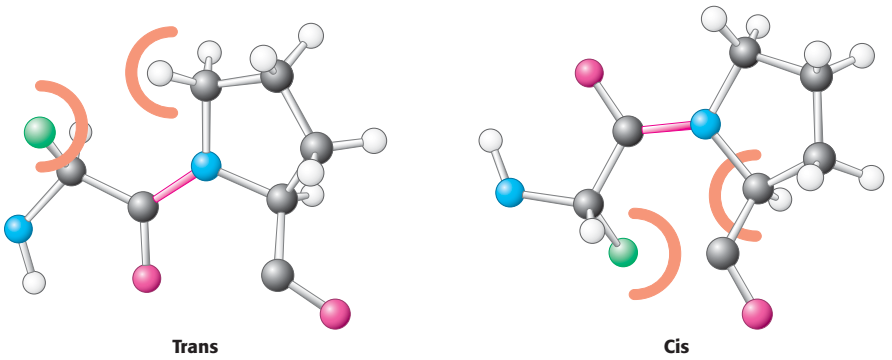


Figure 2.21 Trans and cis X–Pro bonds. The energies of these forms are similar to one another because steric clashes arise in both forms.

because the nitrogen of proline is bonded to two tetrahedral carbon atoms, limiting the steric differences between the trans and cis forms (Figure 2.21).

In contrast with the peptide bond, the bonds between the amino group and the α-carbon atom and between the α-carbon atom and the carbonyl group are pure single bonds. The two adjacent rigid peptide units can rotate about these bonds, taking on various orientations. *This freedom of rotation about two bonds of each amino acid allows proteins to fold in many different ways*. The rotations about these bonds can be specified by *torsion angles* (Figure 2.22). The angle of rotation about the bond between the nitrogen and the α-carbon atoms is called phi (ϕ). The angle of rotation about the bond between the α-carbon and the carbonyl carbon atoms is called psi (ψ). A clockwise rotation about either bond as viewed from the nitrogen atom toward the α-carbon atom or from the carbonyl group toward the α-carbon atom corresponds to a positive value. The ϕ and ψ angles determine the path of the polypeptide chain.

Torsion angle
A measure of the rotation about a bond, usually taken to lie between −180 and +180 degrees. Torsion angles are sometimes called dihedral angles.

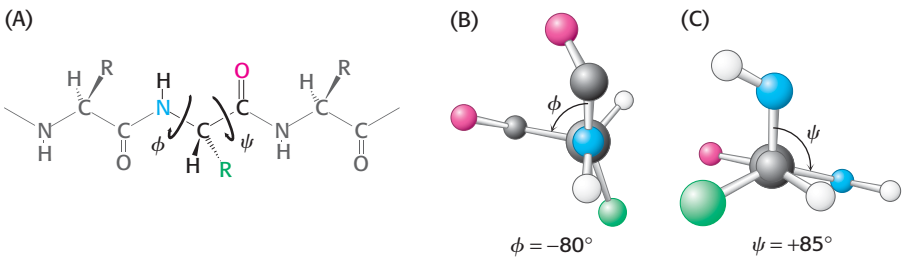


Figure 2.22 Rotation about bonds in a polypeptide. The structure of each amino acid in a polypeptide can be adjusted by rotation about two single bonds. (A) Phi (ϕ) is the angle of rotation about the bond between the nitrogen and the α-carbon atoms, whereas psi (ψ) is the angle of rotation about the bond between the α-carbon and the carbonyl carbon atoms. (B) A view down the bond between the nitrogen and the α-carbon atoms, showing how ϕ is measured. (C) A view down the bond between the α-carbon and the carbonyl carbon atoms, showing how ψ is measured.

Are all combinations of ϕ and ψ possible? Gopalasamudram Ramachandran recognized that many combinations are forbidden because of steric collisions between atoms. The allowed values can be visualized on a two-dimensional plot called a *Ramachandran diagram* (Figure 2.23). Three-quarters of the possible (ϕ, ψ) combinations are excluded simply by local steric clashes. *Steric exclusion, the fact that two atoms cannot be in the same place at the same time, can be a powerful organizing principle.*

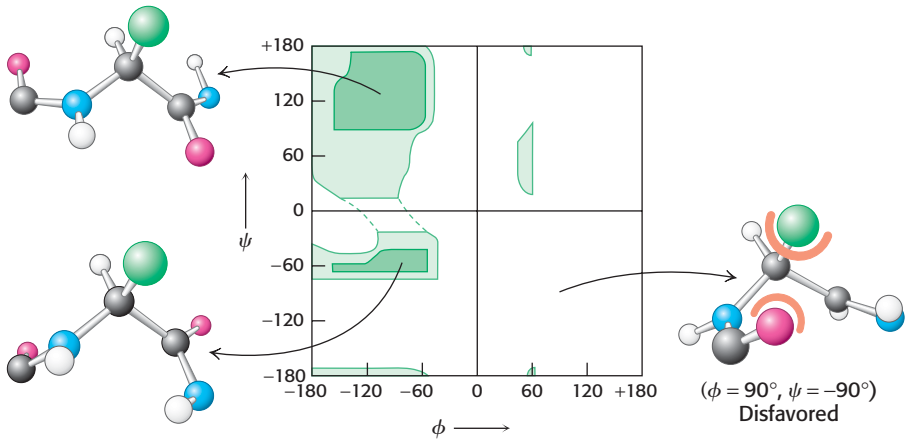


Figure 2.23 A Ramachandran diagram showing the values of ϕ and ψ. Not all ϕ and ψ values are possible without collisions between atoms. The most favorable regions are shown in dark green; borderline regions are shown in light green. The structure on the right is disfavored because of steric clashes.

The ability of biological polymers such as proteins to fold into well-defined structures is remarkable thermodynamically. An unfolded polymer exists as a random coil: each copy of an unfolded polymer will have a different conformation, yielding a mixture of many possible conformations. The favorable entropy associated with a mixture of many conformations opposes folding and must be overcome by interactions favoring the folded form. Thus, highly flexible polymers with a large number of possible conformations do not fold into unique structures. *The rigidity of the peptide unit and the restricted set of allowed* ϕ *and* ψ *angles limits the number of structures accessible to the unfolded form sufficiently to allow protein folding to take place.*

2.3 Secondary Structure: Polypeptide Chains Can Fold into Regular Structures Such As the Alpha Helix, the Beta Sheet, and Turns and Loops

Can a polypeptide chain fold into a regularly repeating structure? In 1951, Linus Pauling and Robert Corey proposed two periodic structures called the *α helix* (alpha helix) and the *β pleated sheet* (beta pleated sheet). Subsequently, other structures such as the *β turn* and *omega* (Ω) *loop* were identified. Although not periodic, these common turn or loop structures are well defined and contribute with α helices and β sheets to form the final protein structure. Alpha helices, β strands, and turns are formed by a regular pattern of hydrogen bonds between the peptide N—H and C═O groups of amino acids that are *near one another in the linear sequence*. Such folded segments are called *secondary structure*.

The alpha helix is a coiled structure stabilized by intrachain hydrogen bonds

In evaluating potential structures, Pauling and Corey considered which conformations of peptides were sterically allowed and which most fully exploited the hydrogen-bonding capacity of the backbone NH and CO groups. The first of their proposed structures, the *α helix*, is a rodlike structure (Figure 2.24). A tightly coiled backbone forms the inner part of the rod and the side chains extend outward in a helical array. The α helix is stabilized by hydrogen bonds between the NH and CO groups of the main chain. In

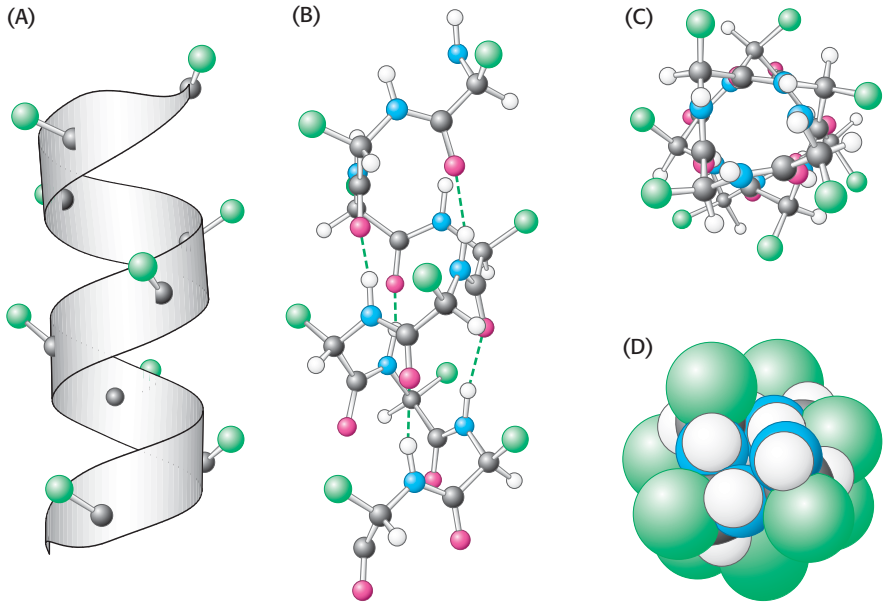


Figure 2.24 Structure of the α helix. (A) A ribbon depiction shows the α-carbon atoms and side chains (green). (B) A side view of a ball-and-stick version depicts the hydrogen bonds (dashed lines) between NH and CO groups. (C) An end view shows the coiled backbone as the inside of the helix and the side chains (green) projecting outward. (D) A space-filling view of part C shows the tightly packed interior core of the helix.

particular, the CO group of each amino acid forms a hydrogen bond with the NH group of the amino acid that is situated four residues ahead in the sequence (Figure 2.25). Thus, except for amino acids near the ends of an α helix, all *the main-chain CO and NH groups are hydrogen bonded*. Each residue is related to the next one by a rise, also called *translation*, of 1.5 Å along the helix axis and a rotation of 100 degrees, which gives 3.6 amino acid residues per turn of helix. Thus, amino acids spaced three and four apart in the sequence are spatially quite close to one another in an α helix. In contrast, amino acids spaced two apart in the sequence are situated on opposite sides of the helix and so are unlikely to make contact. The *pitch* of the α helix is the length of one complete turn along the helix axis and is equal to the product of the rise (1.5 Å) and the number of residues per turn (3.6), or 5.4 Å. The *screw sense* of a helix can be right-handed (clockwise) or left-handed (counterclockwise). The Ramachandran diagram reveals that both the right-handed and the left-handed helices are among allowed conformations

Screw sense

Describes the direction in which a helical structure rotates with respect to its axis. If, viewed down the axis of a helix, the chain turns in a clockwise direction, it has a right-handed screw sense. If the turning is counterclockwise, the screw sense is left-handed.

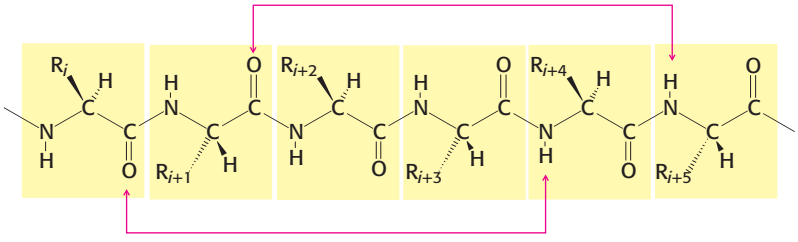


Figure 2.25 Hydrogen-bonding scheme for an α helix. In the α helix, the CO group of residue *i* forms a hydrogen bond with the NH group of residue *i* + 4.

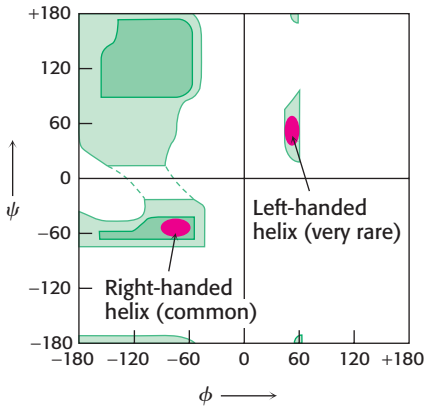


Figure 2.26 Ramachandran diagram for helices. Both right- and left-handed helices lie in regions of allowed conformations in the Ramachandran diagram. However, essentially all α helices in proteins are right-handed.

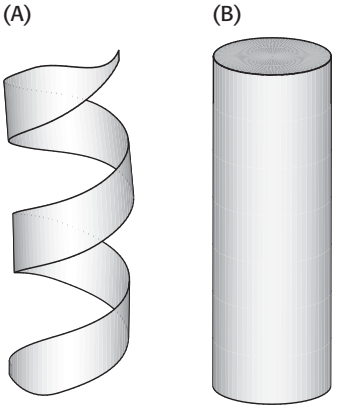


Figure 2.27 Schematic views of α helices. (A) A ribbon depiction. (B) A cylindrical depiction.

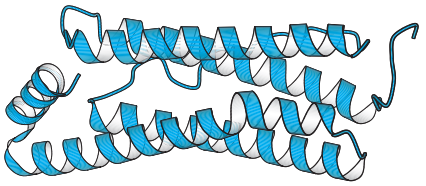


Figure 2.28 A largely α-helical protein. Ferritin, an iron-storage protein, is built from a bundle of α helices. [Drawn from 1AEW.pdb.]

(Figure 2.26). However, right-handed helices are energetically more favorable because there is less steric clash between the side chains and the backbone. *Essentially all α helices found in proteins are right-handed.* In schematic representations of proteins, α helices are depicted as twisted ribbons or rods (Figure 2.27).

Not all amino acids can be readily accommodated in an α helix. Branching at the β-carbon atom, as in valine, threonine, and isoleucine, tends to destabilize α helices because of steric clashes. Serine, aspartate, and asparagine also tend to disrupt α helices because their side chains contain hydrogen-bond donors or acceptors in close proximity to the main chain, where they compete for main-chain NH and CO groups. Proline also is a helix breaker because it lacks an NH group and because its ring structure prevents it from assuming the ϕ value to fit into an α helix.

The α-helical content of proteins ranges widely, from none to almost 100%. For example, about 75% of the residues in ferritin, a protein that helps store iron, are in α helices (Figure 2.28). Indeed, about 25% of all soluble proteins are composed of α helices connected by loops and turns of the polypeptide chain. Single α helices are usually less than 45 Å long. Many proteins that span biological membranes also contain α helices.

Beta sheets are stabilized by hydrogen bonding between polypeptide strands

Pauling and Corey proposed another periodic structural motif, which they named the *β pleated sheet* (β because it was the second structure that they elucidated, the α helix having been the first). The β pleated sheet (or, more simply, the β sheet) differs markedly from the rodlike α helix. It is composed

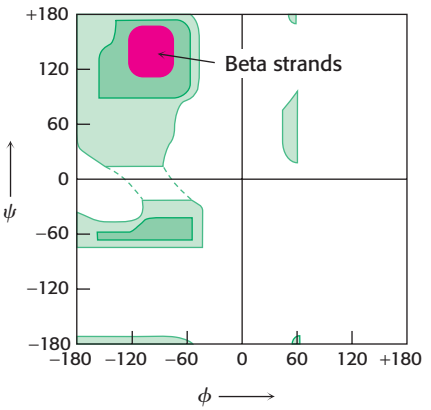


Figure 2.29 Ramachandran diagram for β strands. The red area shows the sterically allowed conformations of extended, β-strand-like structures.

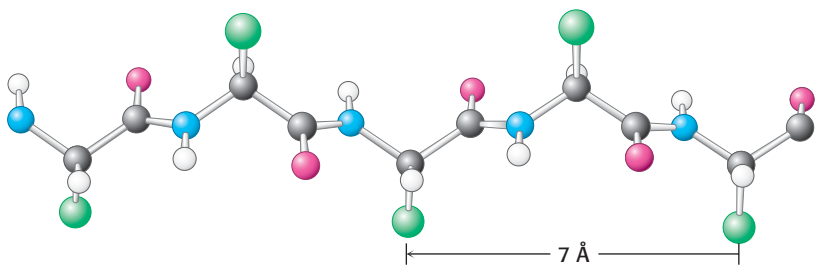


Figure 2.30 Structure of a β strand. The side chains (green) are alternately above and below the plane of the strand.

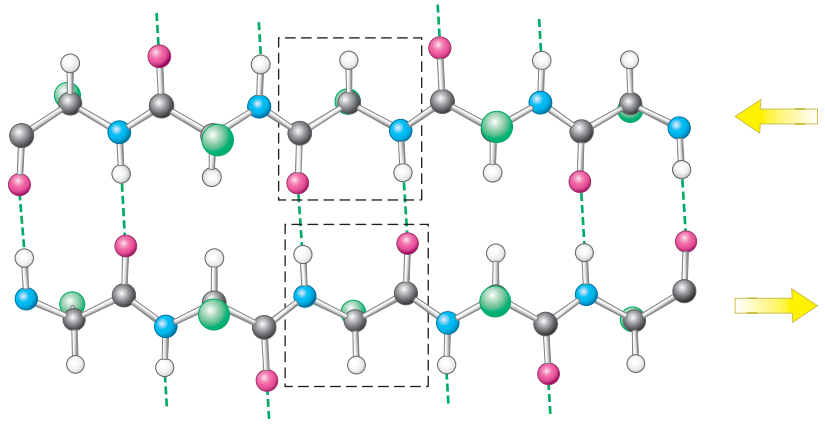


Figure 2.31 An antiparallel β sheet. Adjacent β strands run in opposite directions. Hydrogen bonds between NH and CO groups connect each amino acid to a single amino acid on an adjacent strand, stabilizing the structure.

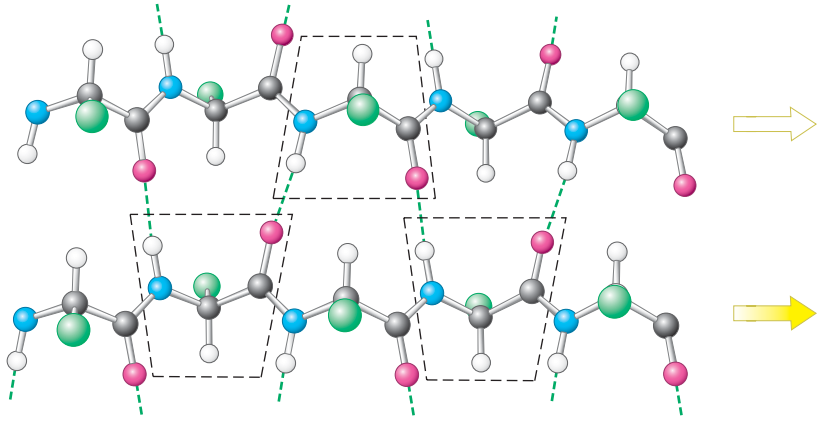


Figure 2.32 A parallel β sheet. Adjacent β strands run in the same direction. Hydrogen bonds connect each amino acid on one strand with two different amino acids on the adjacent strand.

of two or more polypeptide chains called *β strands*. A β strand is almost fully extended rather than being tightly coiled as in the α helix. A range of extended structures are sterically allowed (Figure 2.29).

The distance between adjacent amino acids along a β strand is approximately 3.5 Å, in contrast with a distance of 1.5 Å along an α helix. The side chains of adjacent amino acids point in opposite directions (Figure 2.30). A β sheet is formed by linking two or more β strands lying next to one another through hydrogen bonds. Adjacent chains in a β sheet can run in opposite directions (antiparallel β sheet) or in the same direction (parallel β sheet). In the antiparallel arrangement, the NH group and the CO group of each amino acid are respectively hydrogen bonded to the CO group and the NH group of a partner on the adjacent chain (Figure 2.31). In the parallel arrangement, the hydrogen-bonding scheme is slightly more complicated. For each amino acid, the NH group is hydrogen bonded to the CO group of one amino acid on the adjacent strand, whereas the CO group is hydrogen bonded to the NH group on the amino acid two residues farther along the chain (Figure 2.32). Many strands, typically 4 or 5 but as many as 10 or more, can come together in β sheets. Such β sheets can be purely antiparallel, purely parallel, or mixed (Figure 2.33).

In schematic representations, β strands are usually depicted by broad arrows pointing in the direction of the carboxyl-terminal end to indicate the

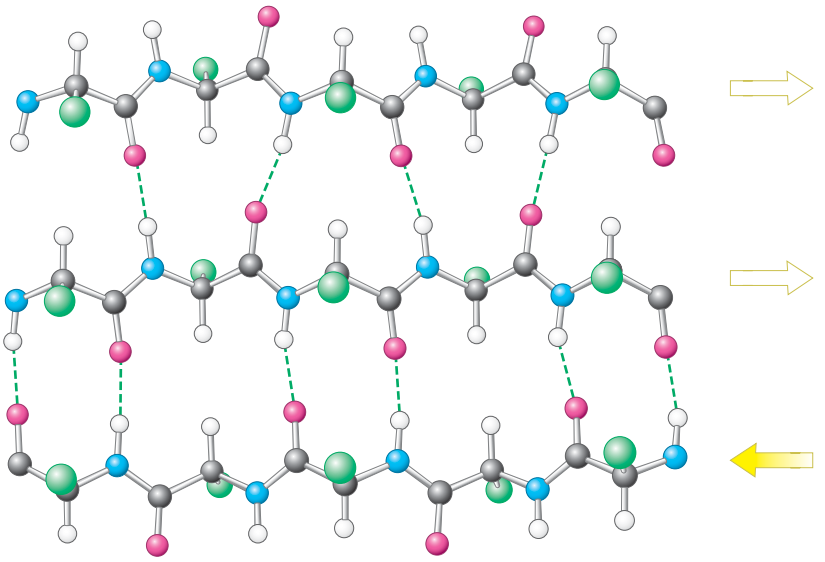


Figure 2.33 Structure of a mixed β sheet.

type of β sheet formed—parallel or antiparallel. More structurally diverse than α helices, β sheets can be almost flat but most adopt a somewhat twisted shape (Figure 2.34). The β sheet is an important structural element in many proteins. For example, fatty acid-binding proteins, important for lipid metabolism, are built almost entirely from β sheets (Figure 2.35).

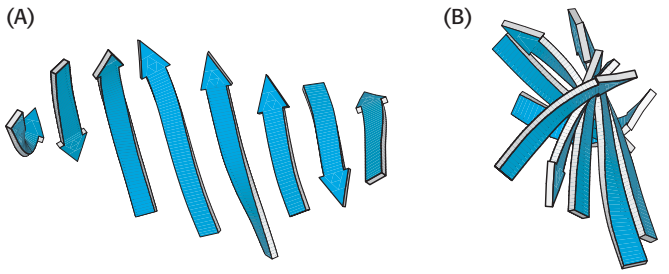


Figure 2.34 A schematic twisted β sheet. (A) A schematic model. (B) The schematic view rotated by 90 degrees to illustrate the twist more clearly.

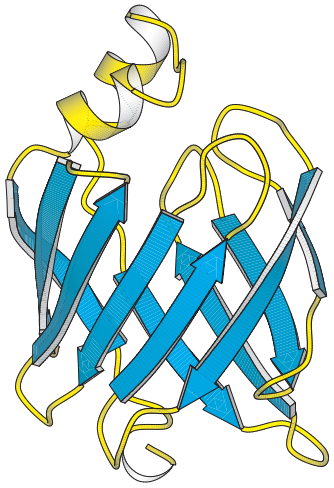


Figure 2.35 A protein rich in β sheets. The structure of a fatty acid-binding protein. [Drawn from 1FTP.pdb.]

Polypeptide chains can change direction by making reverse turns and loops

Most proteins have compact, globular shapes owing to reversals in the direction of their polypeptide chains. Many of these reversals are accomplished by a common structural element called the *reverse turn* (also known as the *β turn* or *hairpin turn*), illustrated in Figure 2.36. In many reverse turns, the CO group of residue i of a polypeptide is hydrogen bonded to the NH group of residue $i + 3$. This interaction stabilizes abrupt changes in direction of the polypeptide chain. In other cases, more-elaborate structures are responsible for chain reversals. These structures are called *loops* or sometimes *Ω loops* (omega loops) to suggest their overall shape. Unlike α helices and β strands, loops do not have regular, periodic structures. Nonetheless, loop structures are often rigid and well defined (Figure 2.37). Turns and loops invariably lie on the surfaces of proteins and thus often participate in interactions between proteins and other molecules.

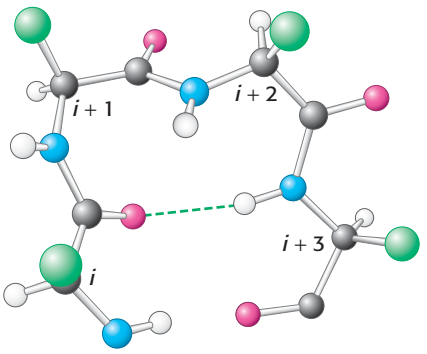


Figure 2.36 Structure of a reverse turn. The CO group of residue *i* of the polypeptide chain is hydrogen bonded to the NH group of residue *i* +3 to stabilize the turn.

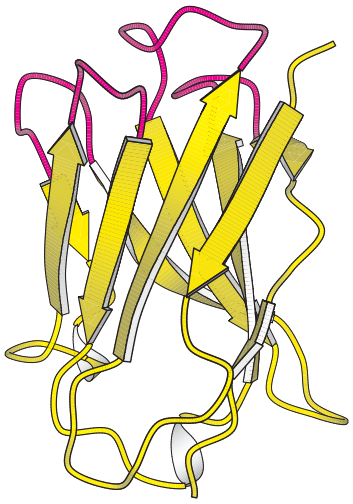


Figure 2.37 Loops on a protein surface. A part of an antibody molecule has surface loops (shown in red) that mediate interactions with other molecules. [Drawn from 7FTP.pdb.]

Fibrous proteins provide structural support for cells and tissues

Special types of helices are present in the two proteins α-keratin and collagen. These proteins form long fibers that serve a structural role.

α-Keratin, which is the primary component of wool, hair, and skin, consists of two right-handed α helices intertwined to form a type of left-handed superhelix called an *α-helical coiled coil*. α-Keratin is a member of a superfamily of proteins referred to as *coiled-coil proteins* (Figure 2.38). In these proteins, two or more α helices can entwine to form a very stable structure, which can have a length of 1000 Å (100 nm, or 0.1 μm) or more. There are approximately 60 members of this family in humans, including intermediate filaments, proteins that contribute to the cell cytoskeleton (internal scaffolding in a cell), and the muscle proteins myosin and tropomyosin (Section 35.2). Members of this family are characterized by a central region of 300 amino acids that contains imperfect repeats of a sequence of seven amino acids called a *heptad repeat*.

The two helices in α-keratin are cross-linked by weak interactions such as van der Waals forces and ionic interactions. These interactions are facilitated by the fact that the left-handed supercoil alters the two right-handed

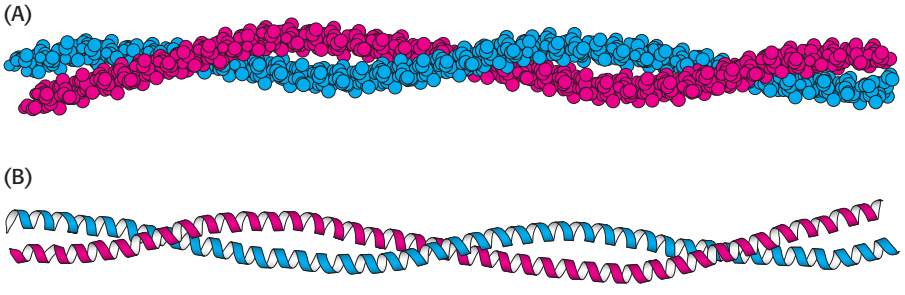


Figure 2.38 An α-helical coiled coil. (A) Space-filling model. (B) Ribbon diagram. The two helices wind around one another to form a superhelix. Such structures are found in many proteins, including keratin in hair, quills, claws, and horns. [Drawn from 1CIG.pdb.]

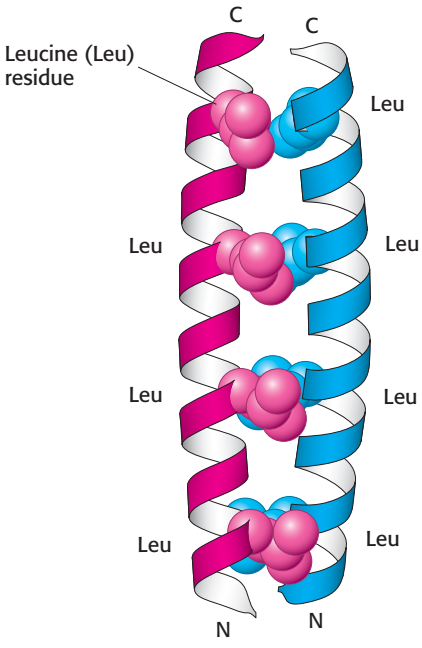


Figure 2.39 Heptad repeats in a coiled-coil protein. Every seventh residue in each helix is leucine. The two helices are held together by van der Waals interactions primarily between the leucine residues. [Drawn from 2ZTA.pdb.]

13
-Gly-Pro-Met-Gly-Pro-Ser-Gly-Pro-Arg-
22
-Gly-Leu-Hyp-Gly-Pro-Hyp-Gly-Ala-Hyp-
31
-Gly-Pro-Gln-Gly-Phe-Gln-Gly-Pro-Hyp-
40
-Gly-Glu-Hyp-Gly-Glu-Hyp-Gly-Ala-Ser-
49
-Gly-Pro-Met-Gly-Pro-Arg-Gly-Pro-Hyp-
58
-Gly-Pro-Hyp-Gly-Lys-Asn-Gly-Asp-Asp-

Figure 2.40 Amino acid sequence of a part of a collagen chain. Every third residue is a glycine. Proline and hydroxyproline also are abundant.

α helices such that there are 3.5 residues per turn instead of 3.6. Thus, the pattern of side-chain interactions can be repeated every seven residues, forming the heptad repeats. Two helices with such repeats are able to interact with one another if the repeats are complementary (Figure 2.39). For example, the repeating residues may be hydrophobic, allowing van der Waals interactions, or have opposite charge, allowing ionic interactions. In addition, the two helices may be linked by disulfide bonds formed by neighboring cysteine residues. The bonding of the helices accounts for the physical properties of wool, an example of an α-keratin. Wool is extensible and can be stretched to nearly twice its length because the α helices stretch, breaking the weak interactions between neighboring helices. However, the covalent disulfide bonds resist breakage and return the fiber to its original state once the stretching force is released. The number of disulfide bond cross-links further defines the fiber's properties. Hair and wool, having fewer cross-links, are flexible. Horns, claws, and hooves, having more cross-links, are much harder.

A different type of helix is present in collagen, the most abundant protein of mammals. Collagen is the main fibrous component of skin, bone, tendon, cartilage, and teeth. This extracellular protein is a rod-shaped molecule, about 3000 Å long and only 15 Å in diameter. It contains three helical polypeptide chains, each nearly 1000 residues long. Glycine appears at every third residue in the amino acid sequence, and the sequence glycine-proline-hydroxyproline recurs frequently (Figure 2.40). Hydroxyproline is a derivative of proline that has a hydroxyl group in place of one of the hydrogen atoms on the pyrrolidine rings.

The collagen helix has properties different from those of the α helix. Hydrogen bonds within a strand are absent. Instead, *the helix is stabilized by steric repulsion of the pyrrolidine rings of the proline and hydroxyproline residues* (Figure 2.41). The pyrrolidine rings keep out of each other's way when the polypeptide chain assumes its helical form, which has about three residues per turn. Three strands wind around one another to form a *superhelical cable* that is stabilized by hydrogen bonds between strands. The hydrogen bonds form between the peptide NH groups of glycine residues and the CO groups of residues on the other chains. The hydroxyl groups of hydroxyproline residues also participate in hydrogen bonding, and the absence of the hydroxyl groups results in the disease scurvy (Section 27.6).

The inside of the triple-stranded helical cable is very crowded and accounts for the requirement that glycine be present at every third position on each strand (Figure 2.42A). *The only residue that can fit in an interior position is glycine*. The amino acid residue on either side of glycine is located on the outside of the cable, where there is room for the bulky rings of proline and hydroxyproline residues (Figure 2.42B).

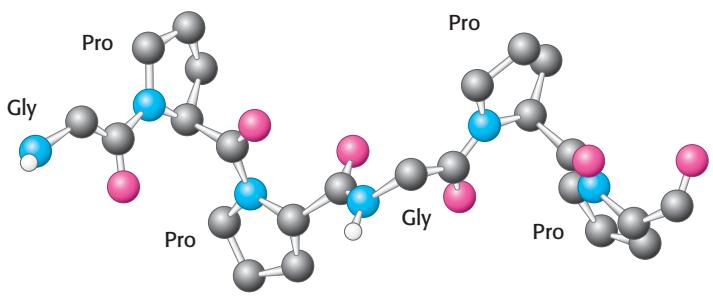


Figure 2.41 Conformation of a single strand of a collagen triple helix.

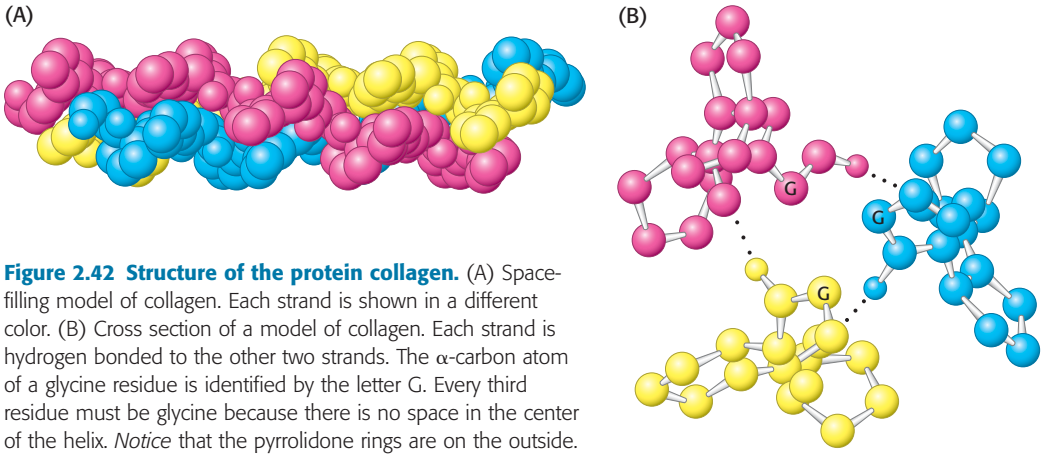


Figure 2.42 Structure of the protein collagen. (A) Space-filling model of collagen. Each strand is shown in a different color. (B) Cross section of a model of collagen. Each strand is hydrogen bonded to the other two strands. The α-carbon atom of a glycine residue is identified by the letter G. Every third residue must be glycine because there is no space in the center of the helix. *Notice* that the pyrrolidone rings are on the outside.

The importance of the positioning of glycine inside the triple helix is illustrated in the disorder osteogenesis imperfecta, also known as brittle bone disease. In this condition, which can vary from mild to very severe, other amino acids replace the internal glycine residue. This replacement leads to a delayed and improper folding of collagen, and the accumulation of defective collagen results in cell death. The most serious symptom is severe bone fragility. Defective collagen in the eyes causes the whites of the eyes to have a blue tint (blue sclera).

2.4 Tertiary Structure: Water-Soluble Proteins Fold into Compact Structures with Nonpolar Cores

Let us now examine how amino acids are grouped together in a complete protein. X-ray crystallographic and nuclear magnetic resonance (NMR) studies (Section 3.6) have revealed the detailed three-dimensional structures of thousands of proteins. We begin here with an examination of *myoglobin,* the first protein to be seen in atomic detail.

Myoglobin, the oxygen carrier in muscle, is a single polypeptide chain of 153 amino acids (see Chapter 7). The capacity of myoglobin to bind oxygen depends on the presence of *heme,* a nonpolypeptide *prosthetic (helper) group* consisting of protoporphyrin IX and a central iron atom. *Myoglobin is an extremely compact molecule.* Its overall dimensions are 45 × 35 × 25 Å, an order of magnitude less than if it were fully stretched out (Figure 2.43). About 70% of the main chain is folded into eight α helices, and much of the rest of the chain forms turns and loops between helices.

The folding of the main chain of myoglobin, like that of most other proteins, is complex and devoid of symmetry. The overall course of the polypeptide chain of a protein is referred to as its *tertiary structure*. A unifying principle emerges from the distribution of side chains. The striking fact is that *the interior consists almost entirely of nonpolar residues* such as leucine, valine, methionine, and phenylalanine (Figure 2.44). Charged residues such as aspartate, glutamate, lysine, and arginine are absent from the inside of myoglobin. The only polar residues inside are two histidine residues, which play critical roles in binding iron and oxygen. The outside of myoglobin, on the other hand, consists of both polar and nonpolar residues. The space-filling model shows that there is very little empty space inside.

This contrasting distribution of polar and nonpolar residues reveals a key facet of protein architecture. In an aqueous environment, protein folding is driven by the strong tendency of hydrophobic residues to be excluded

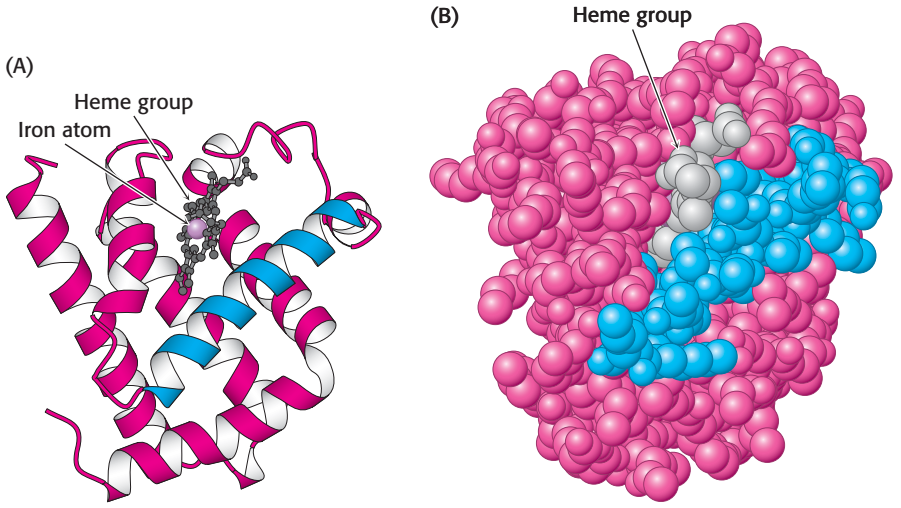


Figure 2.43 Three-dimensional structure of myoglobin. (A) A ribbon diagram shows that the protein consists largely of α helices. (B) A space-filling model in the same orientation shows how tightly packed the folded protein is. *Notice* that the heme group is nestled into a crevice in the compact protein with only an edge exposed. One helix is blue to allow comparison of the two structural depictions. [Drawn from 1A6N.pdb.]

from water. Recall that a system is more thermodynamically stable when hydrophobic groups are clustered rather than extended into the aqueous surroundings (Chapter 1). *The polypeptide chain therefore folds so that its hydrophobic side chains are buried and its polar, charged chains are on the surface*. Many α helices and β strands are amphipathic; that is, the α helix or β strand has a hydrophobic face, which points into the protein interior, and a more polar face, which points into solution. The fate of the main chain accompanying the hydrophobic side chains is important, too. An unpaired peptide NH or CO group markedly prefers water to a nonpolar milieu. The secret of burying a segment of main chain in a hydrophobic environment is to pair all the NH and CO groups by hydrogen bonding. This pairing is neatly accomplished in an α helix or β sheet. Van der Waals interactions between tightly packed hydrocarbon side chains also contribute to the stability of proteins. We can now understand why the set of 20 amino acids contains several that differ subtly in size and shape. They provide a palette from which to choose to fill the interior of a protein neatly and thereby maximize van der Waals interactions, which require intimate contact.

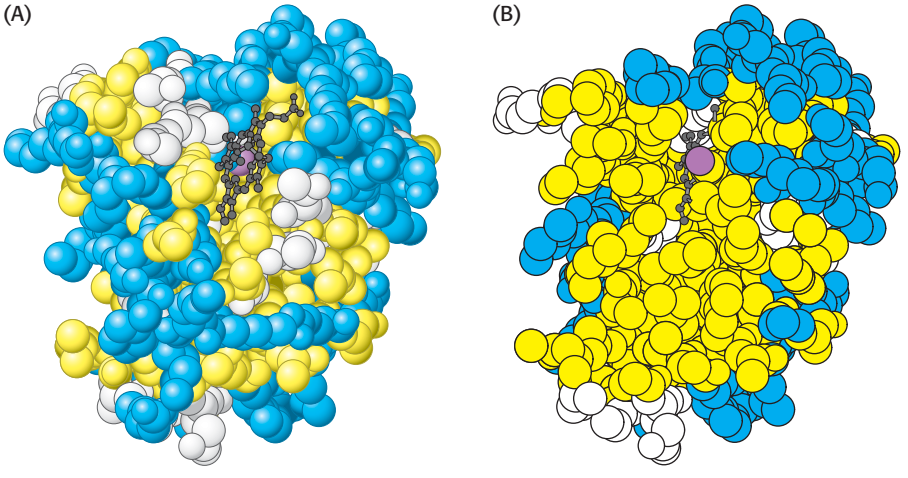


Figure 2.44 Distribution of amino acids in myoglobin. (A) A space-filling model of myoglobin with hydrophobic amino acids shown in yellow, charged amino acids shown in blue, and others shown in white. *Notice* that the surface of the molecule has many charged amino acids, as well as some hydrophobic amino acids. (B) In this cross-sectional view, *notice* that mostly hydrophobic amino acids are found on the inside of the structure, whereas the charged amino acids are found on the protein surface. [Drawn from 1MBD.pdb.]

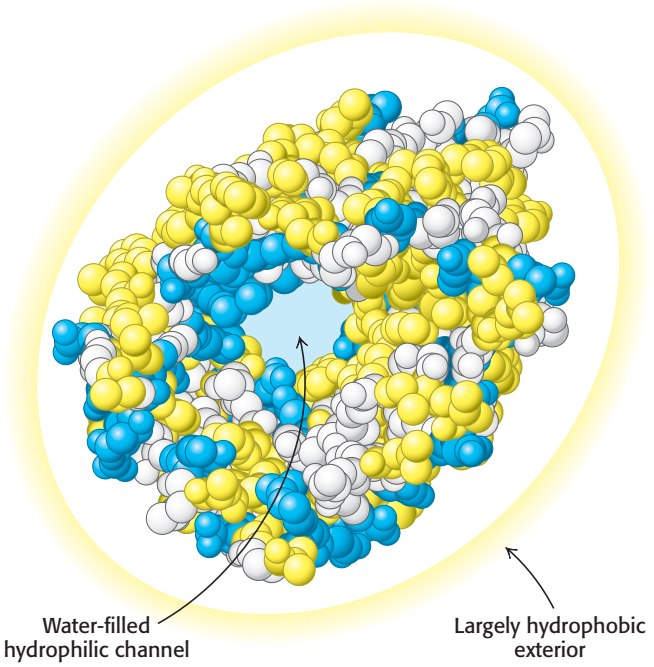


Figure 2.45 "Inside out" amino acid distribution in porin. The outside of porin (which contacts hydrophobic groups in membranes) is covered largely with hydrophobic residues, whereas the center includes a water-filled channel lined with charged and polar amino acids. [Drawn from 1PRN.pdb.]

Some proteins that span biological membranes are "the exceptions that prove the rule" because they have the reverse distribution of hydrophobic and hydrophilic amino acids. For example, consider porins, proteins found in the outer membranes of many bacteria (Figure 2.45). Membranes are built largely of hydrophobic alkane chains (Section 12.2). Thus, porins are covered on the outside largely with hydrophobic residues that interact with the neighboring alkane chains. In contrast, the center of the protein contains many charged and polar amino acids that surround a water-filled channel running through the middle of the protein. Thus, because porins function in hydrophobic environments, they are "inside out" relative to proteins that function in aqueous solution.

Certain combinations of secondary structure are present in many proteins and frequently exhibit similar functions. These combinations are called *motifs* or *supersecondary structures*. For example, an α helix separated from another α helix by a turn, called a *helix-turn-helix* unit, is found in many proteins that bind DNA (Figure 2.46).

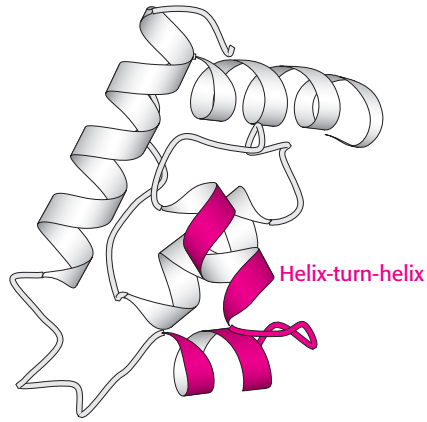


Figure 2.46 The helix-turn-helix motif, a supersecondary structural element. Helix-turn-helix motifs are found in many DNA-binding proteins. [Drawn from 1LMB.pdb.]

Some polypeptide chains fold into two or more compact regions that may be connected by a flexible segment of polypeptide chain, rather like pearls on a string. These compact globular units, called *domains,* range in size from about 30 to 400 amino acid residues. For example, the extracellular part of CD4, the cell-surface protein on certain cells of the immune system to which the human immunodeficiency virus (HIV) attaches itself, comprises four similar domains of approximately 100 amino acids each (Figure 2.47). Proteins may have domains in common even if their overall tertiary structures are different.

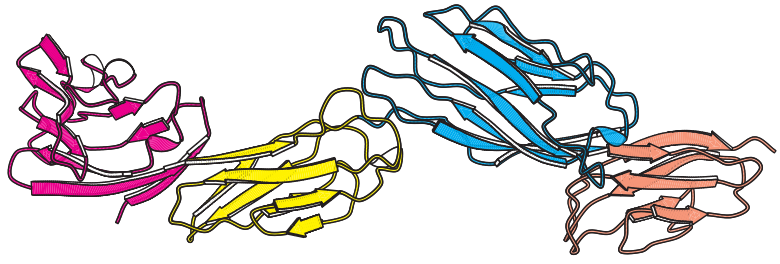


Figure 2.47 Protein domains. The cell-surface protein CD4 consists of four similar domains. [Drawn from 1WIO.pdb.]

2.5 Quaternary Structure: Polypeptide Chains Can Assemble into Multisubunit Structures

Four levels of structure are frequently cited in discussions of protein architecture. So far, we have considered three of them. *Primary structure* is the amino acid sequence. *Secondary structure* refers to the spatial arrangement of amino acid residues that are nearby in the sequence. Some of these arrangements are of a regular kind, giving rise to a periodic structure. The α helix and β strand are elements of secondary structure. *Tertiary structure* refers to the spatial arrangement of amino acid residues that are far apart in the sequence and to the pattern of disulfide bonds. We now turn to proteins containing more than one polypeptide chain. Such proteins exhibit a fourth level of structural organization. Each polypeptide chain in such a protein is called a *subunit*. *Quaternary structure* refers to the spatial arrangement of subunits and the nature of their interactions. The simplest sort of quaternary structure is a *dimer,* consisting of two identical subunits. This organization is present in the DNA-binding protein Cro found in a bacterial virus called λ (Figure 2.48). More-complicated quaternary structures also are common. More than one type of subunit can be present, often in variable numbers. For example, human hemoglobin, the oxygen-carrying protein in blood, consists of two subunits of one type (designated α) and two subunits of another type (designated β), as illustrated in Figure 2.49. Thus, the hemoglobin molecule exists as an $\alpha_2\beta_2$ tetramer. Subtle changes in the arrangement of subunits within the hemoglobin molecule allow it to carry oxygen from the lungs to tissues with great efficiency (Chapter 7).

Viruses make the most of a limited amount of genetic information by forming coats that use the same kind of subunit repetitively in a symmetric array. The coat of rhinovirus, the virus that causes the common cold, includes 60 copies of each of four subunits (Figure 2.50). The subunits come together to form a nearly spherical shell that encloses the viral genome.

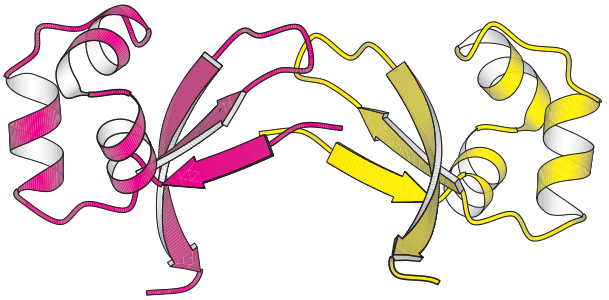


Figure 2.48 Quaternary structure. The Cro protein of bacteriophage λ is a dimer of identical subunits. [Drawn from 5CRO.pdb.]

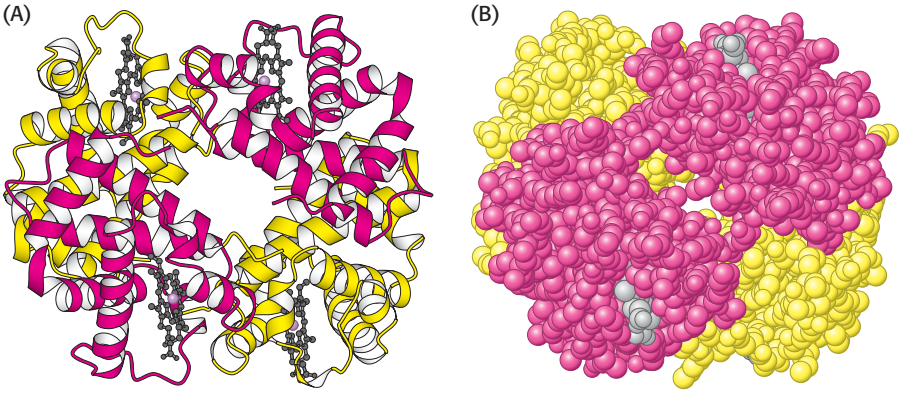


Figure 2.49 The $\alpha_2\beta_2$ tetramer of human hemoglobin. The structure of the two identical α subunits (red) is similar to but not identical with that of the two identical β subunits (yellow). The molecule contains four heme groups (gray with the iron atom shown in purple). (A) The ribbon diagram highlights the similarity of the subunits and shows that they are composed mainly of α helices. (B) The space-filling model illustrates how the heme groups occupy crevices in the protein. [Drawn from 1A3N.pdb.]

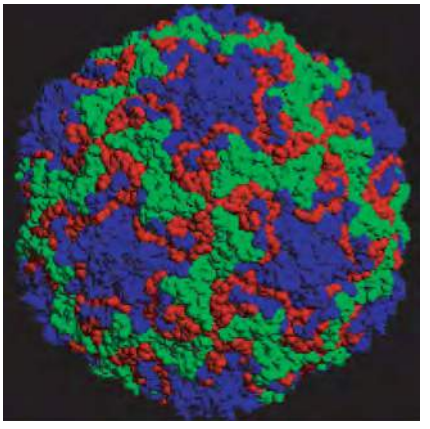


Figure 2.50 Complex quaternary structure. The coat of human rhinovirus, the cause of the common cold, comprises 60 copies of each of four subunits. The three most prominent subunits are shown as different colors.

2.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure

How is the elaborate three-dimensional structure of proteins attained? The classic work of Christian Anfinsen in the 1950s on the enzyme ribonuclease revealed the relation between the amino acid sequence of a protein and its conformation. Ribonuclease is a single polypeptide chain consisting of 124 amino acid residues cross-linked by four disulfide bonds (Figure 2.51). Anfinsen's plan was to destroy the three-dimensional structure of the enzyme and to then determine what conditions were required to restore the structure.

Agents such as urea or guanidinium chloride effectively disrupt a protein's noncovalent bonds. Although the mechanism of action of these agents is not fully understood, computer simulations suggest that they replace water as the molecule solvating the protein and are then able to disrupt the van der Waals interactions stabilizing the protein structure. The disulfide bonds can be cleaved reversibly by reducing them with a reagent such as *β-mercaptoethanol* (Figure 2.52). In the presence of a large excess of β-mercaptoethanol, the disulfides (cystines) are fully converted into sulfhydryls (cysteines).

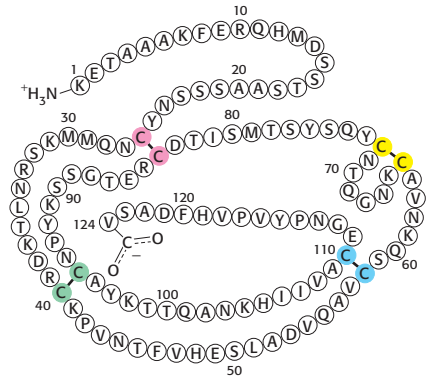
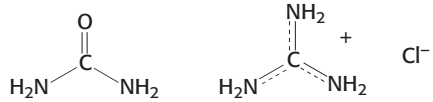
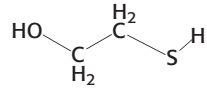


Figure 2.51 Amino acid sequence of bovine ribonuclease. The four disulfide bonds are shown in color. [After C. H. W. Hirs, S. Moore, and W. H. Stein, *J. Biol. Chem.* 235:633–647, 1960.]





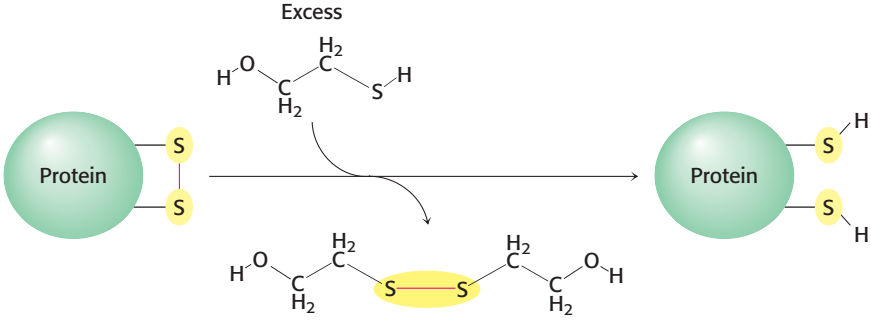


Figure 2.52 Role of β-mercaptoethanol in reducing disulfide bonds. Note that, as the disulfides are reduced, the β-mercaptoethanol is oxidized and forms dimers.

Most polypeptide chains devoid of cross-links assume a *random-coil conformation* in 8 M urea or 6 M guanidinium chloride. When ribonuclease was treated with β-mercaptoethanol in 8 M urea, the product was a fully reduced, randomly coiled polypeptide chain *devoid of enzymatic activity*. When a protein is converted into a randomly coiled peptide without its normal activity, it is said to be *denatured* (Figure 2.53).

Anfinsen then made the critical observation that the denatured ribonuclease, freed of urea and β-mercaptoethanol by dialysis, slowly regained

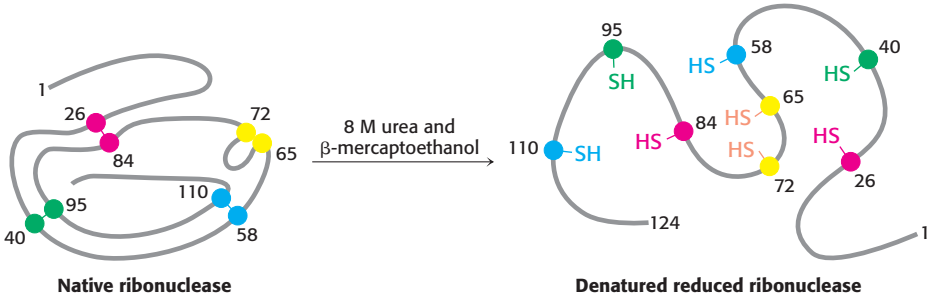


Figure 2.53 Reduction and denaturation of ribonuclease.

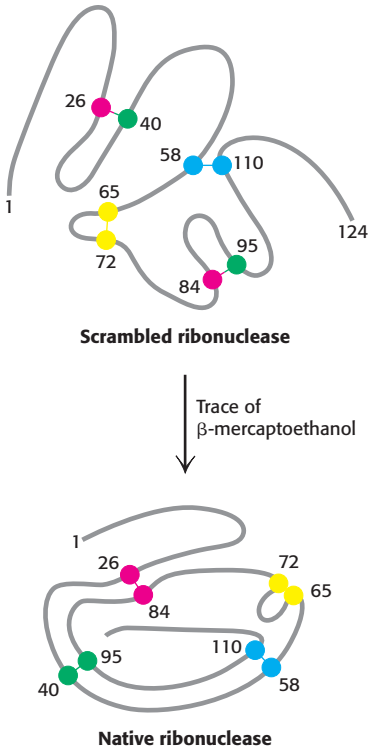


Figure 2.54 Reestablishing correct disulfide pairing. Native ribonuclease can be re-formed from scrambled ribonuclease in the presence of a trace of β-mercaptoethanol.

enzymatic activity. He immediately perceived the significance of this chance finding: the sulfhydryl groups of the denatured enzyme became oxidized by air, and the enzyme spontaneously refolded into a catalytically active form. Detailed studies then showed that nearly all the original enzymatic activity was regained if the sulfhydryl groups were oxidized under suitable conditions. All the measured physical and chemical properties of the refolded enzyme were virtually identical with those of the native enzyme. These experiments showed that *the information needed to specify the catalytically active structure of ribonuclease is contained in its amino acid sequence*. Subsequent studies have established the generality of this central principle of biochemistry: *sequence specifies conformation.* The dependence of conformation on sequence is especially significant because of the intimate connection between conformation and function.

A quite different result was obtained when reduced ribonuclease was reoxidized while it was still in 8 M urea and the preparation was then dialyzed to remove the urea. Ribonuclease reoxidized in this way had only 1% of the enzymatic activity of the native protein. Why were the outcomes so different when reduced ribonuclease was reoxidized in the presence and absence of urea? The reason is that the wrong disulfides formed pairs in urea. There are 105 different ways of pairing eight cysteine molecules to form four disulfides; only one of these combinations is enzymatically active. The 104 wrong pairings have been picturesquely termed "scrambled" ribonuclease. Anfinsen found that scrambled ribonuclease spontaneously converted into fully active, native ribonuclease when trace amounts of β-mercaptoethanol were added to an aqueous solution of the protein (Figure 2.54). The added β-mercaptoethanol catalyzed the rearrangement of disulfide pairings until the native structure was regained in about 10 hours. *This process was driven by the decrease in free energy as the scrambled conformations were converted into the stable, native conformation of the enzyme.* The native disulfide pairings of ribonuclease thus contribute to the stabilization of the thermodynamically preferred structure.

Similar refolding experiments have been performed on many other proteins. In many cases, the native structure can be generated under suitable conditions. For other proteins, however, refolding does not proceed efficiently. In these cases, the unfolding protein molecules usually become tangled up with one another to form aggregates. Inside cells, proteins called *chaperones* block such illicit interactions. Additionally, it is now evident that some proteins do not assume a defined structure until they interact with molecular partners, as we will see shortly.

Amino acids have different propensities for forming alpha helices, beta sheets, and beta turns

How does the amino acid sequence of a protein specify its three-dimensional structure? How does an unfolded polypeptide chain acquire the form of the native protein? These fundamental questions in biochemistry can be approached by first asking a simpler one: What determines whether a particular sequence in a protein forms an α helix, a β strand, or a turn? One source of insight is to examine the frequency of occurrence of particular amino acid residues in these secondary structures (Table 2.3). Residues such as alanine, glutamate, and leucine tend to be present in α helices, whereas valine and isoleucine tend to be present in β strands. Glycine, asparagine, and proline have a propensity for being present in turns.

Studies of proteins and synthetic peptides have revealed some reasons for these preferences. The α helix can be regarded as the default conformation. Branching at the β-carbon atom, as in valine, threonine, and isoleu-

Table 2.3 Relative frequencies of amino acid residues in secondary structures

Amino acid	α helix	β sheet	Reverse turn
Glu	1.59	0.52	1.01
Ala	1.41	0.72	0.82
Leu	1.34	1.22	0.57
Met	1.30	1.14	0.52
Gln	1.27	0.98	0.84
Lys	1.23	0.69	1.07
Arg	1.21	0.84	0.90
His	1.05	0.80	0.81
Val	0.90	1.87	0.41
Ile	1.09	1.67	0.47
Tyr	0.74	1.45	0.76
Cys	0.66	1.40	0.54
Trp	1.02	1.35	0.65
Phe	1.16	1.33	0.59
Thr	0.76	1.17	0.96
Gly	0.43	0.58	1.77
Asn	0.76	0.48	1.34
Pro	0.34	0.31	1.32
Ser	0.57	0.96	1.22
Asp	0.99	0.39	1.24

Note: The amino acids are grouped according to their preference for α helices (top group), β sheets (middle group), or turns (bottom group).
Source: T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2d ed. (W. H. Freeman and Company, 1992), p. 256.

cine, tends to destabilize α helices because of steric clashes. These residues are readily accommodated in β strands, in which their side chains project out of the plane containing the main chain. Serine, aspartate, and asparagine tend to disrupt α helices because their side chains contain hydrogen-bond donors or acceptors in close proximity to the main chain, where they compete for main-chain NH and CO groups. Proline tends to disrupt both α helices and β strands because it lacks an NH group and because its ring structure restricts its ϕ value to near 60 degrees. Glycine readily fits into all structures and for that reason does not favor helix formation in particular.

Can we predict the secondary structure of a protein by using this knowledge of the conformational preferences of amino acid residues? Accurate predictions of secondary structure adopted by even a short stretch of residues have proved to be difficult. What stands in the way of more-accurate prediction? Note that the conformational preferences of amino acid residues are not tipped all the way to one structure (see Table 2.3). For example, glutamate, one of the strongest helix formers, prefers α helix to β strand by only a factor of two. The preference ratios of most other residues are smaller. Indeed, some penta- and hexapeptide sequences have been found to adopt one structure in one protein and an entirely different structure in another (Figure 2.55). Hence, some amino acid sequences do not uniquely determine secondary structure. Tertiary interactions—interactions between residues that are far apart in the sequence—may be decisive in specifying the secondary structure of some segments. The context is often crucial in determining the conformational outcome. The conformation of a protein evolved to work in a particular environment or context. Substantial improvements in secondary

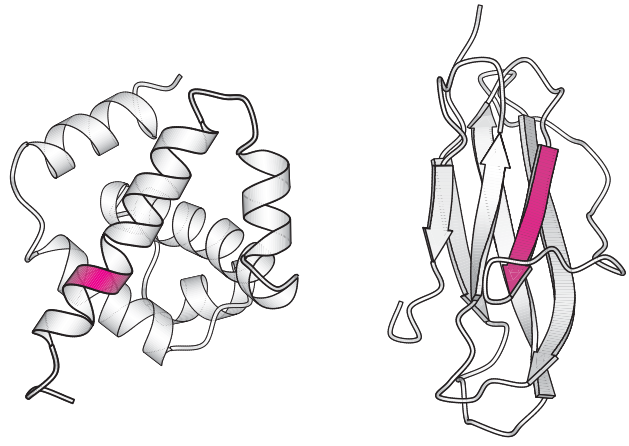


Figure 2.55 Alternative conformations of a peptide sequence. Many sequences can adopt alternative conformations in different proteins. Here the sequence VDLLKN shown in red assumes an α helix in one protein context (left) and a β strand in another (right). [Drawn from (left) 3WRP.pdb and (right) 2HLA.pdb.]

structure prediction can be achieved by using families of related sequences, each of which adopts the same structure.

Protein folding is a highly cooperative process

Proteins can be denatured by any treatment that disrupts the weak bonds stabilizing tertiary structure, such as heating, or by chemical denaturants such as urea or guanidinium chloride. For many proteins, a comparison of the degree of unfolding as the concentration of denaturant increases reveals a sharp transition from the folded, or native, form to the unfolded, or denatured form, suggesting that only these two conformational states are present to any significant extent (Figure 2.56). A similar sharp transition is observed if denaturants are removed from unfolded proteins, allowing the proteins to fold.

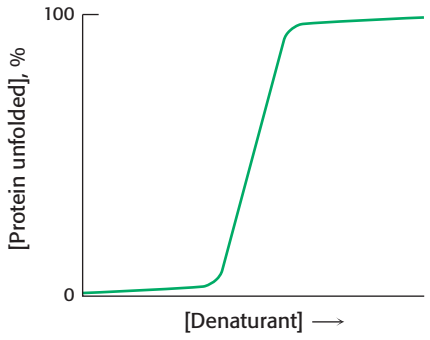


Figure 2.56 Transition from folded to unfolded state. Most proteins show a sharp transition from the folded to the unfolded form on treatment with increasing concentrations of denaturants.

The sharp transition seen in Figure 2.56 suggests that protein folding and unfolding is an *"all or none" process* that results from a *cooperative transition*. For example, suppose that a protein is placed in conditions under which some part of the protein structure is thermodynamically unstable. As this part of the folded structure is disrupted, the interactions between it and the remainder of the protein will be lost. The loss of these interactions, in turn, will destabilize the remainder of the structure. Thus, conditions that lead to the disruption of any part of a protein structure are likely to unravel the protein completely. The structural properties of proteins provide a clear rationale for the cooperative transition.

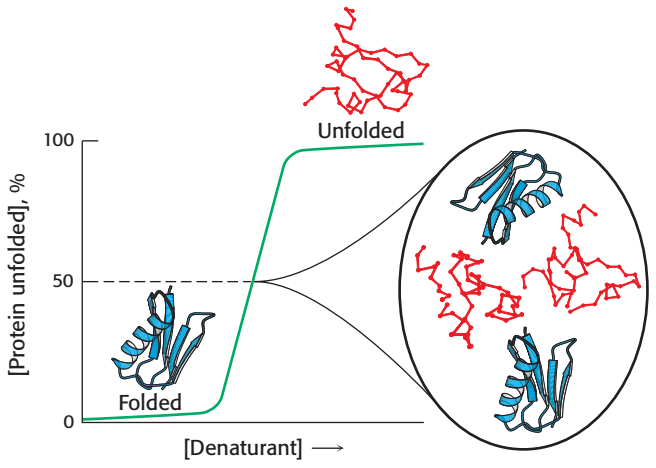


Figure 2.57 Components of a partly denatured protein solution. In a half-unfolded protein solution, half the molecules are fully folded and half are fully unfolded.

The consequences of cooperative folding can be illustrated by considering the contents of a protein solution under conditions corresponding to the middle of the transition between the folded and the unfolded forms. Under these conditions, the protein is "half folded." Yet the solution will appear to have no partly folded molecules but, instead, look as if it is a 50/50 mixture of fully folded and fully unfolded molecules (Figure 2.57). Although the protein may appear to behave as if it exists in only two states, this simple two-state existence is an impossibility at a molecular level. Even simple reactions go through reaction intermediates, and so a complex molecule such as a protein cannot simply switch from a completely unfolded state to the native state in one step. Unstable, transient intermediate structures must exist between the native and denatured state (p. 53). Determining the nature of these intermediate structures is an intense area of biochemical research.

Proteins fold by progressive stabilization of intermediates rather than by random search

How does a protein make the transition from an unfolded structure to a unique conformation in the native form? One possibility a priori would be that all possible conformations are tried out to find the energetically most favorable one. How long would such a random search take? Consider a small protein with 100 residues. Cyrus Levinthal calculated that, if each residue can assume three different conformations, the total number of structures would be 3^{100}, which is equal to 5×10^{47}. If it takes 10^{-13} s to convert one structure into another, the total search time would be $5 \times 10^{47} \times 10^{-13}$ s, which is equal to 5×10^{34} s, or 1.6×10^{27} years. Clearly, it would take much too long for even a small protein to fold properly by randomly trying out all possible conformations. The enormous difference between calculated and

actual folding times is called *Levinthal's paradox*. This paradox clearly reveals that proteins do not fold by trying every possible conformation; instead, they must follow at least a partly defined folding pathway consisting of intermediates between the fully denatured protein and its native structure.

The way out of this paradox is to recognize the power of *cumulative selection*. Richard Dawkins, in *The Blind Watchmaker,* asked how long it would take a monkey poking randomly at a typewriter to reproduce Hamlet's remark to Polonius, "Methinks it is like a weasel" (Figure 2.58). An astronomically large number of keystrokes, of the order of 10^{40}, would be required. However, suppose that we preserved each correct character and allowed the monkey to retype only the wrong ones. In this case, only a few thousand keystrokes, on average, would be needed. The crucial difference between these cases is that the first employs a completely random search, whereas, in the second, *partly correct intermediates are retained*.

The essence of protein folding is the tendency to retain partly correct intermediates. However, the protein-folding problem is much more difficult than the one presented to our simian Shakespeare. First, the criterion of correctness is not a residue-by-residue scrutiny of conformation by an omniscient observer but rather the total free energy of the transient species. Second, proteins are only marginally stable. The free-energy difference between the folded and the unfolded states of a typical 100-residue protein is 42 kJ mol^{-1} (10 kcal mol^{-1}), and thus each residue contributes on average only 0.42 kJ mol^{-1} (0.1 kcal mol^{-1}) of energy to maintain the folded state. This amount is less than the amount of thermal energy, which is 2.5 kJ mol^{-1} (0.6 kcal mol^{-1}) at room temperature. This meager stabilization energy means that correct intermediates, especially those formed early in folding, can be lost. The analogy is that the monkey would be somewhat free to undo its correct keystrokes. Nonetheless, the interactions that lead to cooperative folding can stabilize intermediates as structure builds up. Thus, local regions that have significant structural preference, though not necessarily stable on their own, will tend to adopt their favored structures and, as they form, can interact with one other, leading to increasing stabilization. This conceptual framework is often referred to as the *nucleation-condensation model*.

A simulation of the folding of a protein, based on the nucleation-condensation model, is shown in Figure 2.59. This model suggests that certain pathways may be preferred. Although Figure 2.59 suggests a discrete pathway, each of the intermediates shown represents an ensemble of similar structures, and thus a protein follows a general rather than a precise pathway in its transition from the unfolded to the native state. The energy

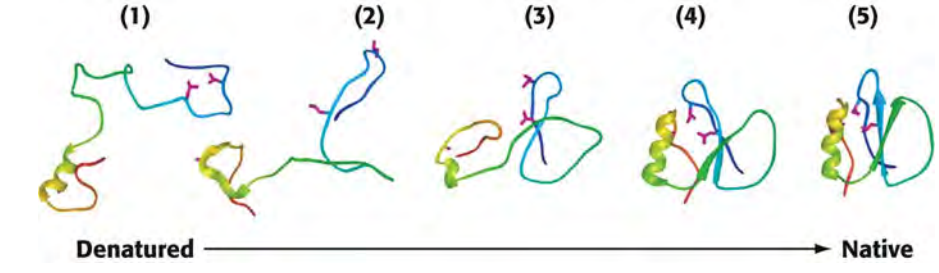


Figure 2.58 Typing-monkey analogy. A monkey randomly poking a typewriter could write a line from Shakespeare's *Hamlet,* provided that correct keystrokes were retained. In the two computer simulations shown, the cumulative number of keystrokes is given at the left of each line.

Figure 2.59 Proposed folding pathway of chymotrypsin inhibitor. Local regions with sufficient structural preference tend to adopt their favored structures initially (1). These structures come together to form a nucleus with a nativelike, but still mobile, structure (4). This structure then fully condenses to form the native, more rigid structure (5). [From A. R. Fersht and V. Daggett. *Cell* 108:573–582, 2002; with permission from Elsevier.]

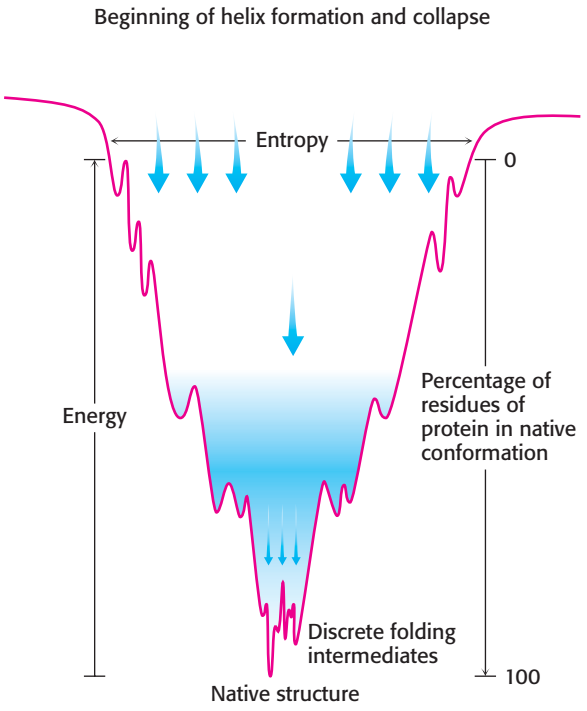


Figure 2.60 Folding funnel. The folding funnel depicts the thermodynamics of protein folding. The top of the funnel represents all possible denatured conformations—that is, maximal conformational entropy. Depressions on the sides of the funnel represent semistable intermediates that can facilitate or hinder the formation of the native structure, depending on their depth. Secondary structures, such as helices, form and collapse onto one another to initiate folding. [After D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry,* 5th ed. (W. H. Freeman and Company, 2008), p. 143.]

surface for the overall process of protein folding can be visualized as a funnel (Figure 2.60). The wide rim of the funnel represents the wide range of structures accessible to the ensemble of denatured protein molecules. As the free energy of the population of protein molecules decreases, the proteins move down into narrower parts of the funnel and fewer conformations are accessible. At the bottom of the funnel is the folded state with its well-defined conformation. Many paths can lead to this same energy minimum.

Prediction of three-dimensional structure from sequence remains a great challenge

The prediction of three-dimensional structure from sequence has proved to be extremely difficult. As we have seen, the local sequence appears to determine only between 60 and 70% of the secondary structure; long-range interactions are required to fix the full secondary structure and the tertiary structure.

Investigators are exploring two fundamentally different approaches to predicting three-dimensional structure from amino acid sequence. The first is *ab initio* (Latin, "from the beginning") *prediction,* which attempts to predict the folding of an amino acid sequence without prior knowledge about similar sequences in known protein structures. Computer-based calculations are employed that attempt to minimize the free energy of a structure with a given amino acid sequence or to simulate the folding process. The utility of these methods is limited by the vast number of possible conformations, the marginal stability of proteins, and the subtle energetics of weak interactions in aqueous solution. The second approach takes advantage of our growing knowledge of the three-dimensional structures of many proteins. In these *knowledge-based methods,* an amino acid sequence of unknown structure is examined for compatibility with known protein structures or fragments therefrom. If a significant match is detected, the known structure can be used as an initial model. Knowledge-based methods have been a source of many insights into the three-dimensional conformation of proteins of known sequence but unknown structure.

Some proteins are inherently unstructured and can exist in multiple conformations

The discussion of protein folding thus far is based on the paradigm that a given protein amino acid sequence will fold into a particular three-dimensional structure. This paradigm holds well for many proteins, such as enzymes and transport proteins. However, it has been known for some time that some proteins can adopt two different structures, one of which results in protein aggregation and pathological conditions (p. 55). Such alternate structures originating from a unique amino acid sequence were thought to be rare, the exception to the paradigm. Recent work has called into question the universality of the idea that each amino acid sequence gives rise to one structure for certain proteins, even under normal cellular conditions.

Our first example is a class of proteins referred to as *intrinsically unstructured proteins* (IUPs). As the name suggests, these proteins, completely or in part, do not have a discrete three-dimensional structure under physiological conditions. Indeed, an estimated 50% of eukaryotic proteins have at least one unstructured region greater than 30 amino acids in length. Unstructured

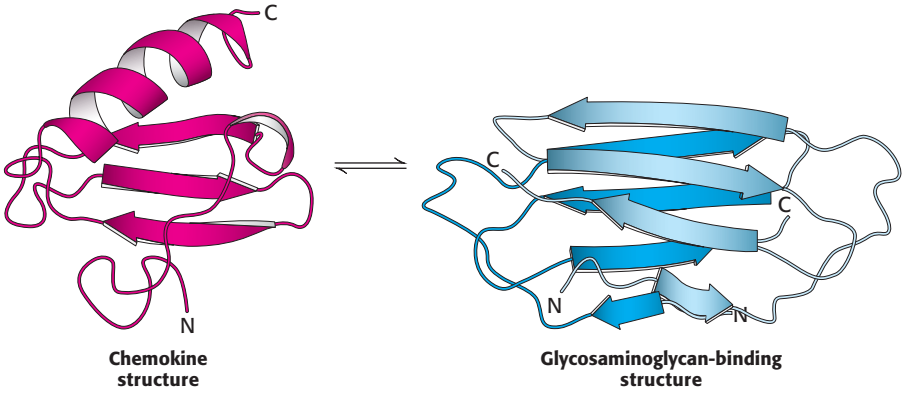


Figure 2.61 Lymphotactin exists in two conformations, which are in equilibrium. [R. L. Tuinstra, F. C. Peterson, S. Kutlesa, E. S. Elgin, M. A. Kron, and B. F. Volkman. *Proc. Natl. Sci. U.S.A.* 105:5057–5062, 2008, Fig. 2A.]

regions are rich in charged and polar amino acids with few hydrophobic residues. These proteins assume a defined structure on interaction with other proteins. This molecular versatility means that one protein can assume different structures and interact with the different partners, yielding different biochemical functions. IUPs appear to be especially important in signaling and regulatory pathways.

Another class of proteins that do not adhere to the paradigm are *metamorphic proteins*. These proteins appear to exist in an ensemble of structures of approximately equal energy that are in equilibrium. Small molecules or other proteins may bind to a particular member of the ensemble, resulting in a complex having a biochemical function that differs from that of another complex formed by the same metamorphic protein bound to a different partner. An especially clear example of a metamorphic protein is the cytokine lymphotactin. Cytokines are signal molecules in the immune system that bind to receptor proteins on the surface of immune-system cells, instigating an immunological response. Lymphotactin exists in two very different structures that are in equilibrium (Figure 2.61). One structure is a characteristic of chemokines, consisting of a three-stranded β sheet and a carboxyl-terminal helix. This structure binds to its receptor and activates it. The alternative structure is an identical dimer of all β sheets. When in this structure, lymphotactin binds to glycosaminglycan, a complex carbohydrate (Chapter 11). The biochemical activities of each structure are mutually exclusive: the cytokine structure cannot bind the glycosaminoglycan, and the β-sheet structure cannot activate the receptor. Yet, remarkably, both activities are required for full biochemical activity of the cytokine.

Note that IUPs and metamorphic proteins effectively expand the protein encoding capacity of the genome. In some cases, a gene can encode a single protein that has more than one structure and function. These examples also illustrate the dynamic nature of the study of biochemistry and its inherent excitement: even well-established ideas are often subject to modifications.

Protein misfolding and aggregation are associated with some neurological diseases

Understanding protein folding and misfolding is of more than academic interest. A host of diseases, including Alzheimer disease, Parkinson disease, Huntington disease, and transmissible spongiform encephalopathies (prion disease), are associated with improperly folded proteins. All of these

diseases result in the deposition of protein aggregates, called *amyloid fibrils* or *plaques*. These diseases are consequently referred to as *amyloidoses*. A common feature of amyloidoses is that normally soluble proteins are converted into insoluble fibrils rich in β sheets. The correctly folded protein is only marginally more stable than the incorrect form. But the incorrect form aggregates, pulling more correct forms into the incorrect form. We will focus on the transmissible spongiform encephalopathies.

One of the great surprises in modern medicine was that certain infectious neurological diseases were found to be transmitted by agents that were similar in size to viruses but consisted only of protein. These diseases include *bovine spongiform encephalopathy* (commonly referred to as *mad cow disease*) and the analogous diseases in other organisms, including *Creutzfeld–Jacob disease* (CJD) in human beings, *scrapie* in sheep, and chronic wasting disease in deer and elk. The agents causing these diseases are termed *prions*. Prions are composed largely or completely of a cellular protein called PrP, which is normally present in the brain but its function has not been identified. Indeed, mice lacking PrP display normal phenotypes. The infectious prions are aggregated forms of the PrP protein termed PrP^{SC}.

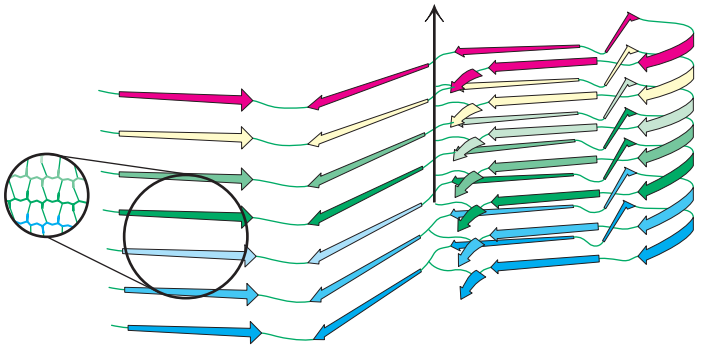


Figure 2.62 A model of the human prion protein amyloid. A detailed model of a human prion amyloid fibril deduced from spin labeling and electron paramagnetic resonance (EPR) spectroscopy studies shows that protein aggregation is due to the formation of large parallel β sheets. The arrow indicates the long axis of the fibril. [N. J. Cobb, F. D. Sönnichsen, H. Mchaourab, and W. K. Surewicz. *Proc. Natl. Acad. Sci. U.S.A.* 104: 18946–18951, 2007, Fig. 4E.]

How does the structure of the protein in the aggregated form differ from that of the protein in its normal state in the brain? The structure of the normal cellular protein PrP contains extensive regions of α helix and relatively little β-strand structure. The structure of the form of the protein present in infected brains, termed PrP^{SC}, has not yet been determined because of challenges posed by its insoluble and heterogeneous nature. However, a variety of evidence indicates that some parts of the protein that had been in α-helical or turn conformations have been converted into β-strand conformations (Figure 2.62). The β strands of largely planar monomers stack on one another with their side chains tightly interwoven. A side view shows the extensive network of hydrogen bonds between the monomers. These fibrous protein aggregates are often referred to as *amyloid* forms.

With the realization that the infectious agent in prion diseases is an aggregated form of a protein that is already present in the brain, a model for disease transmission emerges (Figure 2.63). Protein aggregates built of abnormal forms of PrP act as nuclei to which other PrP molecules attach. Prion diseases can thus be transferred from one individual organism to another through the transfer of an aggregated nucleus, as likely happened in the mad cow disease outbreak

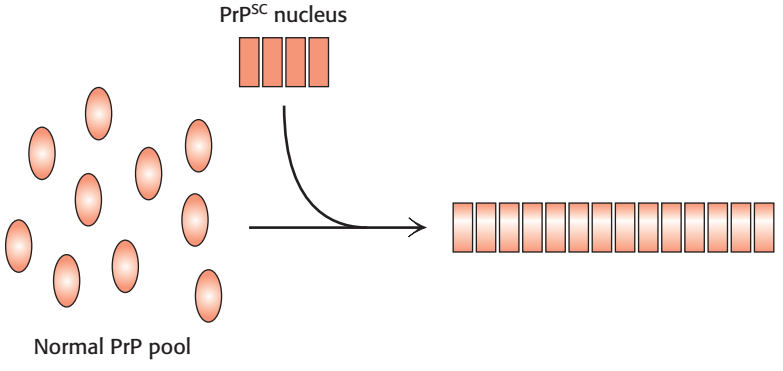


Figure 2.63 The protein-only model for prion-disease transmission. A nucleus consisting of proteins in an abnormal conformation grows by the addition of proteins from the normal pool.

in the United Kingdom in the 1990s. Cattle fed on animal feed containing material from diseased cows developed the disease in turn.

Amyloid fibers are also seen in the brains of patients with certain noninfectious neurodegenerative diseases such as Alzheimer and Parkinson diseases. For example, the brains of patients with Alzheimer disease contain protein aggregates called *amyloid plaques* that consist primarily of a single polypeptide termed Aβ. This polypeptide is derived from a cellular protein called *amyloid precursor protein* (APP) through the action of specific proteases. Polypeptide Aβ is prone to form insoluble aggregates. Despite the difficulties posed by the protein's insolubility, a detailed structural model for Aβ has been derived through the use of NMR techniques that can be applied to solids rather than to materials in solution. As expected, the structure is rich in β strands, which come together to form extended parallel β-sheet structures (see Figure 2.63).

How do such aggregates lead to the death of the cells that harbor them? The answer is still controversial. One hypothesis is that the large aggregates themselves are not toxic but, instead, smaller aggregates of the same proteins may be the culprits, perhaps damaging cell membranes.

Protein modification and cleavage confer new capabilities

Proteins are able to perform numerous functions that rely solely on the versatility of their 20 amino acids. In addition, many proteins are covalently modified, through the attachment of groups other than amino acids, to augment their functions (Figure 2.64). For example, *acetyl groups* are attached to the amino termini of many proteins, a modification that makes these proteins more resistant to degradation. As discussed earlier (p. 44), the addition of *hydroxyl groups* to many proline residues stabilizes fibers of newly synthesized collagen. The biological significance of this modification is evident in the disease scurvy: a deficiency of vitamin C results in insufficient hydroxylation of collagen, and the abnormal collagen fibers that result are unable to maintain normal tissue strength. Another specialized amino acid produced by a finishing touch is *γ-carboxyglutamate*. In vitamin K deficiency, insufficient carboxylation of glutamate in prothrombin, a clotting protein, can lead to hemorrhage (Chapter 10). Many proteins, especially those that are present on the surfaces of cells or are secreted, acquire *carbohydrate units* on specific asparagine residues (see Chapter 11). The addition of sugars makes the proteins more hydrophilic and able to participate in interactions with other proteins. Conversely, the addition of a *fatty acid* to an α-amino group or a cysteine sulfhydryl group produces a more hydrophobic protein.

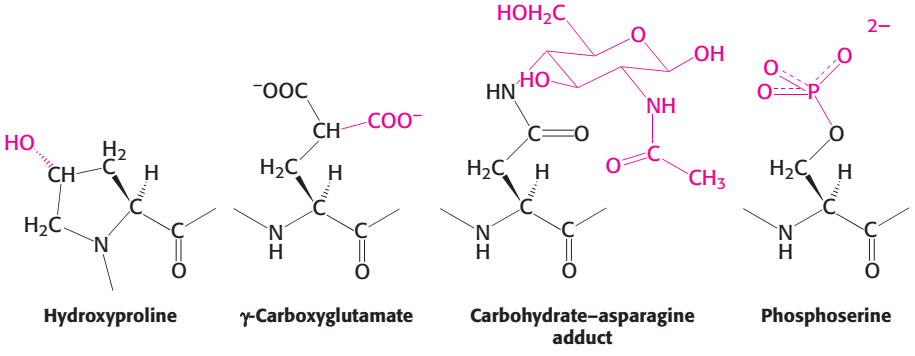


Figure 2.64 Finishing touches. Some common and important covalent modifications of amino acid side chains are shown.

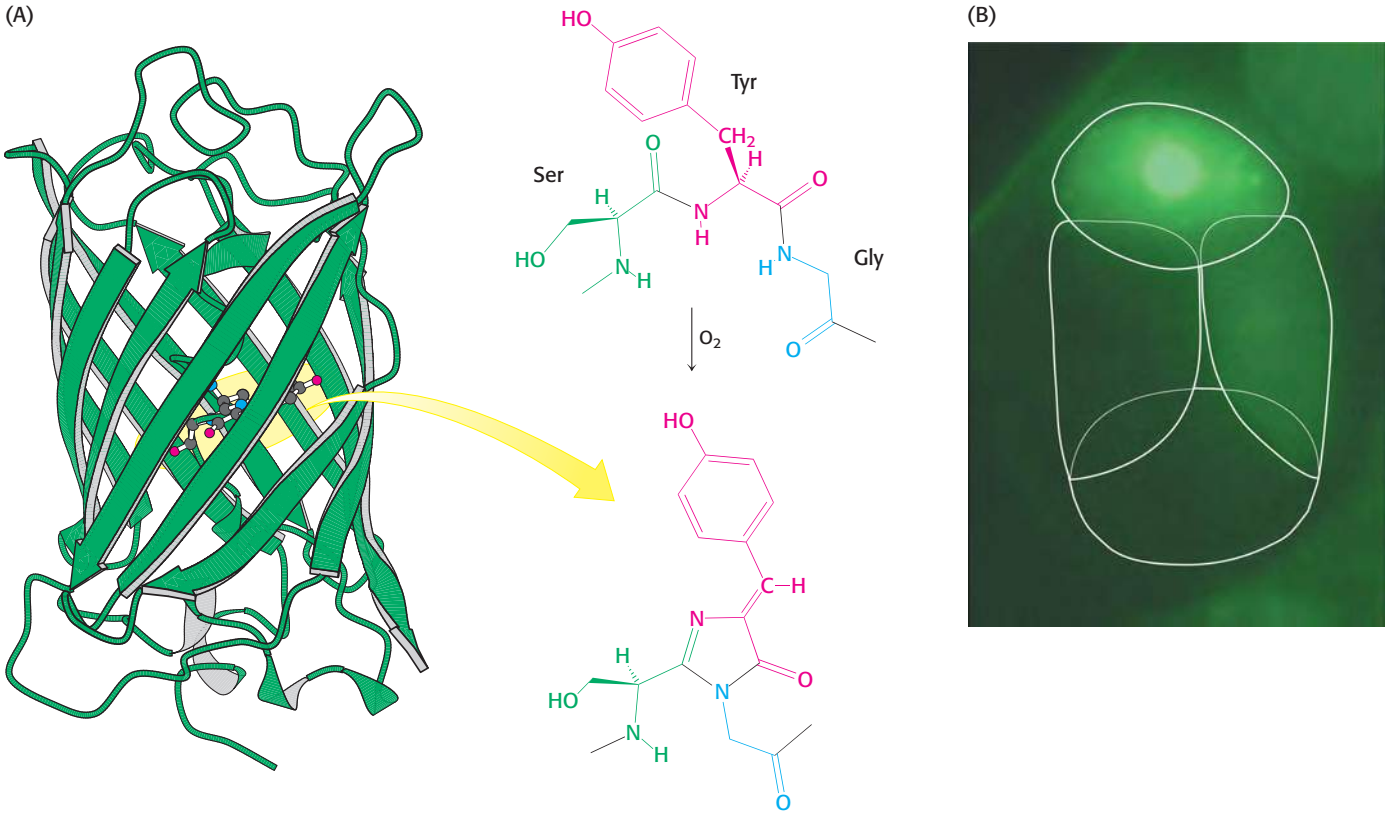


Figure 2.65 Chemical rearrangement in GFP. (A) The structure of green fluorescent protein (GFP). The rearrangement and oxidation of the sequence Ser-Tyr-Gly is the source of fluorescence. (B) Fluorescence micrograph of a four-cell embryo (cells are outlined) from the roundworm *Caenorhabditis elegans* containing a protein, PIE-1, labeled with GFP. The protein is expressed only in the cell (top) that will give rise to the germ line. [(A) Drawn from 1GFL.pdb; (B) courtesy of Dr. Geraldine Seydoux.]

Many hormones, such as epinephrine (adrenaline), alter the activities of enzymes by stimulating the phosphorylation of the hydroxyl amino acids serine and threonine; *phosphoserine* and *phosphothreonine* are the most ubiquitous modified amino acids in proteins. Growth factors such as insulin act by triggering the phosphorylation of the hydroxyl group of tyrosine residues to form *phosphotyrosine*. The phosphoryl groups on these three modified amino acids are readily removed; thus the modified amino acids are able to act as reversible switches in regulating cellular processes. The roles of phosphorylation in signal transduction will be discussed extensively in Chapter 14.

The preceding modifications consist of the addition of special groups to amino acids. Other special groups are generated by chemical rearrangements of side chains and, sometimes, the peptide backbone. For example, certain jellyfish produce a green fluorescent protein (Figure 2.65). The source of the fluorescence is a group formed by the spontaneous rearrangement and oxidation of the sequence Ser-Tyr-Gly within the center of the protein. This protein is of great utility to researchers as a marker within cells.

Finally, many proteins are cleaved and trimmed after synthesis. For example, digestive enzymes are synthesized as inactive precursors that can be stored safely in the pancreas. After release into the intestine, these precursors become activated by peptide-bond cleavage (Section 10.4). In blood clotting, peptide-bond cleavage converts soluble fibrinogen into insoluble fibrin. A number of polypeptide hormones, such as adrenocorticotropic hormone, arise from the splitting of a single large precursor protein. Likewise, many viral proteins are produced by the cleavage of large polyprotein precursors. We shall encounter many more examples of

modification and cleavage as essential features of protein formation and function. Indeed, these finishing touches account for much of the versatility, precision, and elegance of protein action and regulation.

Summary

Protein structure can be described at four levels. The primary structure refers to the amino acid sequence. The secondary structure refers to the conformation adopted by local regions of the polypeptide chain. Tertiary structure describes the overall folding of the polypeptide chain. Finally, quaternary structure refers to the specific association of multiple polypeptide chains to form multisubunit complexes.

2.1 Proteins Are Built from a Repertoire of 20 Amino Acids

Proteins are linear polymers of amino acids. Each amino acid consists of a central tetrahedral carbon atom linked to an amino group, a carboxylic acid group, a distinctive side chain, and a hydrogen atom. These tetrahedral centers, with the exception of that of glycine, are chiral; only the L isomer exists in natural proteins. All natural proteins are constructed from the same set of 20 amino acids. The side chains of these 20 building blocks vary tremendously in size, shape, and the presence of functional groups. They can be grouped as follows: (1) hydrophobic side chains, including the aliphatic amino acids—glycine, alanine, valine, leucine, isoleucine, methionine, and proline—and aromatic side chains—phenylalanine, and tryptophan; (2) polar side chains, including hydroxyl-containing side chains—serine, threonine and tyrosine; the sulfhydryl-containing cysteine; and carboxamide-containing side chains—asparagine and glutamine; (3) basic side chains—lysine, arginine, and histidine; and (4) acidic side chains—aspartic acid and glutamic acid. These groupings are somewhat arbitrary and many other sensible groupings are possible.

2.2 Primary Structure: Amino Acids Are Linked by Peptide Bonds to Form Polypeptide Chains

The amino acids in a polypeptide are linked by amide bonds formed between the carboxyl group of one amino acid and the amino group of the next. This linkage, called a peptide bond, has several important properties. First, it is resistant to hydrolysis, and so proteins are remarkably stable kinetically. Second, the peptide group is planar because the C—N bond has considerable double-bond character. Third, each peptide bond has both a hydrogen-bond donor (the NH group) and a hydrogen-bond acceptor (the CO group). Hydrogen bonding between these backbone groups is a distinctive feature of protein structure. Finally, the peptide bond is uncharged, which allows proteins to form tightly packed globular structures having significant amounts of the backbone buried within the protein interior. Because they are linear polymers, proteins can be described as sequences of amino acids. Such sequences are written from the amino to the carboxyl terminus.

2.3 Secondary Structure: Polypeptide Chains Can Fold into Regular Structures Such As the Alpha Helix, the Beta Sheet, and Turns and Loops

Two major elements of secondary structure are the α helix and the β strand. In the α helix, the polypeptide chain twists into a tightly packed rod. Within the helix, the CO group of each amino acid is hydrogen bonded to the NH group of the amino acid four residues farther along the polypeptide chain. In the β strand, the polypeptide chain

is nearly fully extended. Two or more β strands connected by NH-to-CO hydrogen bonds come together to form β sheets. The strands in β sheets can be antiparallel, parallel, or mixed.

2.4 Tertiary Structure: Water-Soluble Proteins Fold into Compact Structures with Nonpolar Cores

The compact, asymmetric structure that individual polypeptides attain is called tertiary structure. The tertiary structures of water-soluble proteins have features in common: (1) an interior formed of amino acids with hydrophobic side chains and (2) a surface formed largely of hydrophilic amino acids that interact with the aqueous environment. The hydrophobic interactions between the interior residues are the driving force for the formation of the tertiary structure of water-soluble proteins. Some proteins that exist in a hydrophobic environment, such as in membranes, display the inverse distribution of hydrophobic and hydrophilic amino acids. In these proteins, the hydrophobic amino acids are on the surface to interact with the environment, whereas the hydrophilic groups are shielded from the environment in the interior of the protein.

2.5 Quaternary Structure: Polypeptide Chains Can Assemble into Multisubunit Structures

Proteins consisting of more than one polypeptide chain display quaternary structure; each individual polypeptide chain is called a subunit. Quaternary structure can be as simple as two identical subunits or as complex as dozens of different subunits. In most cases, the subunits are held together by noncovalent bonds.

2.6 The Amino Acid Sequence of a Protein Determines Its Three-Dimensional Structure

The amino acid sequence determines the three-dimensional structure and, hence, all other properties of a protein. Some proteins can be unfolded completely yet refold efficiently when placed under conditions in which the folded form of the protein is stable. The amino acid sequence of a protein is determined by the sequences of bases in a DNA molecule. This one-dimensional sequence information is extended into the three-dimensional world by the ability of proteins to fold spontaneously. Protein folding is a highly cooperative process; structural intermediates between the unfolded and folded forms do not accumulate.

Some proteins, such as intrinsically unstructured proteins and metamorphic proteins, do not strictly adhere to the one-sequence–one-structure paradigm. Because of this versatility, these proteins expand the protein encoding capacity of the genome.

The versatility of proteins is further enhanced by covalent modifications. Such modifications can incorporate functional groups not present in the 20 amino acids. Other modifications are important to the regulation of protein activity. Through their structural stability, diversity, and chemical reactivity, proteins make possible most of the key processes associated with life.

APPENDIX: Visualizing Molecular Structures II: Proteins

Scientists have developed powerful techniques for the determination of protein structures, as will be considered in Chapter 3. In most cases, these techniques allow the positions of the thousands of atoms within a protein structure to be determined. The final results from such an experiment include the x, y, and z coordinates for each atom in the structure. These coordinate files are compiled in the Protein Data Bank (http://www.pdb.org) from which they can be readily downloaded. These structures comprise thousands or even tens of thousands

of atoms. The complexity of proteins with thousands of atoms presents a challenge for the depiction of their structure. Several different types of representations are used to portray proteins, each with its own strengths and weaknesses. The types that you will see most often in this book are space-filling models, ball-and-stick models, backbone models, and ribbon diagrams. Where appropriate, structural features of particular importance or relevance are noted in an illustration's legend.

Space-Filling Models

Space-filling models are the most realistic type of representation. Each atom is shown as a sphere with a size corresponding to the van der Waals radius of the atom (Section 1.3). Bonds are not shown explicitly but are represented by the intersection of the spheres shown when atoms are closer together than the sum of their van der Waals radii. All atoms are shown, including those that make up the backbone and those in the side chains. A space-filling model of lysozyme is depicted in Figure 2.66.

Space-filling models convey a sense of how little open space there is in a protein's structure, which always has many atoms in van der Waals contact with one another. These models are particularly useful in showing conformational changes in a protein from one set of circumstances to another. A disadvantage of space-filling models is that the secondary and tertiary structures of the protein are difficult to see. Thus, these models are not very effective in distinguishing one protein from another—many space-filling models of proteins look very much alike.

Ball-and-Stick Models

Ball-and-stick models are not as realistic as space-filling models. Realistically portrayed atoms occupy more space, determined by their van der Waals radii, than do the atoms depicted in ball-and-stick models. However, the bonding arrangement is easier to see because the bonds are explicitly represented as sticks (Figure 2.67). A ball-and-stick model reveals a complex structure more clearly than a space-filling model does. However, the depiction is so complicated that structural features such as α helices or potential binding sites are difficult to discern.

Because space-filling and ball-and-stick models depict protein structures at the atomic level, the large number of atoms in a complex structure makes it difficult to discern the relevant structural features. Thus, representations that are more schematic—such as backbone models and ribbon diagrams—have been developed for the depiction of macromolecular structures. In these representations, most or all atoms are not shown explicitly.

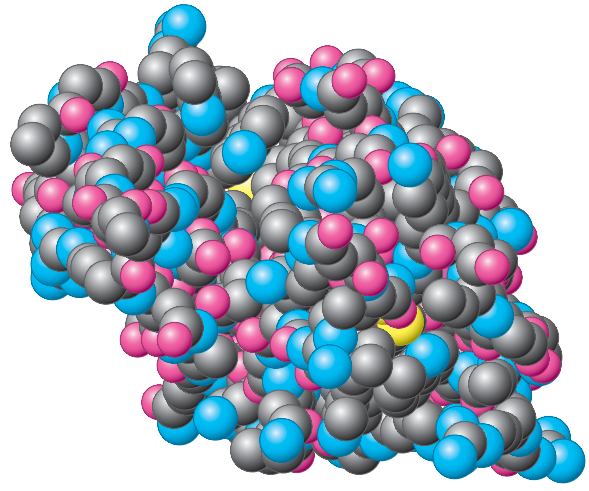


Figure 2.66 Space-filling model of lysozyme. *Notice* how tightly packed the atoms are, with little unfilled space. All atoms are shown with the exception of hydrogen atoms. Hydrogen atoms are often omitted because their positions are not readily determined by x-ray crystallographic methods and because their omission somewhat improves the clarity of the structure's depiction.

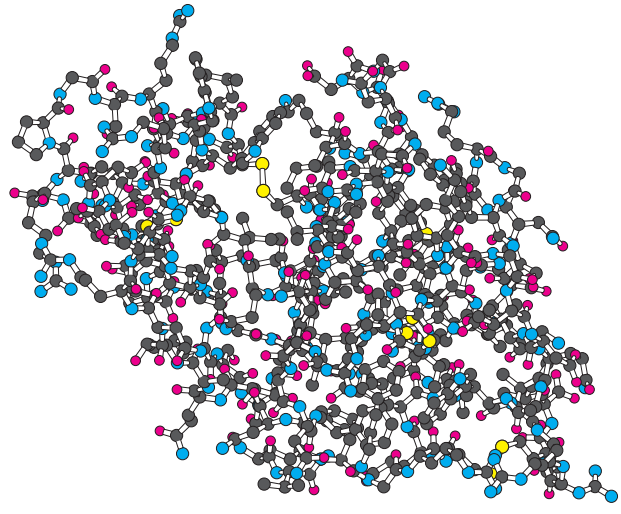


Figure 2.67 Ball-and-stick model of lysozyme. Again, hydrogen atoms are omitted.

Backbone Models

Backbone models show only the backbone atoms of a molecule's polypeptide or even only the α-carbon atom of each amino acid. Atoms are linked by lines representing bonds; if only α-carbon atoms are depicted, lines connect α-carbon atoms of amino acids that are adjacent in the amino acid sequence (Figure 2.68). In this book, backbone models show only the lines

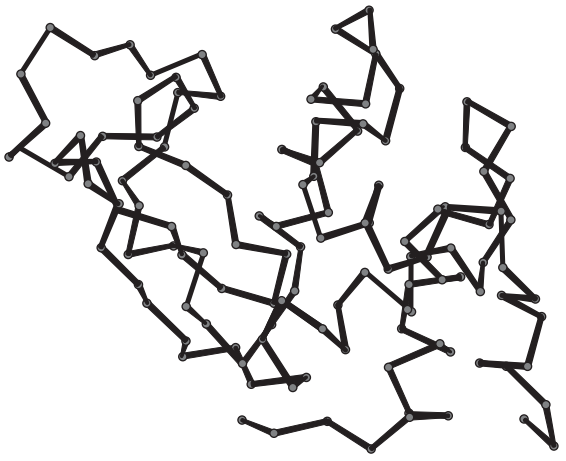


Figure 2.68 Backbone model of lysozyme.

connecting the α-carbon atoms; other carbon atoms are not depicted.

A backbone model shows the overall course of the polypeptide chain much better than a space-filling or ball-and-stick model does. However, secondary structural elements are still difficult to see.

Ribbon Diagrams

Ribbon diagrams are highly schematic and most commonly used to accent a few dramatic aspects of protein structure, such as the α helix (depicted as a coiled ribbon or a cylinder), the β strand (a broad arrow), and loops (thin tubes), to provide clear views of the folding patterns of proteins (Figure 2.69). The ribbon diagram allows the course of a polypeptide chain to be traced and readily shows the secondary structural elements. Thus, ribbon diagrams of proteins that are related to one another by evolutionary divergence appear similar (see Figure 6.14), whereas unrelated proteins are clearly distinct.

In this book, coiled ribbons will be generally used to depict α helices. However, for membrane proteins, which are often quite complex, cylinders will be used rather than coiled ribbons. This convention will also make membrane proteins with their membrane-spanning α helices easy to recognize (see Figure 12.18).

Bear in mind that the open appearance of ribbon diagrams is deceptive. As noted earlier, protein structures are tightly packed and have little open space. The openness of ribbon diagrams makes them particularly useful as frameworks in which to highlight additional aspects of protein structure. Active sites, substrates, bonds, and other structural fragments can be included in ball-and-stick or space-filling form within a ribbon diagram (Figure 2.70).

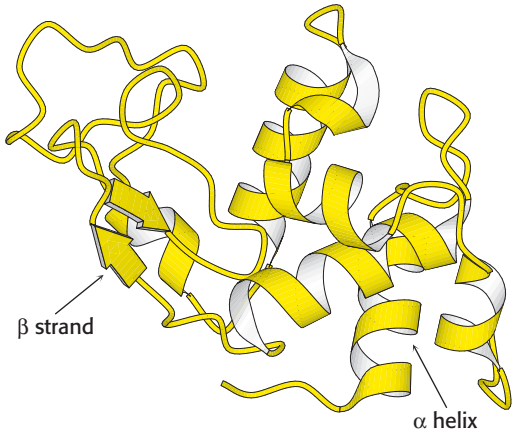


Figure 2.69 Ribbon diagram of lysozyme. The α helices are shown as coiled ribbons; β strands are depicted as arrows. More irregular structures are shown as thin tubes.

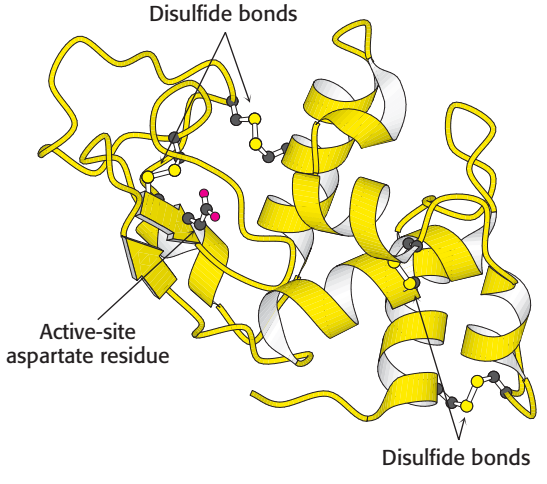


Figure 2.70 Ribbon diagram of lysozyme with highlights. Four disulfide bonds and a functionally important aspartate residue are shown in ball-and-stick form.

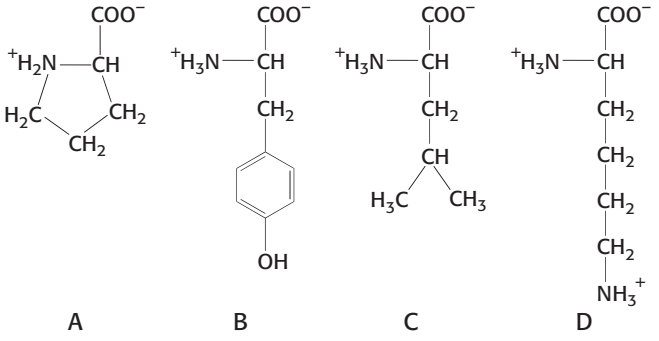
Key Terms

side chain (R group) (p. 27)
L amino acid (p. 27)
dipolar ion (zwitterion) (p. 27)
peptide bond (amide bond) (p. 33)
disulfide bond (p. 35)
primary structure (p. 35)
torsion angle (p. 37)
phi (ϕ) angle (p. 37)
psi (ψ) angle (p. 37)
Ramachandran diagram (p. 37)
secondary structure (p. 38)
α helix (p. 38)
rise (translation) (p. 39)
β pleated sheet (p. 40)
β strand (p. 40)

reverse turn (β turn; hairpin turn) (p. 42)
coiled coil (p.43)
heptad repeat (p. 43)
tertiary structure (p. 45)
motif (supersecondary structure) (p. 47)
domain (p. 47)
subunit (p. 48)
quaternary structure (p. 48)
cooperative transition (p. 52)
intrinsically unstructured protein (IUP) (p. 54)
metamorphic protein (p. 55)
prion (p. 56)

Problems

1. *Identify.* Examine the following four amino acids (A–D):



What are their names, three-letter abbreviations, and one-letter symbols?

2. *Properties.* In reference to the amino acids shown in Problem 1, which are associated with the following characteristics?

(a) Hydrophobic side chain _____

(b) Basic side chain _____

(c) Three ionizable groups _____

(d) pK_a of approximately 10 in proteins _____

(e) Modified form of phenylalanine _____

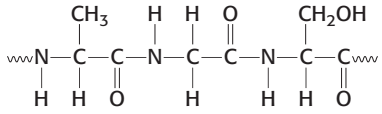
3. *Match 'em.* Match each amino acid in the left-hand column with the appropriate side-chain type in the right-hand column.

(a) Leu	(1) hydroxyl-containing
(b) Glu	(2) acidic
(c) Lys	(3) basic
(d) Ser	(4) sulfur-containing
(e) Cys	(5) nonpolar aromatic
(f) Trp	(6) nonpolar aliphatic

4. *Solubility.* In each of the following pairs of amino acids, identify which amino acid would be most soluble in water: (a) Ala, Leu; (b) Tyr, Phe; (c) Ser, Ala; (d) Trp, His.

5. *Bonding is good.* Which of the following amino acids have R groups that have hydrogen-bonding potential? Ala, Gly, Ser, Phe, Glu, Tyr, Ile, and Thr.

6. *Name those components.* Examine the segment of a protein shown here.



(a) What three amino acids are present?

(b) Of the three, which is the N-terminal amino acid?

(c) Identify the peptide bonds.

(d) Identify the α-carbon atoms.

7. *Who's charged?* Draw the structure of the dipeptide Gly-His. What is the charge on the peptide at pH 5.5? pH 7.5?

8. *Alphabet soup.* How many different polypeptides of 50 amino acids in length can be made from the 20 common amino acids?

9. *Sweet tooth, but calorie conscious.* Aspartame (NutraSweet), an artificial sweetener, is a dipeptide composed of Asp-Phe in which the carboxyl terminus is modified by the attachment of a methyl group. Draw the structure of Aspartame at pH 7.

10. *Vertebrate proteins?* What is meant by the term *polypeptide backbone?*

11. *Not a sidecar.* Define the term *side chain* in the context of amino acid or protein structure.

12. *One from many.* Differentiate between *amino acid composition* and *amino acid sequence*.

13. *Shape and dimension.* (a) Tropomyosin, a 70-kd muscle protein, is a two-stranded α-helical coiled coil. Estimate the length of the molecule. (b) Suppose that a 40-residue segment of a protein folds into a two-stranded antiparallel β structure with a 4-residue hairpin turn. What is the longest dimension of this motif?

14. *Contrasting isomers.* Poly-L-leucine in an organic solvent such as dioxane is α helical, whereas poly-L-isoleucine is not. Why do these amino acids with the same number and kinds of atoms have different helix-forming tendencies?

15. *Active again.* A mutation that changes an alanine residue in the interior of a protein to valine is found to lead to a

loss of activity. However, activity is regained when a second mutation at a different position changes an isoleucine residue to glycine. How might this second mutation lead to a restoration of activity?

16. *Shuffle test*. An enzyme that catalyzes disulfide–sulfhydryl exchange reactions, called protein disulfide isomerase (PDI), has been isolated. PDI rapidly converts inactive scrambled ribonuclease into enzymatically active ribonuclease. In contrast, insulin is rapidly inactivated by PDI. What does this important observation imply about the relation between the amino acid sequence of insulin and its three-dimensional structure?

17. *Stretching a target*. A protease is an enzyme that catalyzes the hydrolysis of the peptide bonds of target proteins. How might a protease bind a target protein so that its main chain becomes fully extended in the vicinity of the vulnerable peptide bond?

18. *Often irreplaceable*. Glycine is a highly conserved amino acid residue in the evolution of proteins. Why?

19. *Potential partners*. Identify the groups in a protein that can form hydrogen bonds or electrostatic bonds with an arginine side chain at pH 7.

20. *Permanent waves*. The shape of hair is determined in part by the pattern of disulfide bonds in keratin, its major protein. How can curls be induced?

21. *Location is everything 1*. Most proteins have hydrophilic exteriors and hydrophobic interiors. Would you expect this structure to apply to proteins embedded in the hydrophobic interior of a membrane? Explain.

22. *Location is everything 2*. Proteins that span biological membranes often contain α helices. Given that the insides of membranes are highly hydrophobic (Section 12.2), predict what type of amino acids would be in such a helix. Why is an α helix particularly suited to existence in the hydrophobic environment of the interior of a membrane?

23. *Neighborhood peer pressure?* Table 2.1 shows the typical pK_a values for ionizable groups in proteins. However, more than 500 pK_a values have been determined for individual groups in folded proteins. Account for this discrepancy.

24. *Maybe size does matter*. Osteo imperfecta displays a wide range of symptoms, from mild to severe. On the basis of your knowledge of amino acid and collagen structure, propose a biochemical basis for the variety of symptoms.

25. *Issues of stability*. Proteins are quite stable. The lifetime of a peptide bond in aqueous solution is nearly 1000 years. However, the free energy of hydrolysis of proteins is negative and quite large. How can you account for the stability of the peptide bond in light of the fact that hydrolysis releases much energy?

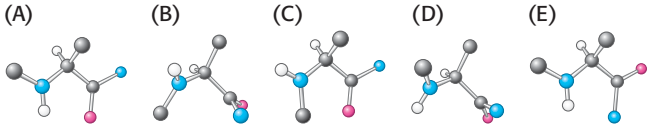
26. *Minor species*. For an amino acid such as alanine, the major species in solution at pH 7 is the zwitterionic form. Assume a pK_a value of 8 for the amino group and a pK_a value of 3 for the carboxylic acid. Estimate the ratio of the concentration of the neutral amino acid species (with the carboxylic acid protonated and the amino group neutral) to that of the zwitterionic species at pH 7 (see Section 1.3).

27. *A matter of convention*. All L amino acids have an *S* absolute configuration except L-cysteine, which has the *R* configuration. Explain why L-cysteine is designated as having the *R* absolute configuration.

28. *Hidden message*. Translate the following amino acid sequence into one-letter code: Glu-Leu-Val-Ile-Ser-Ile-Ser-Leu-Ile-Val-Ile-Asn-Gly-Ile-Asn-Leu-Ala-Ser-Val-Glu-Gly-Ala-Ser.

29. *Who goes first?* Would you expect Pro—X peptide bonds to tend to have cis conformations like those of X—Pro bonds? Why or why not?

30. *Matching*. For each of the amino acid derivatives shown here (A–E), find the matching set of ϕ and ψ values (a–e).

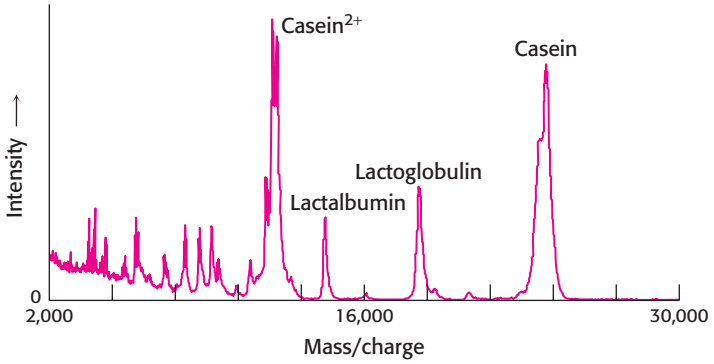


(a)	(b)	(c)	(d)	(e)
$\phi = 120°$, $\psi = 120°$	$\phi = 180°$, $\psi = 0°$	$\phi = 180°$, $\psi = 180°$	$\phi = 0°$, $\psi = 180°$	$\phi = -60°$, $\psi = -40°$

31. *Scrambled ribonuclease.* When performing his experiments on protein refolding, Christian Anfinsen obtained a quite different result when reduced ribonuclease was reoxidized while it was still in 8 M urea and the preparation was then dialyzed to remove the urea. Ribonuclease reoxidized in this way had only 1% of the enzymatic activity of the native protein. Why were the outcomes so different when reduced ribonuclease was reoxidized in the presence and absence of urea?

CHAPTER 3

Exploring Proteins and Proteomes



Milk, a source of nourishment for all mammals, is composed, in part, of a variety of proteins. The protein components of milk are revealed by the technique of MALDI–TOF mass spectrometry, which separates molecules on the basis of their mass-to-charge ratio. [(Left) Okea/istockphoto.com. (Right) Courtesy of Dr. Brian Chait.]

Proteins play crucial roles in nearly all biological processes—in catalysis, signal transmission, and structural support. This remarkable range of functions arises from the existence of thousands of proteins, each folded into a distinctive three-dimensional structure that enables it to interact with one or more of a highly diverse array of molecules. A major goal of biochemistry is to determine how amino acid sequences specify the conformations, and hence functions, of proteins. Other goals are to learn how individual proteins bind specific substrates and other molecules, mediate catalysis, and transduce energy and information.

It is often preferable to study a protein of interest after it has been separated from other components within the cell so that the structure and function of this protein can be probed without any confounding effects from contaminants. Hence, the first step in these studies is the purification of the protein of interest. Proteins can be separated from one another on the basis of solubility, size, charge, and binding ability. After a protein has been purified, its amino acid sequence can be determined. Automated peptide sequencing and the application of recombinant DNA methods are providing a wealth of amino acid sequence data that are opening new vistas. Many protein sequences, often deduced from genome sequences, are now available in vast sequence databases. If the sequence of a purified protein has been archived in a publicly searchable database, the job of the investigator becomes much easier. The investigator need determine only a small stretch of amino acid sequence of the protein to find its match in the database.

OUTLINE

Alternatively, such a protein might be identified by matching its mass to those deduced for proteins in the database. Mass spectrometry provides a powerful method for determining the mass of a protein.

After a protein has been purified and its identity confirmed, the challenge remains to determine its function within a physiologically relevant context. Antibodies are choice probes for locating proteins in vivo and measuring their quantities. Monoclonal antibodies, able to recognize specific proteins, can be obtained in large amounts and used to detect and quantify the protein both in isolation and in cells. Peptides and proteins can be chemically synthesized, providing tools for research and, in some cases, highly pure proteins for use as drugs. Finally, x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the principal techniques for elucidating three-dimensional structure, the key determinant of function.

The exploration of proteins by this array of physical and chemical techniques has greatly enriched our understanding of the molecular basis of life. These techniques make it possible to tackle some of the most challenging questions of biology in molecular terms.

The proteome is the functional representation of the genome

As will be discussed in Chapter 5, the complete DNA base sequences, or *genomes,* of many organisms are now available. For example, the roundworm *Caenorhabditis elegans* has a genome of 97 million bases and about 19,000 protein-encoding genes, whereas that of the fruit fly *Drosophila melanogaster* contains 180 million bases and about 14,000 genes. The completely sequenced human genome contains 3 billion bases and about 23,000 genes. However, these genomes are simply inventories of the genes that *could* be expressed within a cell under specific conditions. Only a subset of the proteins encoded by these genes will actually be present in a given biological context. The *proteome*—derived from *prote*ins expressed by the gen*ome*—of an organism signifies a more complex level of information content, encompassing the types, functions, and interactions of proteins within its biological environment.

The proteome is not a fixed characteristic of the cell. Because it represents the functional expression of information, it varies with cell type, developmental stage, and environmental conditions, such as the presence of hormones. The proteome is much larger than the genome because almost all gene products are proteins that can be chemically modified in a variety of ways. Furthermore, these proteins do not exist in isolation; they often interact with one another to form complexes with specific functional properties. Whereas the genome is "hard wired," the proteome is highly dynamic. An understanding of the proteome is acquired by investigating, characterizing, and cataloging proteins. In some, but not all, cases, this process begins by separating a particular protein from all other biomolecules in the cell.

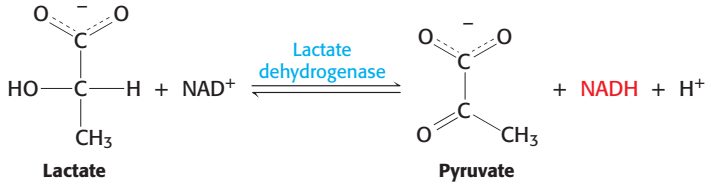
3.1 The Purification of Proteins Is an Essential First Step in Understanding Their Function

An adage of biochemistry is "Never waste pure thoughts on an impure protein." Starting from pure proteins, we can determine amino acid sequences and investigate a protein's biochemical function. From the amino acid sequences, we can map evolutionary relationships between proteins in diverse organisms (Chapter 6). By using crystals grown from pure protein, we can obtain x-ray data that will provide us with a picture of the protein's tertiary structure—the shape that determines function.

The assay: How do we recognize the protein that we are looking for?

Purification should yield a sample containing only one type of molecule—the protein in which the biochemist is interested. This protein sample may be only a fraction of 1% of the starting material, whether that starting material consists of one type of cell in culture or a particular organ from a plant or animal. How is the biochemist able to isolate a particular protein from a complex mixture of proteins?

A protein can be purified by subjecting the impure mixture of the starting material to a series of separations based on physical properties such as size and charge. To monitor the success of this purification, the biochemist needs a test, called an *assay,* for some unique identifying property of the protein. A positive result on the assay indicates that the protein is present. Although assay development can be a challenging task, the more specific the assay, the more effective the purification. For enzymes, which are protein catalysts (Chapter 8), the assay usually measures *enzyme activity*—that is, the ability of the enzyme to promote a particular chemical reaction. This activity is often measured indirectly. Consider the enzyme lactate dehydrogenase, which catalyzes the following reaction in the synthesis of glucose:



Reduced nicotinamide adenine dinucleotide (NADH, see Figure 15.13) absorbs light at 340 nm, whereas oxidized nicotinamide adenine dinucleotide (NAD^+) does not. Consequently, we can follow the progress of the reaction by examining how much light-absorbing ability is developed by a sample in a given period of time—for instance, within 1 minute after the addition of the enzyme. Our assay for enzyme activity during the purification of lactate dehydrogenase is thus the increase in the absorbance of light at 340 nm observed in 1 minute.

To analyze how our purification scheme is working, we need one additional piece of information—the amount of protein present in the mixture being assayed. There are various rapid and reasonably accurate means of determining protein concentration. With these two experimentally determined numbers—enzyme activity and protein concentration—we then calculate the *specific activity,* the ratio of enzyme activity to the amount of protein in the mixture. Ideally, the specific activity will rise as the purification proceeds and the protein mixture being assayed consists to a greater and greater extent of lactate dehydrogenase. In essence, the overall goal of the purification is to maximize the specific activity. For a pure enzyme, the specific activity will have a constant value.

Proteins must be released from the cell to be purified

Having found an assay and chosen a source of protein, we now fractionate the cell into components and determine which component is enriched in the protein of interest. In the first step, a *homogenate* is formed by disrupting the cell membrane, and the mixture is fractionated by centrifugation, yielding a dense pellet of heavy material at the bottom of the centrifuge tube and a lighter supernatant above (Figure 3.1). The supernatant is

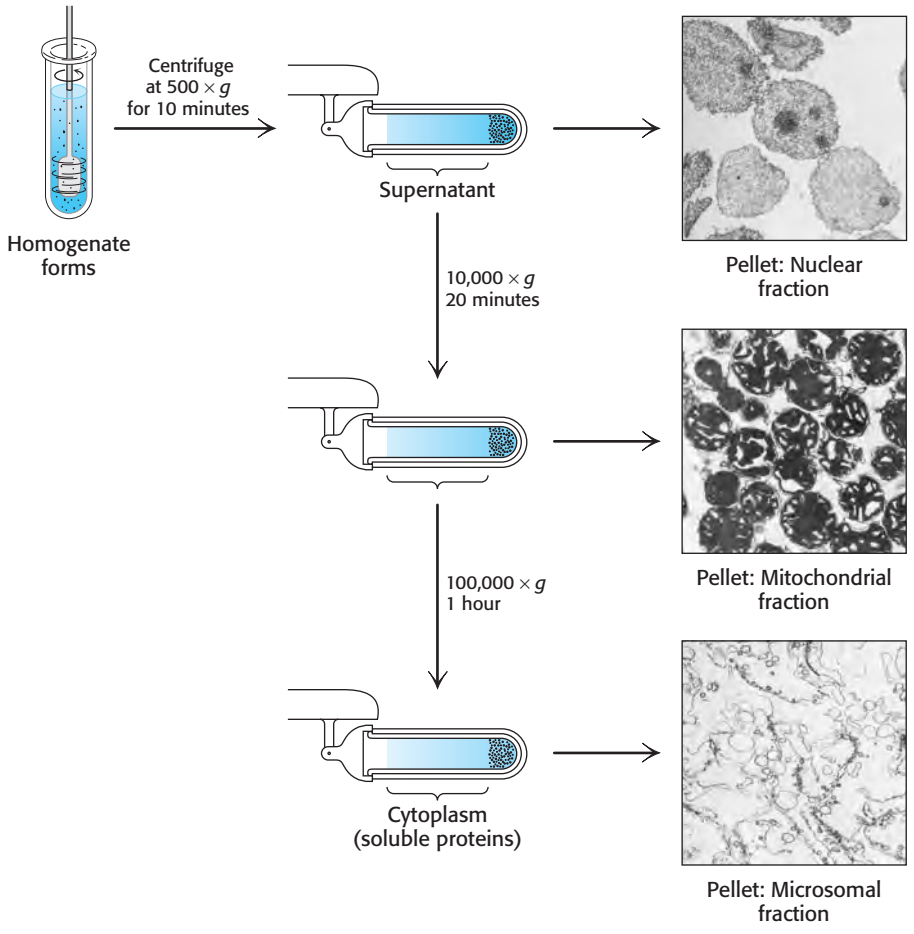


Figure 3.1 Differential centrifugation. Cells are disrupted in a homogenizer and the resulting mixture, called the homogenate, is centrifuged in a step-by-step fashion of increasing centrifugal force. The denser material will form a pellet at lower centrifugal force than will the less-dense material. The isolated fractions can be used for further purification. [Photographs courtesy of Dr. S. Fleischer and Dr. B. Fleischer.]

again centrifuged at a greater force to yield yet another pellet and supernatant. The procedure, called *differential centrifugation*, yields several fractions of decreasing density, each still containing hundreds of different proteins. The fractions are each separately assayed for the desired activity. Usually, one fraction will be enriched for such activity, and it then serves as the source of material to which more-discriminating purification techniques are applied.

Proteins can be purified according to solubility, size, charge, and binding affinity

Several thousand proteins have been purified in active form on the basis of such characteristics as *solubility, size, charge,* and *specific binding affinity*. Usually, protein mixtures are subjected to a series of separations, each based on a different property. At each step in the purification, the preparation is assayed and its specific activity is determined. A variety of purification techniques are available.

Salting out. Most proteins are less soluble at high salt concentrations, an effect called *salting out*. The salt concentration at which a protein precipitates differs from one protein to another. Hence, salting out can be used to fractionate proteins. For example, 0.8 M ammonium sulfate precipitates fibrinogen, a blood-clotting protein, whereas a concentration of 2.4 M is needed to precipitate serum albumin. Salting out is also useful for concentrating dilute solutions of proteins, including active fractions obtained from other purification steps. Dialysis can be used to remove the salt if necessary.

Dialysis. Proteins can be separated from small molecules such as salt by *dialysis* through a semipermeable membrane, such as a cellulose membrane with pores (Figure 3.2). The protein mixture is placed inside the dialysis bag, which is then submerged in a buffer solution that is devoid of the small molecules to be separated away. Molecules having dimensions significantly greater than the pore diameter are retained inside the dialysis bag. Smaller molecules and ions capable of passing through the pores of the membrane diffuse down their concentration gradients and emerge in the solution outside the bag. This technique is useful for removing a salt or other small molecule from a cell fractionate, but it will not distinguish between proteins effectively.

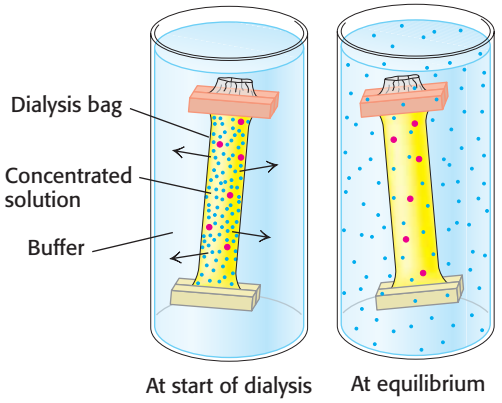


Figure 3.2 Dialysis. Protein molecules (red) are retained within the dialysis bag, whereas small molecules (blue) diffuse down their concentration gradient into the surrounding medium.

Gel-filtration chromatography. More-discriminating separations on the basis of size can be achieved by the technique of *gel-filtration chromatography,* also known as molecular exclusion chromatography (Figure 3.3). The sample is applied to the top of a column consisting of porous beads made of an insoluble but highly hydrated polymer such as dextran or agarose (which are carbohydrates) or polyacrylamide. Sephadex, Sepharose, and Biogel are commonly used commercial preparations of these beads, which are typically 100 μm (0.1 mm) in diameter. Small molecules can enter these beads, but large ones cannot. The result is that small molecules are distributed in the aqueous solution both inside the beads and between them, whereas large molecules are located only in the solution between the beads. *Large molecules flow more rapidly through this column and emerge first because a smaller volume is accessible to them.* Molecules that are of a size to occasionally enter a bead will flow from the column at an intermediate position, and small molecules, which take a longer, tortuous path, will exit last.

Ion-exchange chromatography. To obtain a protein of high purity, one chromatography step is usually not sufficient, because other proteins in the crude mixture will likely co-elute with the desired material. Additional

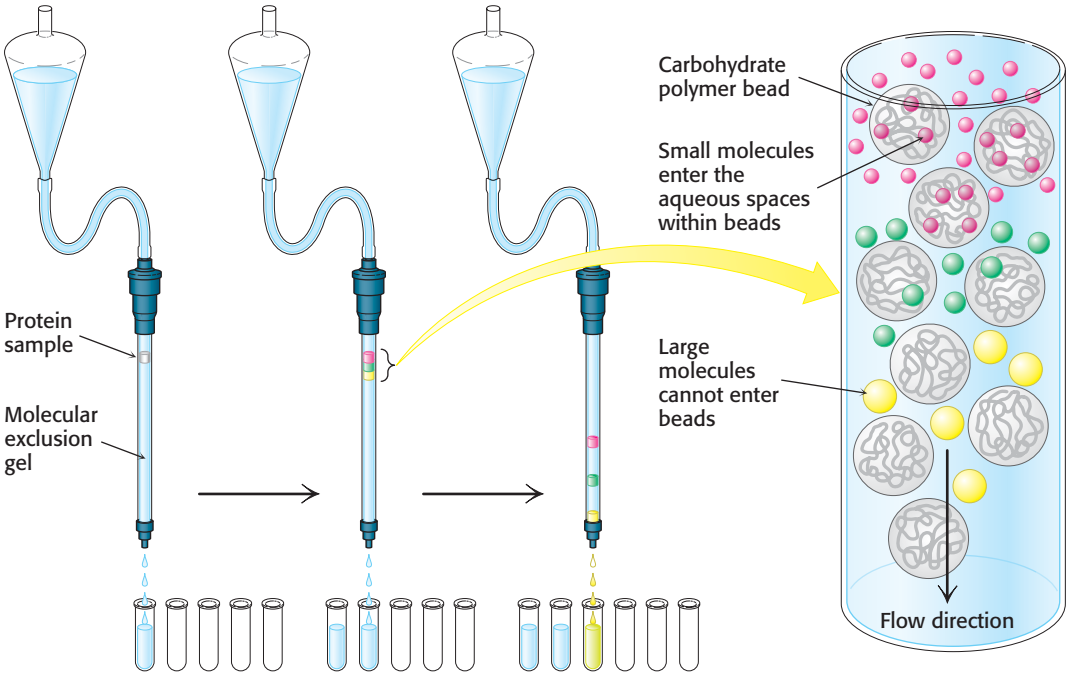


Figure 3.3 Gel-filtration chromatography. A mixture of proteins in a small volume is applied to a column filled with porous beads. Because large proteins cannot enter the internal volume of the beads, they emerge sooner than do small ones.

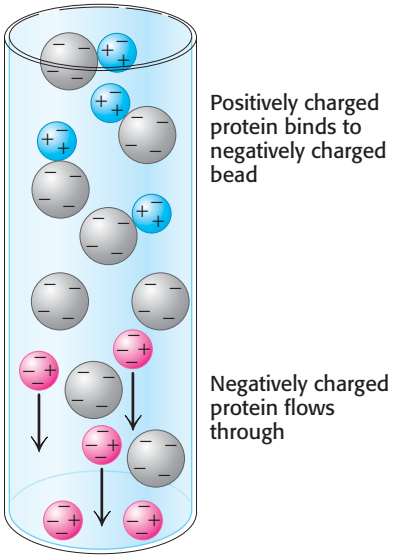
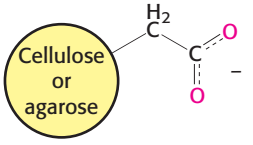
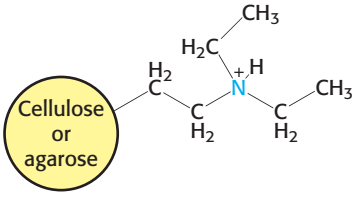


Figure 3.4 Ion-exchange chromatography. This technique separates proteins mainly according to their net charge.

purity can be achieved by performing sequential separations that are based on distinct molecular properties. For example, in addition to size, proteins can be separated on the basis of their net charge by *ion-exchange chromatography*. If a protein has a net positive charge at pH 7, it will usually bind to a column of beads containing carboxylate groups, whereas a negatively charged protein will not (Figure 3.4). The bound protein can then be eluted (released) by increasing the concentration of sodium chloride or another salt in the eluting buffer; sodium ions compete with positively charged groups on the protein for binding to the column. Proteins that have a low density of net positive charge will tend to emerge first, followed by those having a higher charge density. This procedure is also referred to as *cation exchange* to indicate that positively charged groups will bind to the anionic beads. Positively charged proteins (cationic proteins) can be separated by chromatography on negatively charged carboxymethylcellulose (CM-cellulose) columns. Conversely, negatively charged proteins (anionic proteins) can be separated by *anion exchange* on positively charged diethylaminoethylcellulose (DEAE-cellulose) columns.



Carboxymethyl (CM) group (ionized form)



Diethylaminoethyl (DEAE) group (protonated form)

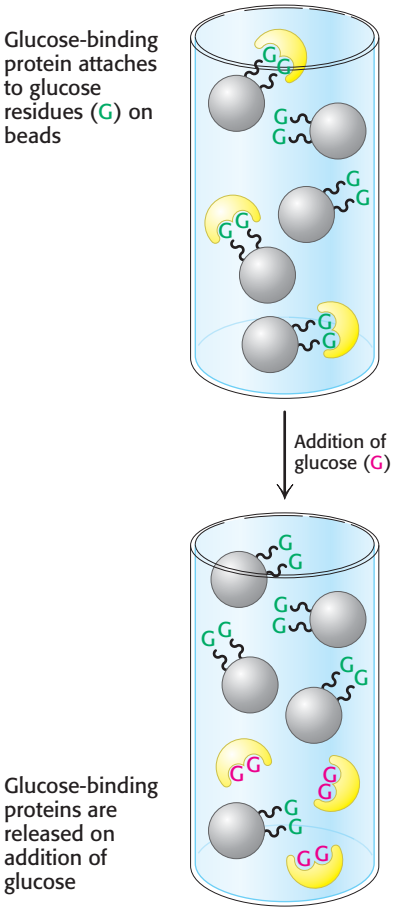


Figure 3.5 Affinity chromatography. Affinity chromatography of concanavalin A (shown in yellow) on a solid support containing covalently attached glucose residues (G).

Affinity chromatography. *Affinity chromatography* is another powerful means of purifying proteins that is highly selective for the protein of interest. This technique takes advantage of the high affinity of many proteins for specific chemical groups. For example, the plant protein concanavalin A is a carbohydrate-binding protein, or lectin (Section 11.4), that has affinity for glucose. When a crude extract is passed through a column of beads containing covalently attached glucose residues, concanavalin A binds to the beads, whereas most other proteins do not (Figure 3.5). The bound concanavalin A can then be released from the column by adding a concentrated solution of glucose. The glucose in solution displaces the column-attached glucose residues from binding sites on concanavalin A. Affinity chromatography is a powerful means of isolating transcription factors—proteins that regulate gene expression by binding to specific DNA sequences. A protein mixture is passed through a column containing specific DNA sequences attached to a matrix; proteins with a high affinity for the sequence will bind and be retained. In this instance, the transcription factor is released by washing with a solution containing a high concentration of salt.

In general, affinity chromatography can be effectively used to isolate a protein that recognizes group X by (1) covalently attaching X or a derivative of it to a column; (2) adding a mixture of proteins to this column, which is then washed with buffer to remove unbound proteins; and (3) eluting the desired protein by adding a high concentration of a soluble form of X or altering the conditions to decrease binding affinity. Affinity chromatography is most effective when the interaction of the protein and the molecule that is used as the bait is highly specific.

The process of standard affinity chromatography can isolate proteins expressed from cloned genes (Section 5.2). Extra amino acids are encoded

in the cloned gene that, when expressed, serve as an affinity tag that can be readily trapped. For example, repeats of the codon for histidine may be added such that the expressed protein has a string of histidine residues (called a *His tag*) on one end. The tagged proteins are then passed through a column of beads containing covalently attached, immobilized nickel(II) or other metal ions. The His tags bind tightly to the immobilized metal ions, binding the desired protein, while other proteins flow through the column. The protein can then be eluted from the column by the addition of imidazole or some other chemical that binds to the metal ions and displaces the protein.

High-pressure liquid chromatography. A technique called *high-pressure liquid chromatography* (HPLC) is an enhanced version of the column techniques already discussed. The column materials are much more finely divided and, as a consequence, possess more interaction sites and thus greater resolving power. Because the column is made of finer material, pressure must be applied to the column to obtain adequate flow rates. The net result is both high resolution and rapid separation. In a typical HPLC setup, a detector that monitors the absorbance of the eluate at a particular wavelength is placed immediately after the column. In the sample HPLC elution profile shown in Figure 3.6, proteins are detected by setting the detector to 220 nm (the characteristic absorbance wavelength of the peptide bond) In a short span of 10 minutes, a number of sharp peaks representing individual proteins can be readily identified.

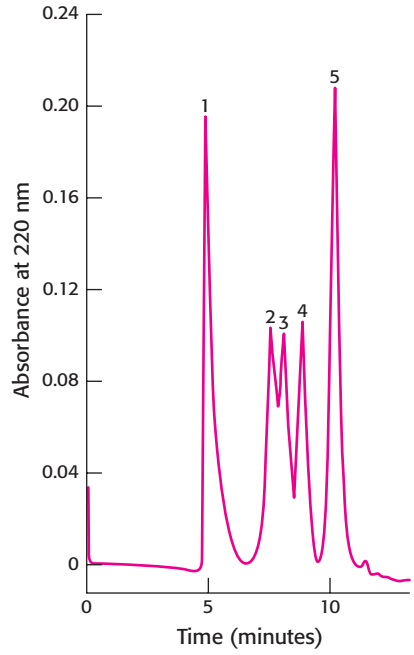


Figure 3.6 High-pressure liquid chromatography (HPLC). Gel filtration by HPLC clearly defines the individual proteins because of its greater resolving power: (1) thyroglobulin (669 kd), (2) catalase (232 kd), (3) bovine serum albumin (67 kd), (4) ovalbumin (43 kd), and (5) ribonuclease (13.4 kd). [After K. J. Wilson and T. D. Schlabach. In *Current Protocols in Molecular Biology,* vol. 2, suppl. 41, F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl, Eds. (Wiley, 1998), p. 10.14.1.]

Proteins can be separated by gel electrophoresis and displayed

How can we tell that a purification scheme is effective? One way is to ascertain that the specific activity rises with each purification step. Another is to determine that the number of different proteins in each sample declines at each step. The technique of electrophoresis makes the latter method possible.

Gel electrophoresis. A molecule with a net charge will move in an electric field. This phenomenon, termed *electrophoresis,* offers a powerful means of separating proteins and other macromolecules, such as DNA and RNA. The velocity of migration (v) of a protein (or any molecule) in an electric field depends on the electric field strength (E), the net charge on the protein (z), and the frictional coefficient (f).

$$v = Ez/f \qquad (1)$$

The electric force Ez driving the charged molecule toward the oppositely charged electrode is opposed by the viscous drag fv arising from friction between the moving molecule and the medium. The frictional coefficient f depends on both the mass and shape of the migrating molecule and the viscosity (η) of the medium. For a sphere of radius r,

$$f = 6\pi\eta r \qquad (2)$$

Electrophoretic separations are nearly always carried out in porous gels (or on solid supports such as paper) because the gel serves as a molecular sieve that enhances separation (Figure 3.7). Molecules that are small compared with the pores in the gel readily move through the gel, whereas molecules much larger than the pores are almost immobile. Intermediate-size molecules move through the gel with various degrees of facility. The electric field is applied such that proteins migrate from the negative to the positive electrodes, typically from top to bottom. Electrophoresis is performed in a

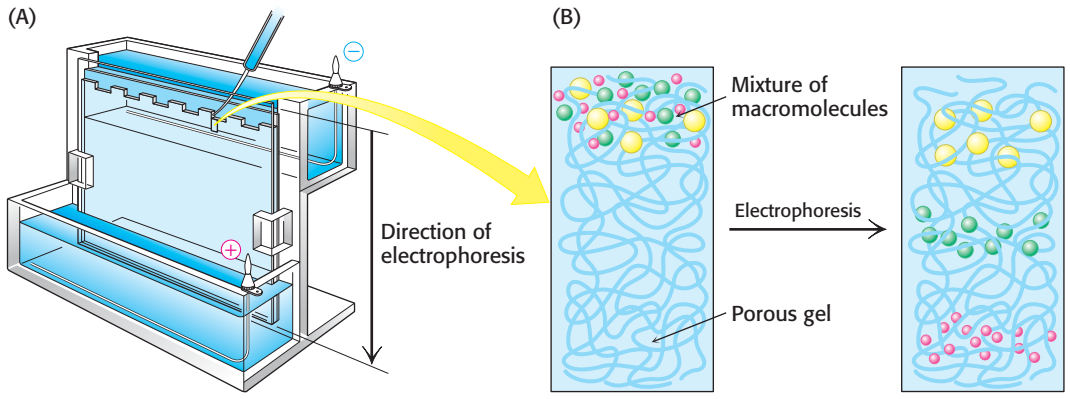


Figure 3.7 Polyacrylamide gel electrophoresis.
(A) Gel-electrophoresis apparatus. Typically, several samples undergo electrophoresis on one flat polyacrylamide gel. A microliter pipette is used to place solutions of proteins in the wells of the slab. A cover is then placed over the gel chamber and voltage is applied. The negatively charged SDS (sodium dodecyl sulfate)–protein complexes migrate in the direction of the anode, at the bottom of the gel.
(B) The sieving action of a porous polyacrylamide gel separates proteins according to size, with the smallest moving most rapidly.

thin, vertical slab of polyacrylamide gel. Polyacrylamide gels are choice supporting media for electrophoresis because they are chemically inert and readily formed by the polymerization of acrylamide with a small amount of the cross-linking agent methylenebisacrylamide to make a three-dimensional mesh (Figure 3.8). Electrophoresis is distinct from gel filtration in that, because of the electric field, all of the molecules, regardless of size, are forced to move through the same matrix.

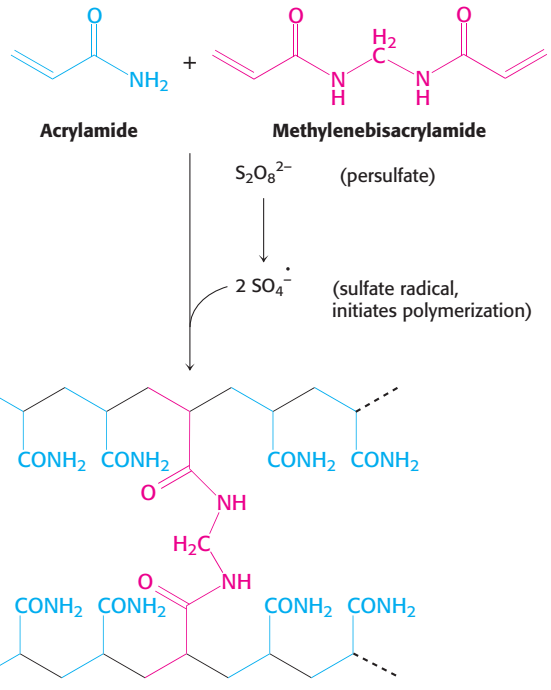


Figure 3.8 Formation of a polyacrylamide gel. A three-dimensional mesh is formed by copolymerizing activated monomer (blue) and cross-linker (red).

Sodium dodecyl sulfate (SDS)

Proteins can be separated largely on the basis of mass by electrophoresis in a polyacrylamide gel under denaturing conditions. The mixture of proteins is first dissolved in a solution of sodium dodecyl sulfate (SDS), an anionic detergent that disrupts nearly all noncovalent interactions in native proteins. β-Mercaptoethanol (2-thioethanol) or dithiothreitol is added to reduce disulfide bonds. Anions of SDS bind to main chains at a ratio of about one SDS anion for every two amino acid residues. The negative charge

acquired on binding SDS is usually much greater than the charge on the native protein; the contribution of the protein to the total charge of the SDS–protein complex is thus rendered insignificant. As a result, this complex of SDS with a denatured protein has a large net negative charge that is roughly proportional to the mass of the protein. The SDS–protein complexes are then subjected to electrophoresis. When the electrophoresis is complete, the proteins in the gel can be visualized by staining them with silver or a dye such as Coomassie blue, which reveals a series of bands (Figure 3.9). Radioactive labels, if they have been incorporated into proteins, can be detected by placing a sheet of x-ray film over the gel, a procedure called *autoradiography*.

Small proteins move rapidly through the gel, whereas large proteins stay at the top, near the point of application of the mixture. The mobility of most polypeptide chains under these conditions is linearly proportional to the logarithm of their mass (Figure 3.10). Some carbohydrate-rich proteins and membrane proteins do not obey this empirical relation, however. SDS–polyacrylamide gel electrophoresis (often referred to as SDS-PAGE) is rapid, sensitive, and capable of a high degree of resolution. As little as 0.1 μg (~2 pmol) of a protein gives a distinct band when stained with Coomassie blue, and even less (~0.02 μg) can be detected with a silver stain. Proteins that differ in mass by about 2% (e.g., 50 and 51 kd, arising from a difference of about 10 amino acids) can usually be distinguished with SDS-PAGE.

We can examine the efficacy of our purification scheme by analyzing a part of each fraction by electrophoresis. The initial fractions will display dozens to hundreds of proteins. As the purification progresses, the number of bands will diminish, and the prominence of one of the bands should increase. This band should correspond to the protein of interest.

Isoelectric focusing. Proteins can also be separated electrophoretically on the basis of their relative contents of acidic and basic residues. The *isoelectric point* (pI) of a protein is the pH at which its net charge is zero. At this pH, its electrophoretic mobility is zero because z in equation 1 is equal to zero. For example, the pI of cytochrome *c*, a highly basic electron-transport protein, is 10.6, whereas that of serum albumin, an acidic protein in blood, is 4.8. Suppose that a mixture of proteins undergoes electrophoresis in a pH gradient in a gel in the absence of SDS. Each protein will move until it reaches a position in the gel at which the pH is equal to the pI of the protein. This method of separating proteins according to their isoelectric point is called *isoelectric focusing*. The pH gradient in the gel is formed first by subjecting a mixture of *polyampholytes* (small multicharged polymers) having many different pI values to electrophoresis. Isoelectric focusing can readily resolve proteins that differ in pI by as little as 0.01, which means that proteins differing by one net charge can be separated (Figure 3.11).

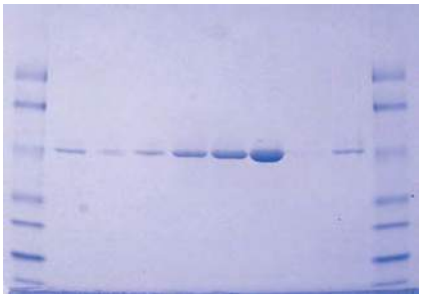


Figure 3.9 Staining of proteins after electrophoresis. Proteins subjected to electrophoresis on an SDS–polyacrylamide gel can be visualized by staining with Coomassie blue. [Courtesy of Kodak Scientific Imaging Systems.]

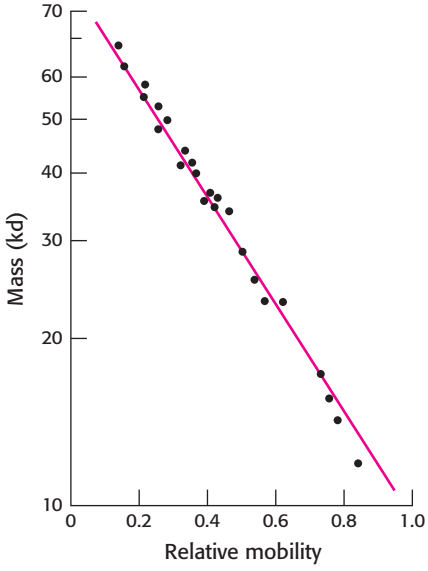


Figure 3.10 Electrophoresis can determine mass. The electrophoretic mobility of many proteins in SDS–polyacrylamide gels is inversely proportional to the logarithm of their mass. [After K. Weber and M. Osborn, *The Proteins,* vol. 1, 3d ed. (Academic Press, 1975), p. 179.]

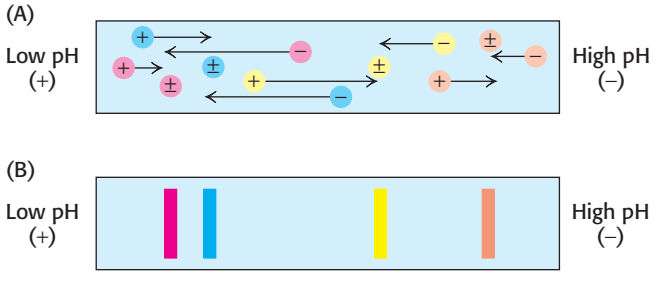


Figure 3.11 The principle of isoelectric focusing. A pH gradient is established in a gel before loading the sample. (A) The sample is loaded and voltage is applied. The proteins will migrate to their isoelectric pH, the location at which they have no net charge. (B) The proteins form bands that can be excised and used for further experimentation.

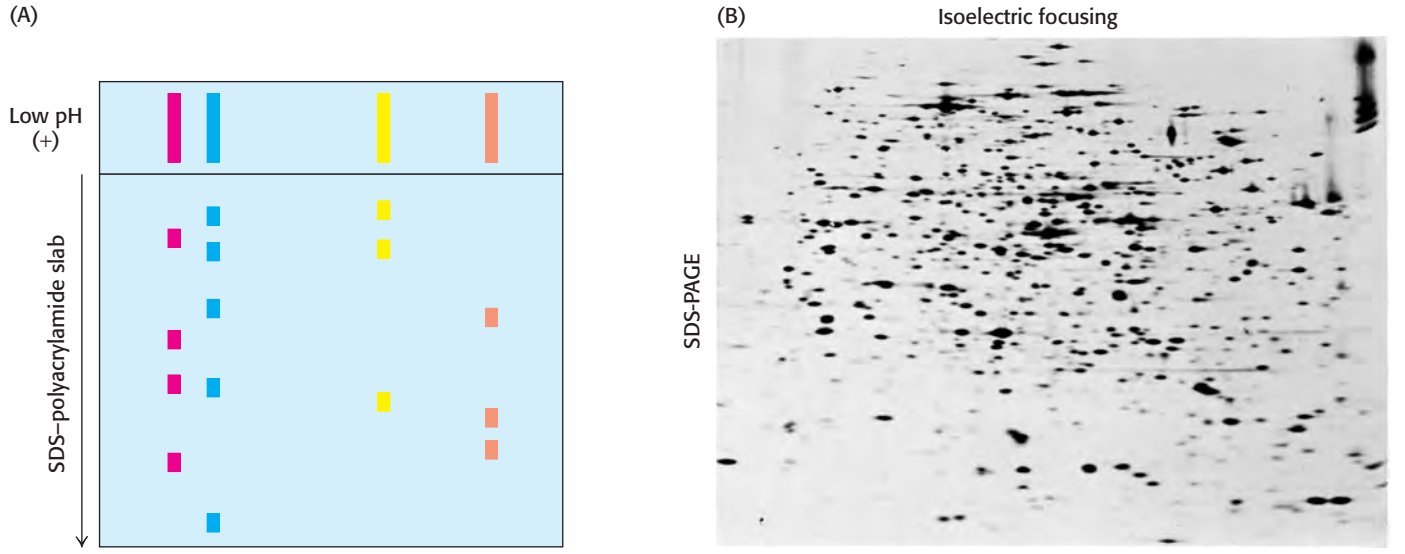


Figure 3.12 Two-dimensional gel electrophoresis. (A) A protein sample is initially fractionated in one dimension by isoelectric focusing as described in Figure 3.11. The isoelectric focusing gel is then attached to an SDS–polyacrylamide gel, and electrophoresis is performed in the second dimension, perpendicular to the original separation. Proteins with the same pI are now separated on the basis of mass. (B) Proteins from *E. coli* were separated by two-dimensional gel electrophoresis, resolving more than a thousand different proteins. The proteins were first separated according to their isoelectric pH in the horizontal direction and then by their apparent mass in the vertical direction. [(B) Courtesy of Dr. Patrick H. O'Farrell.]

Two-dimensional electrophoresis. Isoelectric focusing can be combined with SDS-PAGE to obtain very high resolution separations. A single sample is first subjected to isoelectric focusing. This single-lane gel is then placed horizontally on top of an SDS–polyacrylamide slab. The proteins are thus spread across the top of the polyacrylamide gel according to how far they migrated during isoelectric focusing. They then undergo electrophoresis again in a perpendicular direction (vertically) to yield a two-dimensional pattern of spots. In such a gel, proteins have been separated in the horizontal direction on the basis of isoelectric point and in the vertical direction on the basis of mass. Remarkably, more than a thousand different proteins in the bacterium *Escherichia coli* can be resolved in a single experiment by two-dimensional electrophoresis (Figure 3.12).

Proteins isolated from cells under different physiological conditions can be subjected to two-dimensional electrophoresis. The intensities of individual spots on the gels can then be compared, which indicates that the concentrations of specific proteins have changed in response to the physiological state (Figure 3.13). How can we discover the identity of a protein that is showing such responses? Although many proteins are displayed on a two-dimensional gel, they are not identified. It is now possible to identify proteins by coupling two-dimensional gel electrophoresis with mass spectrometric techniques. We will examine these powerful techniques shortly (Section 3.4).

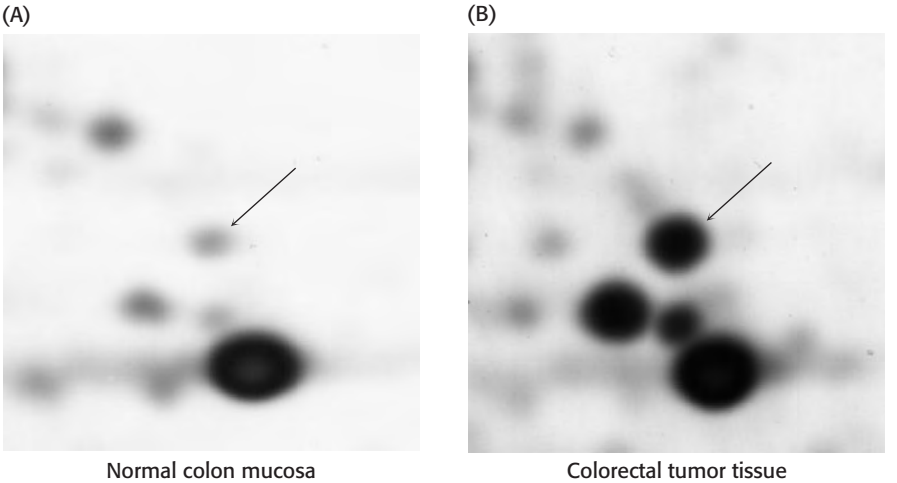


Figure 3.13 Alterations in protein levels detected by two-dimensional gel electrophoresis. Samples of normal colon mucosa and colorectal tumor tissue from the same person were analyzed by two-dimensional gel electrophoresis. In the gel section shown, changes in the intensity of several spots are evident, including a dramatic increase in levels of the protein indicated by the arrow, corresponding to the enzyme glyceraldehyde-3-phosphate dehydrogenase. [Courtesy of Lin Quinsong © 2010, The American Society for Biochemistry and Molecular Biology.]

Table 3.1 Quantification of a purification protocol for a fictitious protein

Step	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Yield (%)	Purification level
Homogenization	15,000	150,000	10	100	1
Salt fractionation	4,600	138,000	30	92	3
Ion-exchange chromatography	1,278	115,500	90	77	9
Gel-filtration chromatography	68.8	75,000	1,100	50	110
Affinity chromatography	1.75	52,500	30,000	35	3,000

A protein purification scheme can be quantitatively evaluated

To determine the success of a protein purification scheme, we monitor each step of the procedure by determining the specific activity of the protein mixture and by subjecting it to SDS-PAGE analysis. Consider the results for the purification of a fictitious protein, summarized in Table 3.1 and Figure 3.14. At each step, the following parameters are measured:

Total Protein. The quantity of protein present in a fraction is obtained by determining the protein concentration of a part of each fraction and multiplying by the fraction's total volume.

Total Activity. The enzyme activity for the fraction is obtained by measuring the enzyme activity in the volume of fraction used in the assay and multiplying by the fraction's total volume.

Specific Activity. This parameter is obtained by dividing total activity by total protein.

Yield. This parameter is a measure of the activity retained after each purification step as a percentage of the activity in the crude extract. The amount of activity in the initial extract is taken to be 100%.

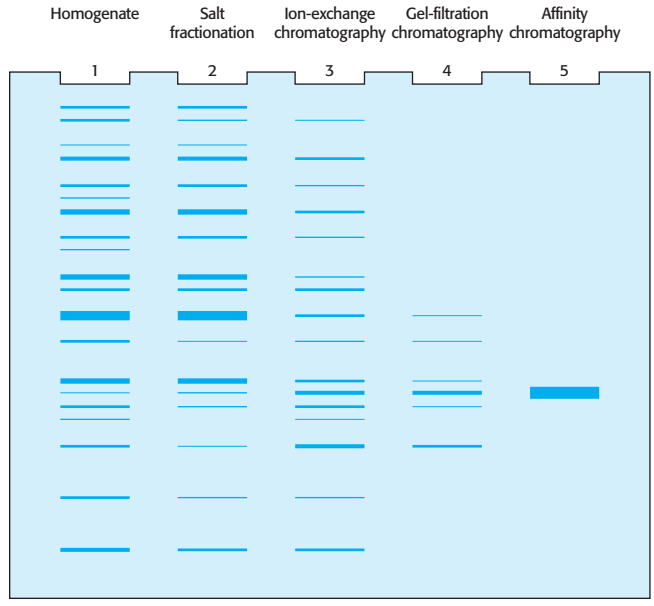


Figure 3.14 Electrophoretic analysis of a protein purification. The purification scheme in Table 3.1 was analyzed by SDS-PAGE. Each lane contained 50 μg of sample. The effectiveness of the purification can be seen as the band for the protein of interest becomes more prominent relative to other bands.

Purification Level. This parameter is a measure of the increase in purity and is obtained by dividing the specific activity, calculated after each purification step, by the specific activity of the initial extract.

As we see in Table 3.1, the first purification step, salt fractionation, leads to an increase in purity of only 3-fold, but we recover nearly all the target protein in the original extract, given that the yield is 92%. After dialysis to lower the high concentration of salt remaining from the salt fractionation, the fraction is passed through an ion-exchange column. The purification now increases to 9-fold compared with the original extract, whereas the yield falls to 77%. Gel-filtration chromatography brings the level of purification to 110-fold, but the yield is now at 50%. The final step is affinity chromatography with the use of a ligand specific for the target enzyme. This step, the most powerful of these purification procedures, results in a purification level of 3000-fold but lowers the yield to 35%. The SDS-PAGE analysis in Figure 3.14 shows that, if we load a constant amount of protein onto each lane after each step, the number of bands decreases in proportion

to the level of purification, and the amount of protein of interest increases as a proportion of the total protein present.

A good purification scheme takes into account both purification levels and yield. A high degree of purification and a poor yield leave little protein with which to experiment. A high yield with low purification leaves many contaminants (proteins other than the one of interest) in the fraction and complicates the interpretation of subsequent experiments.

Ultracentrifugation is valuable for separating biomolecules and determining their masses

We have already seen that centrifugation is a powerful and generally applicable method for separating a crude mixture of cell components. This technique is also valuable for the analysis of the physical properties of biomolecules. Using centrifugation, we can determine such parameters as mass and density, learn something about the shape of a molecule, and investigate the interactions between molecules. To deduce these properties from the centrifugation data, we require a mathematical description of how a particle behaves when a centrifugal force is applied.

A particle will move through a liquid medium when subjected to a centrifugal force. A convenient means of quantifying the rate of movement is to calculate the sedimentation coefficient, s, of a particle by using the following equation:

$$s = m(1 - \bar{\nu}\rho)/f$$

where m is the mass of the particle, $\bar{\nu}$ is the partial specific volume (the reciprocal of the particle density), ρ is the density of the medium, and f is the frictional coefficient (a measure of the shape of the particle). The $(1 - \bar{\nu}\rho)$ term is the buoyant force exerted by liquid medium.

Sedimentation coefficients are usually expressed in *Svedberg units* (S), equal to 10^{-13} s. The smaller the S value, the more slowly a molecule moves in a centrifugal field. The S values for a number of biomolecules and cellular components are listed in Table 3.2 and Figure 3.15.

Several important conclusions can be drawn from the preceding equation:

1. The sedimentation velocity of a particle depends in part on its mass. A more massive particle sediments more rapidly than does a less massive particle of the same shape and density.

2. Shape, too, influences the sedimentation velocity because it affects the viscous drag. The frictional coefficient f of a compact particle is smaller than that of an extended particle of the same mass. Hence, elongated particles sediment more slowly than do spherical ones of the same mass.

Table 3.2 S values and molecular weights of sample proteins

Protein	S value (Svedberg units)	Molecular weight
Pancreatic trypsin inhibitor	1	6,520
Cytochrome *c*	1.83	12,310
Ribonuclease A	1.78	13,690
Myoglobin	1.97	17,800
Trypsin	2.5	23,200
Carbonic anhydrase	3.23	28,800
Concanavalin A	3.8	51,260
Malate dehydrogenase	5.76	74,900
Lactate dehydrogenase	7.54	146,200

Source: T. Creighton, *Proteins,* 2d ed. (W. H. Freeman and Company, 1993), Table 7.1.

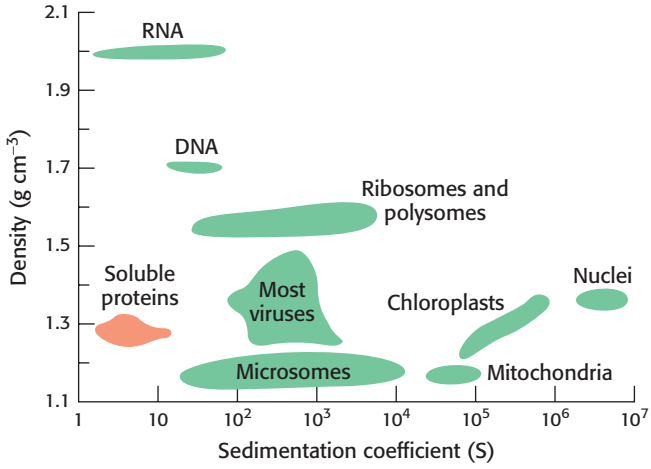


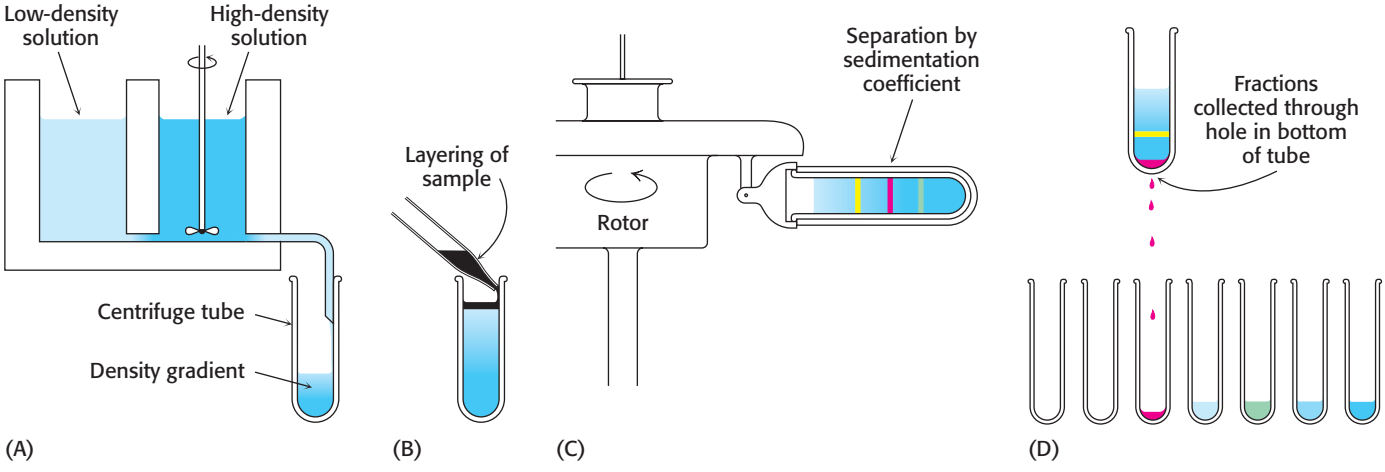
Figure 3.15 Density and sedimentation coefficients of cellular components. [After L. J. Kleinsmith and V. M. Kish, *Principles of Cell and Molecular Biology,* 2d ed. (HarperCollins, 1995), p. 138.]

3. A dense particle moves more rapidly than does a less dense one because the opposing buoyant force $(1 - \bar{\nu}\rho)$ is smaller for the denser particle.

4. The sedimentation velocity also depends on the density of the solution (ρ). Particles sink when $\bar{\nu}\rho < 1$, float when $\bar{\nu}\rho > 1$, and do not move when $\bar{\nu}\rho = 1$,

A technique called *zonal, band,* or most commonly *gradient* centrifugation can be used to separate proteins with different sedimentation coefficients. The first step is to form a density gradient in a centrifuge tube. Differing proportions of a low-density solution (such as 5% sucrose) and a high-density solution (such as 20% sucrose) are mixed to create a linear gradient of sucrose concentration ranging from 20% at the bottom of the tube to 5% at the top (Figure 3.16). The role of the gradient is to prevent convective flow. A small volume of a solution containing the mixture of proteins to be separated is placed on top of the density gradient. When the rotor is spun, proteins move through the gradient and separate according to their sedimentation coefficients. The time and speed of the centrifugation is determined empirically. The separated bands, or zones, of protein can be harvested by making a hole in the bottom of the tube and collecting drops.

Figure 3.16 Zonal centrifugation. The steps are as follows: (A) form a density gradient, (B) layer the sample on top of the gradient, (C) place the tube in a swinging-bucket rotor and centrifuge it, and (D) collect the samples. [After D. Freifelder, *Physical Biochemistry,* 2d ed. (W. H. Freeman and Company, 1982), p. 397.]



The drops can be measured for protein content and catalytic activity or another functional property. This sedimentation-velocity technique readily separates proteins differing in sedimentation coefficient by a factor of two or more.

The mass of a protein can be directly determined by *sedimentation equilibrium,* in which a sample is centrifuged at low speed such that a concentration gradient of the sample is formed. However, this sedimentation is counterbalanced by the diffusion of the sample from regions of high to low concentration. When equilibrium has been achieved, the shape of the final gradient depends solely on the mass of the sample. *The sedimentation-equilibrium technique for determining mass is very accurate and can be applied without denaturing the protein. Thus the native quaternary structure of multimeric proteins is preserved.* In contrast, SDS–polyacrylamide gel electrophoresis provides an *estimate* of the mass of dissociated polypeptide chains under *denaturing* conditions. Note that, if we know the mass of the dissociated components of a multimeric protein as determined by SDS–polyacrylamide analysis and the mass of the intact multimer as determined by sedimentation-equilibrium analysis, we can determine the number of copies of each polypeptide chain present in the protein complex.

Protein purification can be made easier with the use of recombinant DNA technology

In Chapter 5, we shall consider the widespread effect of recombinant DNA technology on all areas of biochemistry and molecular biology. The application of recombinant methods to the overproduction of proteins has enabled dramatic advances in our understanding of their structure and function. Before the advent of this technology, proteins were isolated solely from their native sources, often requiring a large amount of tissue to obtain a sufficient amount of protein for analytical study. For example, the purification of bovine deoxyribonuclease in 1946 required nearly ten pounds of beef pancreas to yield one gram of protein. As a result, biochemical studies on purified material were often limited to abundant proteins.

Armed with the tools of recombinant technology, the biochemist is now able to enjoy a number of significant advantages:

1. *Proteins can be expressed in large quantities.* The homogenate serves as the starting point in a protein purification scheme. For recombinant systems, a host organism that is amenable to genetic manipulation, such as the bacterium *Escherichia coli* or the yeast *Pichia pastoris*, is utilized to express a protein of interest. The biochemist can exploit the short doubling times and ease of genetic manipulation of such organisms to produce large amounts of protein from manageable amounts of culture. As a result, purification can begin with a homogenate that is often highly enriched with the desired molecule. Moreover, a protein can be easily obtained regardless of its natural abundance or its species of origin.

2. *Affinity tags can be fused to proteins.* As described earlier, affinity chromatography can be a highly selective step within a protein purification scheme. Recombinant DNA technology enables the attachment of any one of a number of possible affinity tags to a protein (such as the "His tag" mentioned earlier). Hence, the benefits of affinity chromatography can be realized even for those proteins for which a binding partner is unknown or not easily determined.

3. *Proteins with modified primary structures can be readily generated.* A powerful aspect of recombinant DNA technology as applied to protein

purification is the ability to manipulate genes to generate variants of a native protein sequence (Section 5.2). We learned in Section 2.4 that many proteins consist of compact domains connected by flexible linker regions. With the use of genetic-manipulation strategies, fragments of a protein that encompass single domains can be generated, an advantageous approach when expression of the entire protein is limited by its size or solubility. Additionally, as we will see in Section 9.1, amino acid substitutions can be introduced into the active site of an enzyme to precisely probe the roles of specific residues within its catalytic cycle.

3.2 Amino Acid Sequences of Proteins Can Be Determined Experimentally

The amino acid sequence of a protein can be a valuable source of insight into its function, structure, and history.

1. *The sequence of a protein of interest can be compared with all other known sequences to ascertain whether significant similarities exist.* A search for kinship between a newly sequenced protein and the millions of previously sequenced ones takes only a few seconds on a personal computer (Chapter 6). If the newly isolated protein is a member of an established class of protein, we can begin to infer information about the protein's structure and function. For instance, chymotrypsin and trypsin are members of the serine protease family, a clan of proteolytic enzymes that have a common catalytic mechanism based on a reactive serine residue (Chapter 9). If the sequence of the newly isolated protein shows sequence similarity with trypsin or chymotrypsin, the result suggests that it may be a serine protease.

2. *Comparison of sequences of the same protein in different species yields a wealth of information about evolutionary pathways.* Genealogical relationships between species can be inferred from sequence differences between their proteins. If we assume that the random mutation rate of proteins over time is constant, then careful sequence comparison of related proteins between two organisms can provide an estimate for when these two evolutionary lines diverged. For example, a comparison of serum albumins found in primates indicates that human beings and African apes diverged 5 million years ago, not 30 million years ago as was once thought. Sequence analyses have opened a new perspective on the fossil record and the pathway of human evolution.

3. *Amino acid sequences can be searched for the presence of internal repeats.* Such internal repeats can reveal the history of an individual protein itself. Many proteins apparently have arisen by duplication of primordial genes followed by their diversification. For example, calmodulin, a ubiquitous calcium sensor in eukaryotes, contains four similar calcium-binding modules that arose by gene duplication (Figure 3.17).

4. *Many proteins contain amino acid sequences that serve as signals designating their destinations or controlling their processing.* For example, a protein destined for export from a cell or for location in a membrane contains a *signal sequence,* a stretch of about 20 hydrophobic residues near the amino terminus that directs the protein to the appropriate membrane. Another protein may contain a stretch of amino acids that functions as a *nuclear localization signal,* directing the protein to the nucleus.

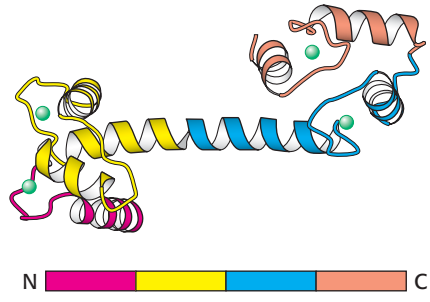


Figure 3.17 Repeating motifs in a protein chain. Calmodulin, a calcium sensor, contains four similar units (shown in red, yellow, blue, and orange) in a single polypeptide chain. *Notice* that each unit binds a calcium ion (shown in green). [Drawn from 1CLL.pdb.]

5. *Sequence data provide a basis for preparing antibodies specific for a protein of interest*. One or more parts of the amino acid sequence of a protein will elicit an antibody when injected into a mouse or rabbit. These specific antibodies can be very useful in determining the amount of a protein present in solution or in the blood, ascertaining its distribution within a cell, or cloning its gene (Section 3.3).

6. *Amino acid sequences are valuable for making DNA probes that are specific for the genes encoding the corresponding proteins*. Knowledge of a protein's primary structure permits the use of reverse genetics. DNA sequences that correspond to a part of the amino acid sequence can be constructed on the basis of the genetic code. These DNA sequences can be used as probes to isolate the gene encoding the protein so that the entire sequence of the protein can be determined. The gene in turn can provide valuable information about the physiological regulation of the protein. Protein sequencing is an integral part of molecular genetics, just as DNA cloning is central to the analysis of protein structure and function. We will revisit some of these topics in more detail in Chapter 5.

Peptide sequences can be determined by automated Edman degradation

Given the importance of determining the amino acid sequence of a protein, let us consider one of the methods available to the biochemist for determining this information. Consider a simple peptide, whose composition is unknown to the researcher:

Ala-Gly-Asp-Phe-Arg-Gly

The first step is to determine the *amino acid composition* of the peptide. The peptide is hydrolyzed into its constituent amino acids by heating it in 6 M HCl at 110°C for 24 hours. The amino acids in solution can then be separated by ion-exchange chromatography. The identity of each amino acid is revealed by its elution volume, which is the volume of buffer used to remove the amino acid from the column (Figure 3.18), and its quantity is revealed

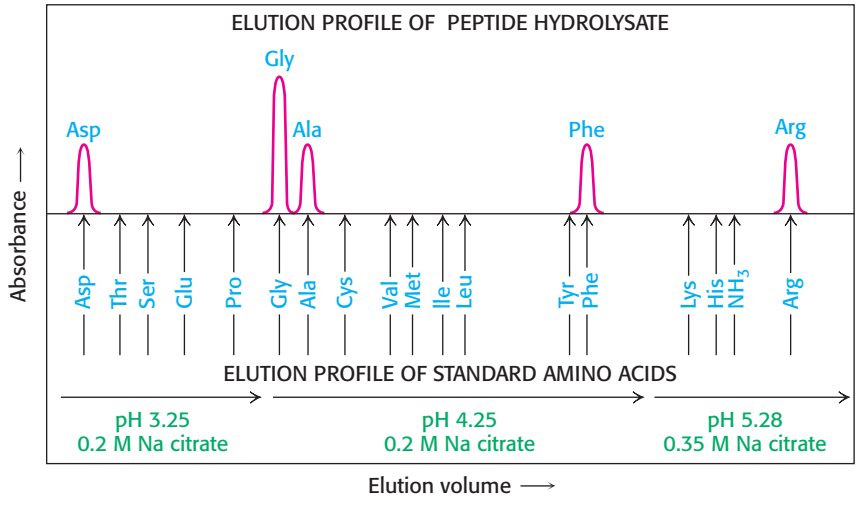
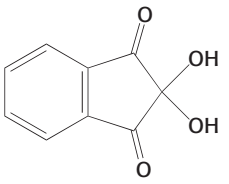


Figure 3.18 Determination of amino acid composition. Different amino acids in a peptide hydrolysate can be separated by ion-exchange chromatography on a sulfonated polystyrene resin (such as Dowex-50). Buffers (in this case, sodium citrate) of increasing pH are used to elute the amino acids from the column. The amount of each amino acid present is determined from the absorbance. Aspartate, which has an acidic side chain, is first to emerge, whereas arginine, which has a basic side chain, is the last. The original peptide is revealed to be composed of one aspartate, one alanine, one phenylalanine, one arginine, and two glycine residues.

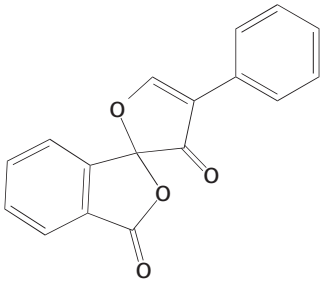
by reaction with an indicator dye such as *ninhydrin* or *fluorescamine*. After conjugation to the indicator, the amino acid exhibits a color with an intensity that is proportional to its concentration. A comparison of the chromatographic patterns of our sample hydrolysate with that of a standard mixture of amino acids would show that the amino acid composition of the peptide is



Ninhydrin

$$(\text{Ala, Arg, Asp, Gly}_2\text{, Phe})$$

The parentheses denote that this is the amino acid composition of the peptide, not its sequence.



Fluorescamine

The next step is to identify the N-terminal amino acid. Pehr Edman devised a method for labeling the amino-terminal residue and cleaving it from the peptide without disrupting the peptide bonds between the other amino acid residues. The *Edman degradation* sequentially removes one residue at a time from the amino end of a peptide (Figure 3.19). *Phenyl isothiocyanate* reacts with the uncharged terminal amino group of the peptide to form a phenylthiocarbamoyl derivative. Then, under mildly acidic conditions, a cyclic derivative of the terminal amino acid is liberated, which leaves an intact peptide shortened by one amino acid. The cyclic compound is a phenylthiohydantoin (PTH)–amino acid, which can be identified by chromatographic methods. The Edman procedure can then be repeated on the shortened peptide, yielding another PTH–amino acid, which can again be identified by chromatography. Three more rounds of the Edman degradation will reveal the complete sequence of the original hexapeptide.

The development of automated sequencers has markedly decreased the time required to determine protein sequences. By repeated Edman degradations, the amino acid sequence of some 50 residues in a protein can be

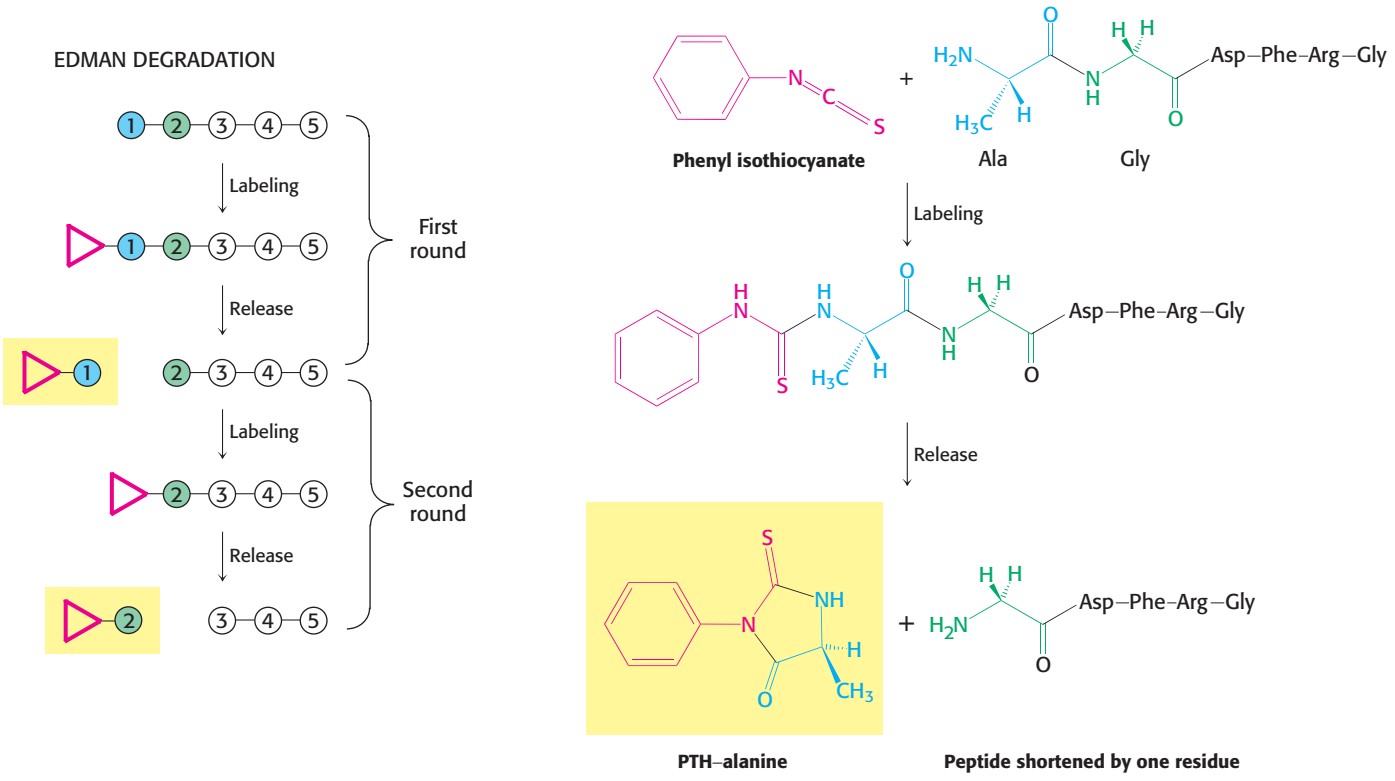


Figure 3.19 The Edman degradation. The labeled amino-terminal residue (PTH–alanine in the first round) can be released without hydrolyzing the rest of the peptide. Hence, the amino-terminal residue of the shortened peptide (Gly-Asp-Phe-Arg-Gly) can be determined in the second round. Three more rounds of the Edman degradation reveal the complete sequence of the original peptide.

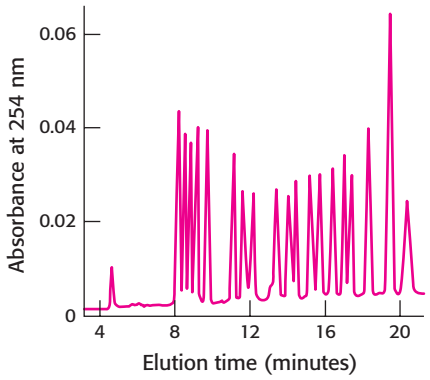


Figure 3.20 Separation of PTH–amino acids. PTH–amino acids can be rapidly separated by high-pressure liquid chromatography (HPLC). In this HPLC profile, a mixture of PTH–amino acids is clearly resolved into its components. An unknown amino acid can be identified by its elution position relative to the known ones.

determined. Gas-phase sequenators can analyze picomole quantities of peptides and proteins with the use of high-pressure liquid chromatography to identify each amino acid as it is released (Figure 3.20). This high sensitivity makes it feasible to analyze the sequence of a protein sample eluted from a single band of an SDS–polyacrylamide gel.

Proteins can be specifically cleaved into small peptides to facilitate analysis

In principle, it should be possible to sequence an entire protein by using the Edman method. In practice, the peptides cannot be much longer than about 50 residues, because not all peptides in the reaction mixture release the amino acid derivative at each step. For instance, if the efficiency of release for each round were 98%, the proportion of "correct" amino acid released after 60 rounds would be (0.98^{60}), or 0.3—a hopelessly impure mix. This obstacle can be circumvented by cleaving a protein into smaller peptides that can be sequenced. Protein cleavage can be achieved by chemical reagents, such as cyanogen bromide, or proteolytic enzymes, such as trypsin. Table 3.3 gives several other ways of specifically cleaving polypeptide chains. Note that these methods are sequence specific: they disrupt the protein backbone at particular amino acid residues in a predictable manner.

Table 3.3 Specific cleavage of polypeptides

Reagent	Cleavage site
Chemical cleavage	
Cyanogen bromide	Carboxyl side of methionine residues
O-Iodosobenzoate	Carboxyl side of tryptophan residues
Hydroxylamine	Asparagine–glycine bonds
2-Nitro-5-thiocyanobenzoate	Amino side of cysteine residues
Enzymatic cleavage	
Trypsin	Carboxyl side of lysine and arginine residues
Clostripain	Carboxyl side of arginine residues
Staphylococcal protease	Carboxyl side of aspartate and glutamate residues (glutamate only under certain conditions)
Thrombin	Carboxyl side of arginine
Chymotrypsin	Carboxyl side of tyrosine, tryptophan, phenylalanine, leucine, and methionine
Carboxypeptidase A	Amino side of C-terminal amino acid (not arginine, lysine, or proline)

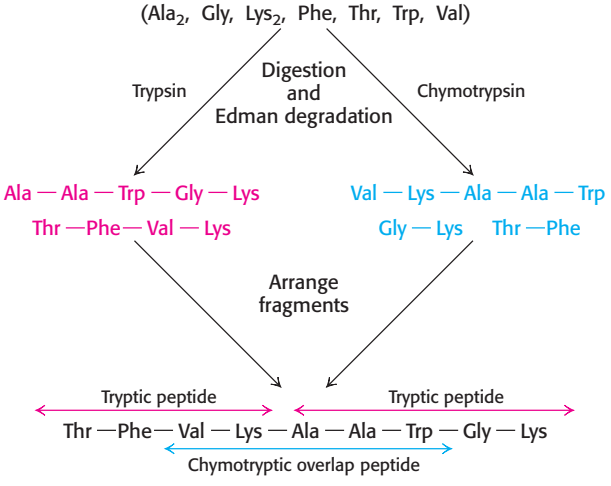


Figure 3.21 Overlap peptides. The peptide obtained by chymotryptic digestion overlaps two tryptic peptides, establishing their order.

The peptides obtained by specific chemical or enzymatic cleavage are separated by some type of chromatography. The sequence of each purified peptide is then determined by the Edman method. At this point, the amino acid sequences of segments of the protein are known, but the order of these segments is not yet defined. How can we order the peptides to obtain the primary structure of the original protein? The necessary additional information is obtained from *overlap peptides* (Figure 3.21). A second enzyme is used to split the polypeptide chain at different linkages. For example, chymotrypsin cleaves preferentially on the carboxyl side of aromatic and some other bulky nonpolar residues (Chapter 9). Because these chymotryptic peptides overlap two or more tryptic peptides, they can be used to establish the order of the peptides. The entire amino acid sequence of the polypeptide chain is then known.

Additional steps are necessary if the initial protein sample is actually several polypeptide chains. SDS–gel electrophoresis under reducing conditions should display the number of chains. Alternatively, the number of distinct N-terminal amino acids could be determined. After a protein has been identified as being made up of two or more polypeptide chains, denaturing agents, such as urea or guanidine hydrochloride, are used to dissociate chains held together by noncovalent bonds. The dissociated chains must be separated from one another before sequence determination can begin. Polypeptide chains linked by disulfide bonds are separated by reduction with thiols such as β-mercaptoethanol or dithiothreitol. To prevent the cysteine residues from recombining, they are then alkylated with iodoacetate to form stable *S*-carboxymethyl derivatives (Figure 3.22). Sequencing can then be performed as already described.

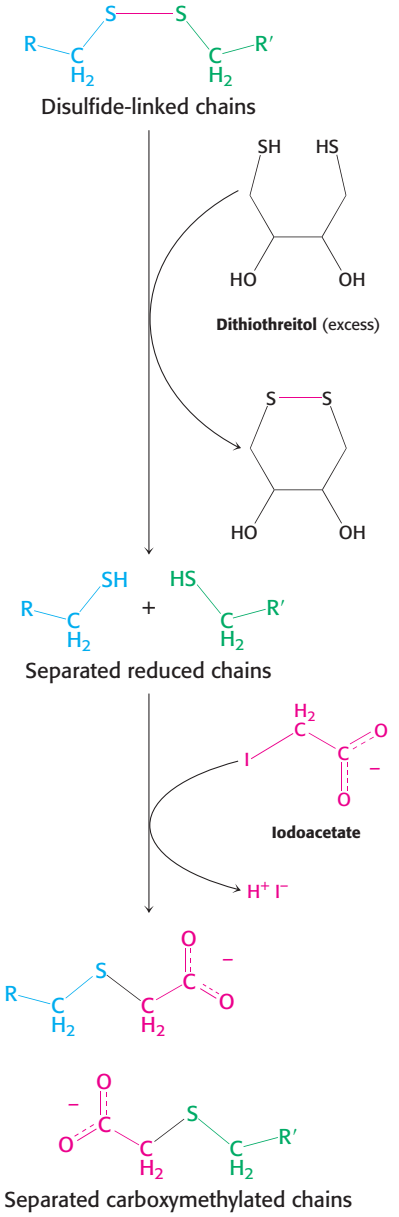


Figure 3.22 Disulfide-bond reduction. Polypeptides linked by disulfide bonds can be separated by reduction with dithiothreitol followed by alkylation to prevent them from re-forming.

Genomic and proteomic methods are complementary

Thousands of proteins have been sequenced by the Edman degradation of peptides derived from specific cleavages. Nevertheless, heroic effort is required to elucidate the sequence of large proteins, those with more than 1000 residues. For sequencing such proteins, a complementary experimental approach based on recombinant DNA technology is often more efficient. As will be discussed in Chapter 5, long stretches of DNA can be cloned and sequenced, and the nucleotide sequence can be translated to reveal the amino acid sequence of the protein encoded by the gene (Figure 3.23). Recombinant DNA technology is producing a wealth of amino acid sequence information at a remarkable rate.

Figure 3.23 DNA sequence yields the amino acid sequence. The complete nucleotide sequence of HIV-1 (human immunodeficiency virus), the cause of AIDS (acquired immune deficiency syndrome), was determined within a year after the isolation of the virus. A part of the DNA sequence specified by the RNA genome of the virus is shown here with the corresponding amino acid sequence (deduced from a knowledge of the genetic code).

Even with the use of the DNA base sequence to determine primary structure, there is still a need to work with isolated proteins. The amino acid sequence deduced by reading the DNA sequence is that of the nascent protein, the direct product of the translational machinery. However, many proteins undergo *posttranslational modifications* after their syntheses. Some have their ends trimmed, and others arise by cleavage of a larger initial polypeptide chain. Cysteine residues in some proteins are oxidized to form disulfide links, connecting either parts within a chain or separate polypeptide chains. Specific side chains of some proteins are altered. Amino acid sequences derived from DNA sequences are rich in information, but they do not disclose these modifications. Chemical analyses of proteins in their mature form are needed to delineate the nature of these changes, which are critical for the biological activities of most proteins. *Thus, genomic and proteomic analyses are complementary approaches to elucidating the structural basis of protein function.*

3.3 Immunology Provides Important Techniques with Which to Investigate Proteins

The purification of a protein enables the biochemist to explore its function and structure within a precisely controlled environment. However, the isolation of a protein removes it from its native context within the cell, where its activity is most physiologically relevant. Advances in the field of immunology (Chapter 34) have enabled the use of antibodies as critical reagents for exploring the functions of proteins within the cell. The exquisite specificity of antibodies for their target proteins provides a means to tag a specific protein so that it can be isolated, quantified, or visualized.

Antibodies to specific proteins can be generated

Immunological techniques begin with the generation of antibodies to a particular protein. An *antibody* (also called an *immunoglobulin,* Ig) is itself a protein (Figure 3.24); it is synthesized by an animal in response to the presence

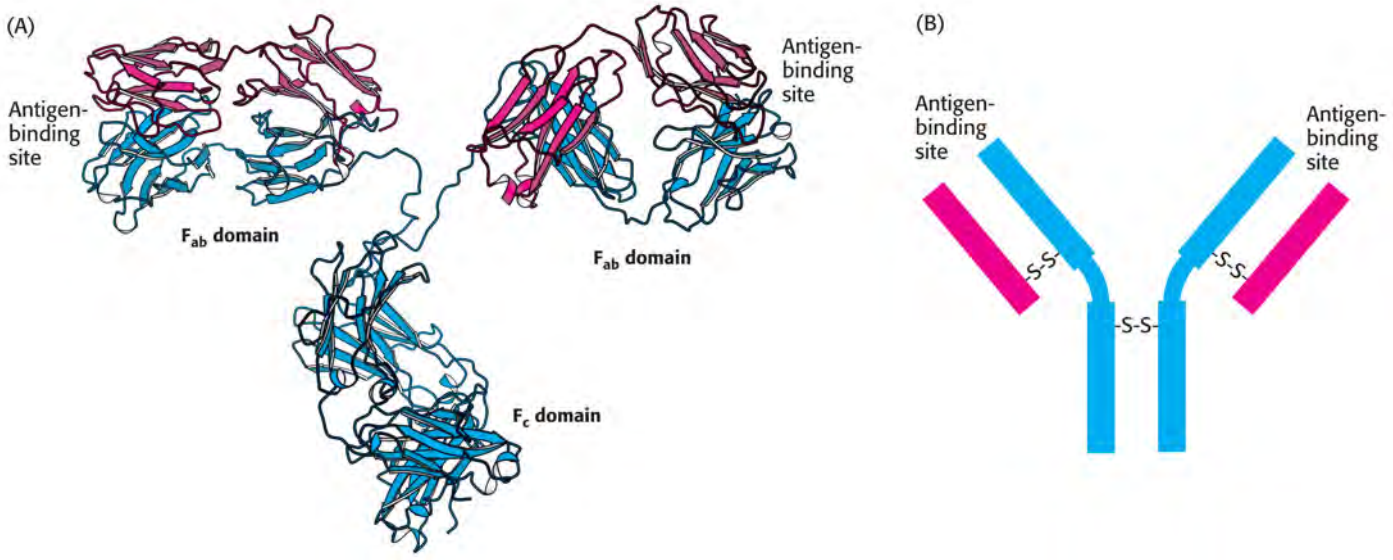


Figure 3.24 Antibody structure. (A) Immunoglobulin G (IgG) consists of four chains, two heavy chains (blue) and two light chains (red), linked by disulfide bonds. The heavy and light chains come together to form F_{ab} domains, which have the antigen-binding sites at the ends. The two heavy chains form the F_c domain. *Notice* that the F_{ab} domains are linked to the F_c domain by flexible linkers. (B) A more schematic representation of an IgG molecule. [Drawn from 1IGT.pdb.]

of a foreign substance, called an *antigen*. Antibodies have specific and high affinity for the antigens that elicited their synthesis. The binding of antibody and antigen is a step in the immune response that protects the animal from infection (Chapter 34). Foreign proteins, polysaccharides, and nucleic acids can be antigens. Small foreign molecules, such as synthetic peptides, also can elicit antibodies, provided that the small molecule is attached to a macromolecular carrier. An antibody recognizes a specific group or cluster of amino acids on the target molecule called an *antigenic determinant* or *epitope*. The specificity of the antibody–antigen interaction is a consequence of the shape complementarity between the two surfaces (Figure 3.25). Animals have a very large repertoire of antibody-producing cells, each producing an

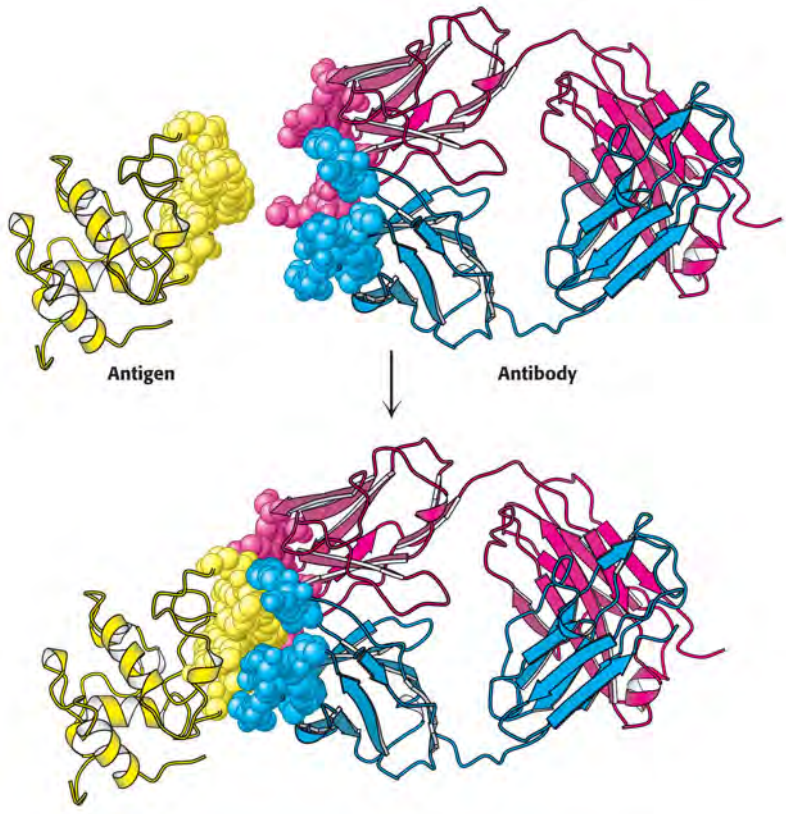


Figure 3.25 Antigen–antibody interactions. A protein antigen, in this case lysozyme, binds to the end of an F_{ab} domain of an antibody. *Notice* that the end of the antibody and the antigen have complementary shapes, allowing a large amount of surface to be buried on binding. [Drawn from 3HFL.pdb.]

antibody that contains a unique surface for antigen recognition. When an antigen is introduced into an animal, it is recognized by a select few cells from this population, stimulating the proliferation of these cells. This process ensures that more antibodies of the appropriate specificity are produced.

Immunological techniques depend on the ability to generate antibodies to a specific antigen. To obtain antibodies that recognize a particular protein, a biochemist injects the protein into a rabbit twice, 3 weeks apart. The injected protein acts as an antigen, stimulating the reproduction of cells producing antibodies that recognize it. Blood is drawn from the immunized rabbit several weeks later and centrifuged to separate blood cells from the supernatant, or serum. The serum, called an *antiserum,* contains antibodies to all antigens to which the rabbit has been exposed. Only some of them will be antibodies to the injected protein. Moreover, antibodies that recognize a particular antigen are not a single molecular species. For instance, 2,4-dinitrophenol (DNP) was used as an antigen to generate antibodies. Analyses of anti-DNP antibodies revealed a wide range of binding affinities; the dissociation constants ranged from about 0.1 nM to 1 μM. Correspondingly, a large number of bands were evident when anti-DNP antibody was subjected to isoelectric focusing. These results indicate that cells are producing many different antibodies, each recognizing a different surface feature of the same antigen. These antibodies are termed *polyclonal,* referring to the fact that they are derived from multiple antibody-producing cell populations (Figure 3.26). The heterogeneity of polyclonal antibodies can be advantageous for certain applications, such as the detection of a protein of low abundance, because each protein molecule can be bound by more than one antibody at multiple distinct antigenic sites.

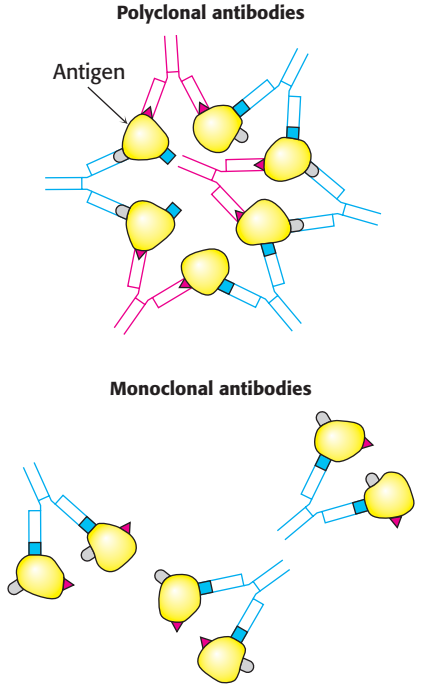


Figure 3.26 Polyclonal and monoclonal antibodies. Most antigens have several epitopes. Polyclonal antibodies are heterogeneous mixtures of antibodies, each specific for one of the various epitopes on an antigen. Monoclonal antibodies are all identical, produced by clones of a single antibody-producing cell. They recognize one specific epitope. [After R. A. Goldsby, T. J. Kindt, and B. A. Osborne, *Kuby Immunology,* 4th ed. (W. H. Freeman and Company, 2000), p. 154.]

Monoclonal antibodies with virtually any desired specificity can be readily prepared

The discovery of a means of producing *monoclonal antibodies* of virtually any desired specificity was a major breakthrough that intensified the power of immunological approaches. As with impure proteins, working with an

impure mixture of antibodies makes it difficult to interpret data. The ideal would be to isolate a clone of cells producing a single, identical antibody. The problem is that antibody-producing cells isolated from an organism have short life spans.

Immortal cell lines that produce monoclonal antibodies do exist. These cell lines are derived from a type of cancer, *multiple myeloma,* which is a malignant disorder of antibody-producing cells. In this cancer, a single transformed plasma cell divides uncontrollably, generating a very large number of *cells of a single kind*. Such a group of cells is a *clone* because the cells are descended from the same cell and have identical properties. The identical cells of the myeloma secrete large amounts of *immunoglobulin of a single kind* generation after generation. These antibodies were useful for elucidating antibody structure, but nothing is known about their specificity and so they are useless for the immunological methods described in the next pages.

César Milstein and Georges Köhler discovered that large amounts of antibodies of nearly any desired specificity can be obtained by fusing a short-lived antibody-producing cell with an immortal myeloma cell. An antigen is injected into a mouse, and its spleen is removed several weeks later (Figure 3.27). A mixture of plasma cells from this spleen is fused in vitro with myeloma cells. Each of the resulting hybrid cells, called hybridoma cells, indefinitely produces the identical antibody specified by the parent cell from the spleen. Hybridoma cells can then be screened by a specific assay for the antigen–antibody interaction to determine which ones

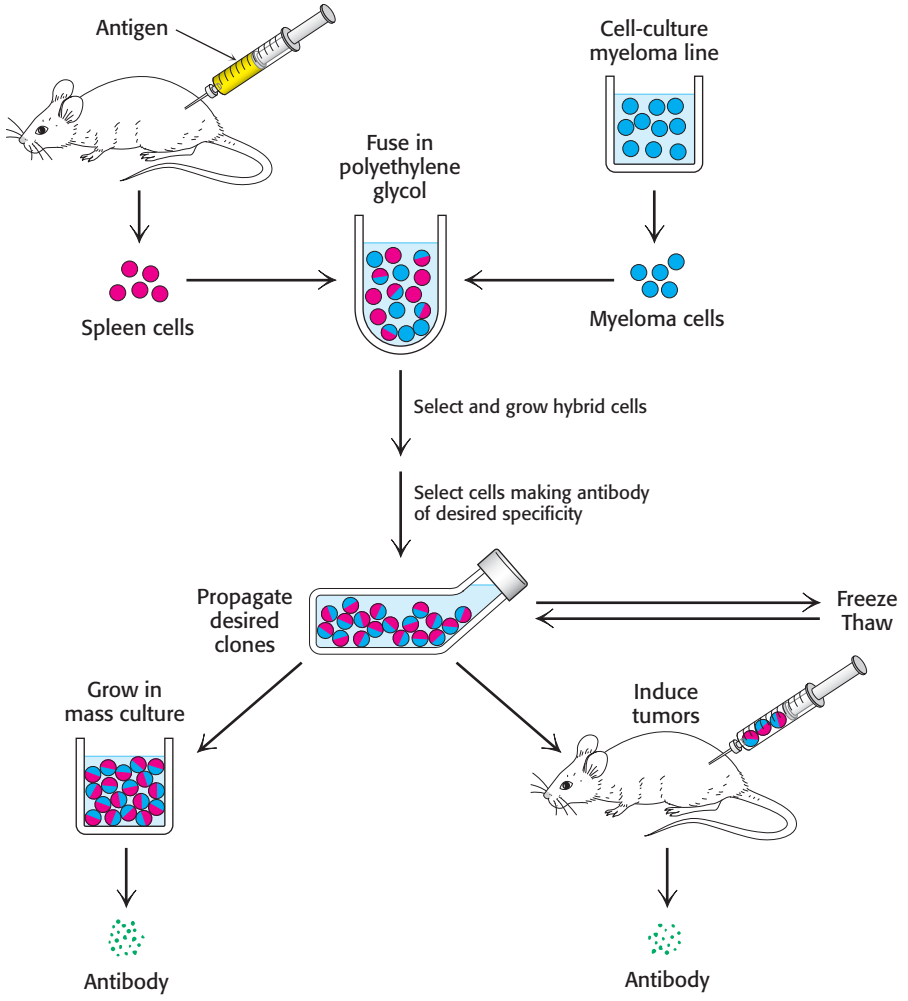


Figure 3.27 Preparation of monoclonal antibodies. Hybridoma cells are formed by the fusion of antibody-producing cells and myeloma cells. The hybrid cells are allowed to proliferate by growing them in selective medium. They are then screened to determine which ones produce antibody of the desired specificity. [After C. Milstein. Monoclonal antibodies. Copyright © 1980 by Scientific American, Inc. All rights reserved.]

Figure 3.28 Fluorescence micrograph of a developing *Drosophila* embryo. The embryo was stained with a fluorescence-labeled monoclonal antibody for the DNA-binding protein encoded by *engrailed,* an essential gene in specifying the body plan. [Courtesy of Dr. Nipam Patel and Dr. Corey Goodman.]

produce antibodies of the preferred specificity. Collections of cells shown to produce the desired antibody are subdivided and reassayed. This process is repeated until a pure cell line, a clone producing a single antibody, is isolated. These positive cells can be grown in culture medium or injected into mice to induce myelomas. Alternatively, the cells can be frozen and stored for long periods.

The hybridoma method of producing monoclonal antibodies has opened new vistas in biology and medicine. Large amounts of identical antibodies with tailor-made specificities can be readily prepared. They are sources of insight into relations between antibody structure and specificity. Moreover, monoclonal antibodies can serve as precise analytical and preparative reagents. Proteins that guide development have been identified with the use of monoclonal antibodies as tags (Figure 3.28). Monoclonal antibodies attached to solid supports can be used as affinity columns to purify scarce proteins. This method has been used to purify interferon (an antiviral protein) 5000-fold from a crude mixture. Clinical laboratories are using monoclonal antibodies in many assays. For example, the detection in blood of isozymes that are normally localized in the heart points to a myocardial infarction (heart attack). Blood transfusions have been made safer by antibody screening of donor blood for viruses that cause AIDS (acquired immune deficiency syndrome), hepatitis, and other infectious diseases. Monoclonal antibodies can be used as therapeutic agents. For example, trastuzumab (Herceptin) is a monoclonal antibody useful for treating some forms of breast cancer.

Proteins can be detected and quantified by using an enzyme-linked immunosorbent assay

Antibodies can be used as exquisitely specific analytic reagents to quantify the amount of a protein or other antigen present in a biological sample. The *enzyme-linked immunosorbent assay* (ELISA) makes use of an enzyme that reacts with a colorless substrate to produce a colored product. The enzyme is covalently linked to a specific antibody that recognizes a target antigen. If the antigen is present, the antibody–enzyme complex will bind to it and, on addition of the substrate, the enzyme will catalyze the reaction, generating the colored product. Thus, the presence of the colored product indicates the presence of the antigen. Rapid and convenient, ELISAs can detect less than a nanogram (10^{-9} g) of a specific protein. ELISA can be performed with either polyclonal or monoclonal antibodies, but the use of monoclonal antibodies yields more-reliable results.

We will consider two among the several types of ELISA. *The indirect ELISA is used to detect the presence of antibody* and is the basis of the test for HIV infection. The HIV test detects the presence of antibodies that recognize viral core protein antigens. Viral core proteins are adsorbed to the bottom of a well. Antibodies from the person being tested are then added to the coated well. Only someone infected with HIV will have antibodies that bind to the antigen. Finally, enzyme-linked antibodies to human antibodies (e.g., enzyme-linked goat antibodies that recognize human antibodies) are allowed to react in the well, and unbound antibodies are removed by washing. Substrate is then applied. An enzyme reaction yielding a colored product suggests that the enzyme-linked antibodies were bound to human antibodies, which in turn implies that the patient has antibodies to the viral antigen (Figure 3.29A). Moreover, this assay is quantitative: the rate of the color-formation reaction is proportional to the amount of antibody originally present.

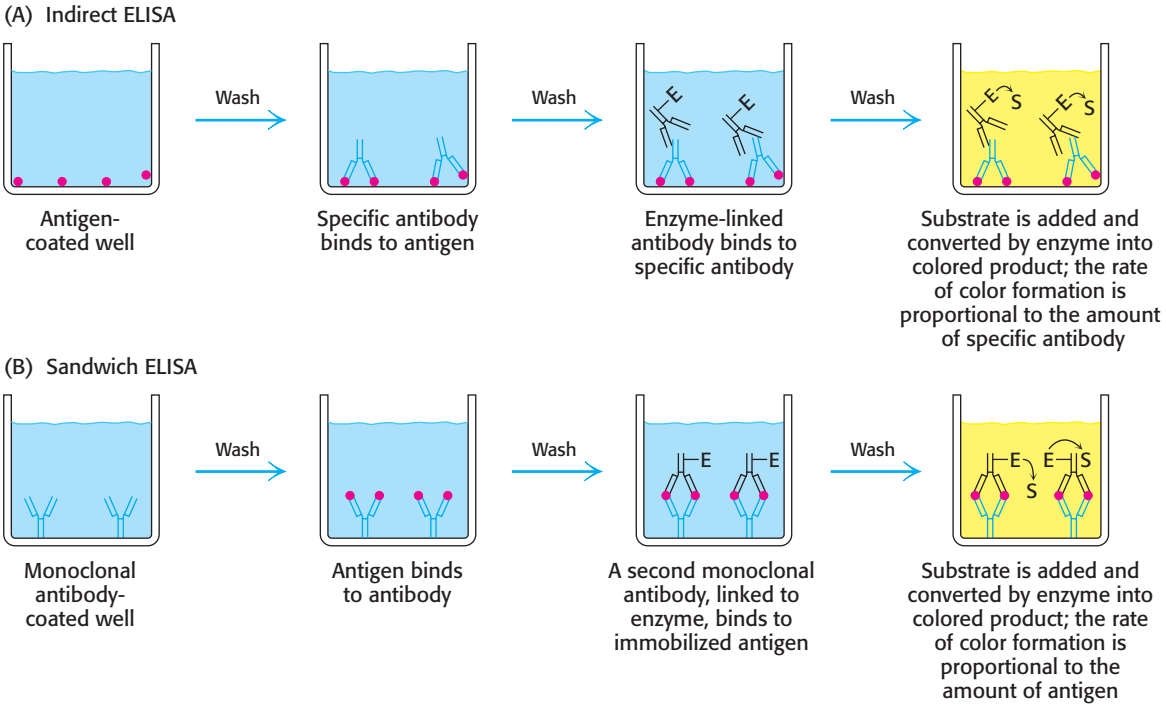


Figure 3.29 Indirect ELISA and sandwich ELISA. (A) In indirect ELISA, the production of color indicates the amount of an antibody to a specific antigen. (B) In sandwich ELISA, the production of color indicates the quantity of antigen. [After R. A. Goldsby, T. J. Kindt, and B. A. Osborne, *Kuby Immunology,* 4th ed. (W. H. Freeman and Company, 2000), p. 162.]

The sandwich ELISA is used to detect antigen rather than antibody. Antibody to a particular antigen is first adsorbed to the bottom of a well. Next, solution containing the antigen (such as blood or urine, in medical diagnostic tests) is added to the well and binds to the antibody. Finally, a second, different antibody to the antigen is added. This antibody is enzyme linked and is processed as described for indirect ELISA. In this case, the rate of color formation is directly proportional to the amount of antigen present. Consequently, it permits the measurement of small quantities of antigen (Figure 3.29B).

Western blotting permits the detection of proteins separated by gel electrophoresis

Very small quantities of a protein of interest in a cell or in body fluid can be detected by an immunoassay technique called *western blotting* (Figure 3.30). A sample is subjected to electrophoresis on an SDS–polyacrylamide gel. A polymer sheet is pressed against the gel, transferring the resolved proteins on the gel to the sheet, which makes the proteins more accessible for reaction. An antibody that is specific for the protein of interest is added to the sheet and reacts with the antigen. The antibody–antigen complex on the sheet can then be detected by rinsing the sheet with a second antibody specific for the first (e.g., goat antibody that recognizes mouse antibody). A radioactive or fluorescent label on the second antibody enables the identification and quantitation of the protein of interest. Alternatively, an enzyme on the second antibody generates a colored product, as in the ELISA method. Western blotting makes it possible to find a protein in a complex mixture, the proverbial needle in a haystack. It is the basis for the test for infection by hepatitis C, where it is used to detect a core protein of the virus. This technique is also very useful in monitoring protein purification and in the cloning of genes.

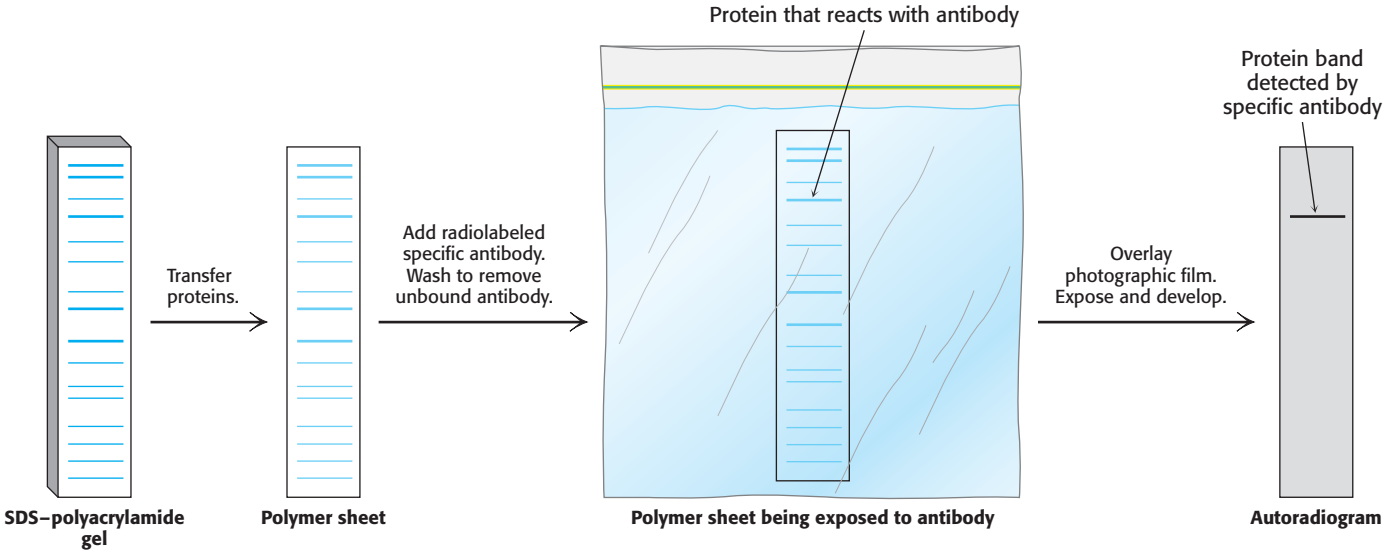


Figure 3.30 Western blotting. Proteins on an SDS–polyacrylamide gel are transferred to a polymer sheet and stained with radioactive antibody. A band corresponding to the protein to which the antibody binds appears in the autoradiogram.

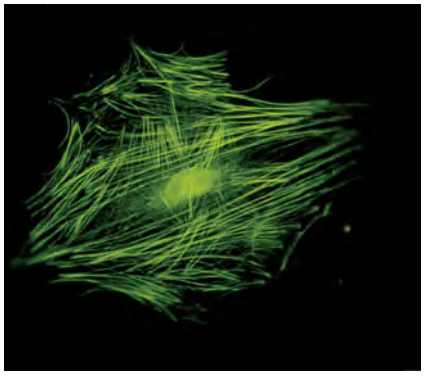


Figure 3.31 Actin filaments. Fluorescence micrograph of actin filaments in a cell stained with an antibody specific to actin. [Courtesy of Dr. Elias Lazarides.]

Fluorescent markers make the visualization of proteins in the cell possible

Biochemistry is often performed in test tubes or polyacrylamide gels. However, most proteins function in the context of a cell. Fluorescent markers provide a powerful means of examining proteins in their biological context. Cells can be stained with fluorescence-labeled antibodies and examined by *fluorescence microscopy* to reveal the location of a protein of interest. For example, arrays of parallel bundles are evident in cells stained with antibody specific for actin, a protein that polymerizes into filaments (Figure 3.31). Actin filaments are constituents of the cytoskeleton, the internal scaffolding of cells that controls their shape and movement. By tracking protein location, fluorescent markers also provide clues to protein function. For instance, the glucocorticoid receptor protein binds to the steroid hormone cortisone. The receptor was linked to *green fluorescent protein* (GFP), a naturally fluorescent protein isolated from the jellyfish *Aequorea victoria* (Chapter 2). Fluorescence microscopy revealed that, in the absence of the hormone, the receptor is located in the cytoplasm (Figure 3.32A). On addition of the steroid, the receptor is translocated to the

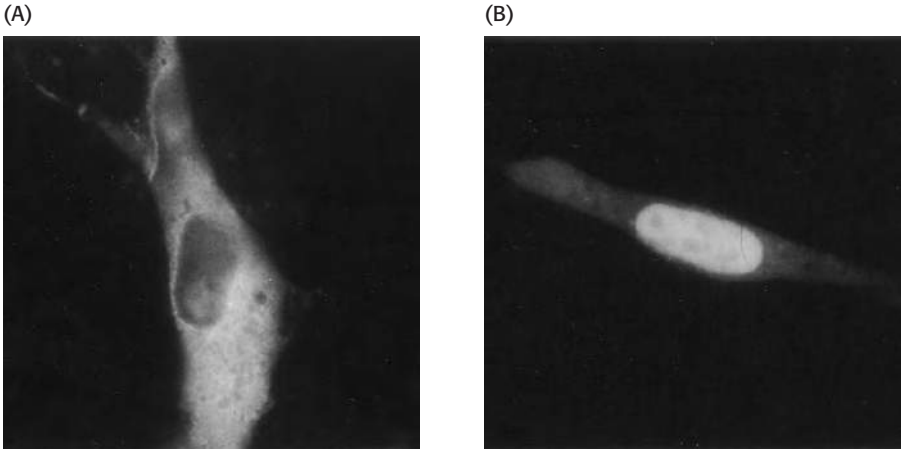


Figure 3.32 Nuclear localization of a steroid receptor. (A) The receptor, made visible by attachment of the green fluorescent protein, is located predominantly in the cytoplasm of the cultured cell. (B) Subsequent to the addition of corticosterone (a glucocorticoid steroid), the receptor moves into the nucleus. [Courtesy of Dr. William B. Pratt.]

nucleus, where it binds to DNA (Figure 3.32B). These results suggested that glucocorticoid receptor protein is a transcription factor that controls gene expression.

The highest resolution of fluorescence microscopy is about 0.2 μm (200 nm, or 2000 Å), the wavelength of visible light. Finer spatial resolution can be achieved by electron microscopy if the antibodies are tagged with electron-dense markers. For example, antibodies conjugated to clusters of gold or to ferritin (which has an electron-dense core rich in iron) are highly visible under the electron microscope. *Immunoelectron microscopy* can define the position of antigens to a resolution of 10 nm (100 Å) or finer (Figure 3.33).

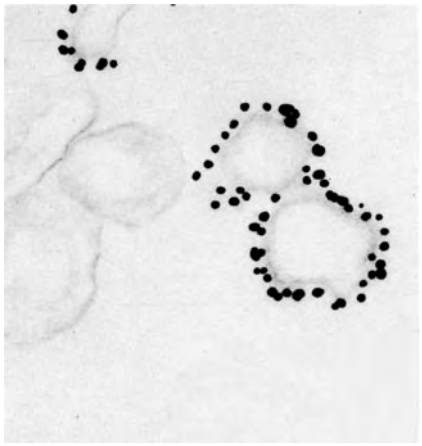


Figure 3.33 Immunoelectron microscopy. The opaque particles (150-Å, or 15-nm, diameter) in this electron micrograph are clusters of gold atoms bound to antibody molecules. These membrane vesicles from the synapses of neurons contain a channel protein that is recognized by the specific antibody. [Courtesy of Dr. Peter Sargent.]

3.4 Mass Spectrometry Is a Powerful Technique for the Identification of Peptides and Proteins

In many instances, the study of a particular biological process in its native context is advantageous. For example, if we are interested in a pathway that is localized to the nucleus of a cell, we might conduct studies on an isolated nuclear extract. In these experiments, identification of the proteins present in the sample is often critical. Antibody-based techniques, such as the ELISA method described in Section 3.3, can be very helpful toward this goal. However, these techniques are limited to the detection of proteins for which an antibody is already available. Mass spectrometry enables the highly precise and sensitive measurement of the atomic composition of a particular molecule, or *analyte,* without prior knowledge of its identity. Originally, this method was relegated to the study of the chemical composition and molecular mass of gases or volatile liquids. However, technological advances in the past two decades have dramatically expanded the utility of mass spectrometry to the study of proteins, even those found at very low concentrations within highly complex mixtures, such as the contents of a particular cell type.

The mass of a protein can be precisely determined by mass spectrometry

Mass spectrometry enables the highly accurate and sensitive detection of the mass of an analyte. This information can be used to determine the identity and chemical state of the molecule of interest. Mass spectrometers operate by converting analyte molecules into gaseous, charged forms (*gas-phase ions*). Through the application of electrostatic potentials, the ratio of the mass of each ion to its charge (the *mass-to-charge ratio*, or m/z) can be measured. Although a wide variety of techniques employed by mass spectrometers are used in current practice, each of them comprises three essential components: the ion source, the mass analyzer, and the detector. Let us consider the first two in greater detail, because improvements in them have contributed most significantly to the analysis of biological samples.

The *ion source* achieves the first critical step in mass spectrometric analysis: conversion of the analyte into gas-phase ions (*ionization*). Until recently, proteins could not be ionized efficiently because of their high molecular weights and low volatility. However, the development of techniques such as *matrix-assisted laser desorption/ionization* (MALDI) and *electrospray ionization* (ESI) has enabled the clearing of this significant hurdle. In MALDI, the analyte is evaporated to dryness in the presence of a volatile, aromatic compound (the *matrix*) that can absorb light at specific wavelengths. A laser pulse tuned to one of these wavelengths excites and

vaporizes the matrix, converting some of the analyte into the gas phase. Subsequent gaseous collisions enable the intermolecular transfer of charge, ionizing the analyte. In ESI, a solution of the analyte is passed through an electrically charged nozzle. Droplets of the analyte, now charged, emerge from the nozzle into a chamber of very low pressure, evaporating the solvent and ultimately yielding the ionized analyte.

The newly formed analyte ions then enter the *mass analyzer,* where they are distinguished on the basis of their mass-to-charge ratios. There are a number of different types of mass analyzers. For this discussion, we will consider one of the simplest, the *time-of-flight* (TOF) *mass analyzer*, in which ions are accelerated through an elongated chamber under a fixed electrostatic potential. Given two ions of identical net charge, the smaller ion will require less time to traverse the chamber than will the larger ion. The mass of each ion can be determined by measuring the time required for each ion to pass through the chamber.

The sequential action of the ion source and the mass analyzer enables the highly sensitive measurement of the mass of potentially massive ions, such as those of proteins. Consider an example of a MALDI ion source coupled to a TOF mass analyzer: the MALDI-TOF mass spectrometer (Figure 3.34). Gas-phase ions generated by the MALDI ion source pass directly into the TOF analyzer, where the mass-to-charge ratios are recorded. In Figure 3.35, the MALDI-TOF mass spectrum of a mixture of 5 pmol each of insulin and lactoglobulin is shown. The masses determined by MALDI-TOF are 5733.9 and 18,364, respectively. A comparison with the calculated values of 5733.5 and 18,388 reveals that MALDI-TOF is clearly an accurate means of determining protein mass.

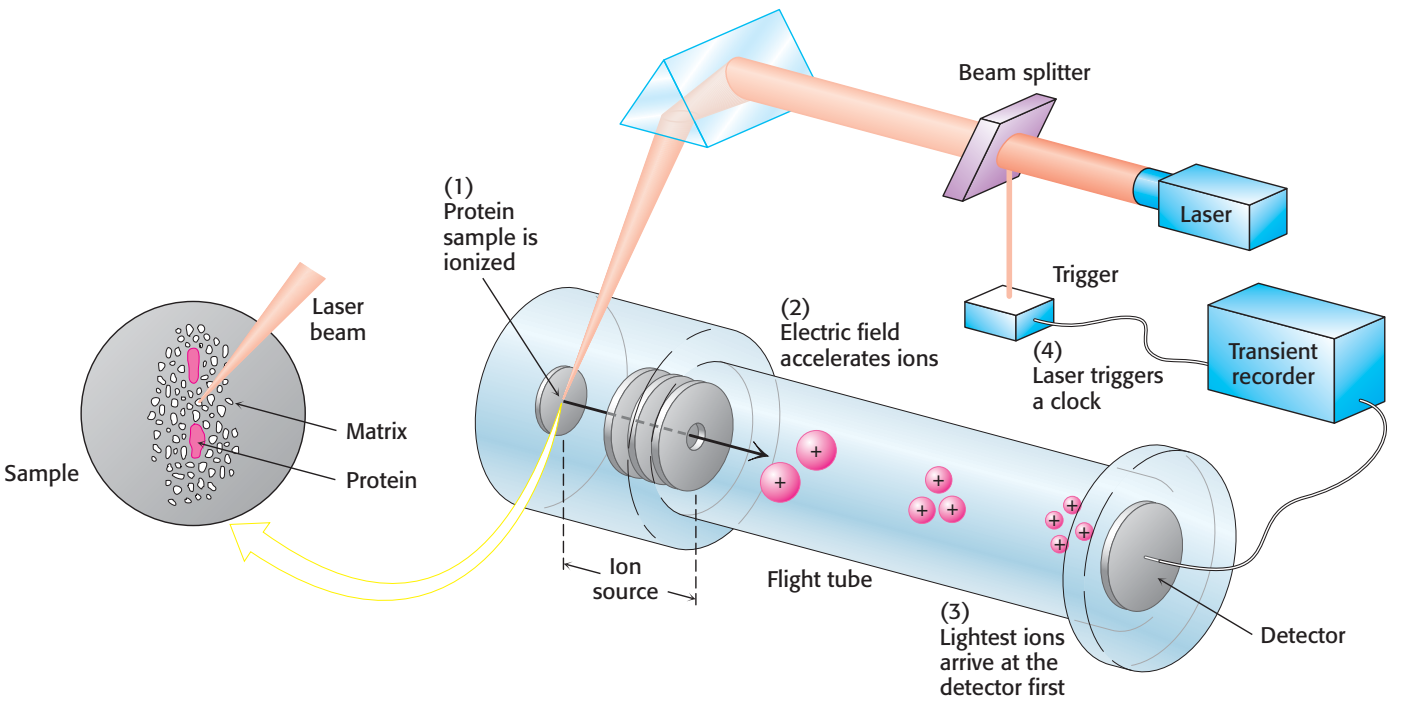


Figure 3.34 MALDI-TOF mass spectrometry. (1) The protein sample, embedded in an appropriate matrix, is ionized by the application of a laser beam. (2) An electric field accelerates the ions through the flight tube toward the detector. (3) The lightest ions arrive first. (4) The ionizing laser pulse also triggers a clock that measures the time of flight (TOF) for the ions. [After J. T. Watson, *Introduction to Mass Spectrometry,* 3d ed. (Lippincott-Raven, 1997), p. 279.]

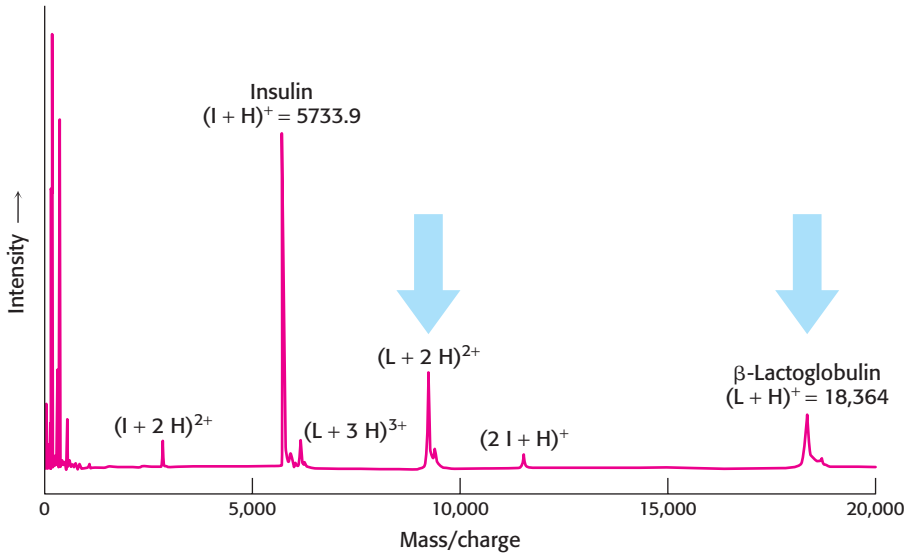


Figure 3.35 MALDI-TOF mass spectrum of insulin and β-lactoglobulin. A mixture of 5 pmol each of insulin (I) and β-lactoglobulin (L) was ionized by MALDI, which produces predominately singly charged molecular ions from peptides and proteins—the insulin ion $(I + H)^+$ and the lactoglobulin ion $(L + H)^+$. Molecules with multiple charges, such as those for β-lactoglobulin indicated by the blue arrows, as well as small quantities of a singly charged dimer of insulin $(2\ I + H)^+$ also are produced. [After J. T. Watson, *Introduction to Mass Spectrometry,* 3d ed. (Lippincott-Raven, 1997), p. 282.]

In the ionization process, a family of ions, each of the same mass but carrying different total net charges, is formed from a single analyte. Because the mass spectrometer detects ions on the basis of their mass-to-charge ratio, these ions will appear as separate peaks in the mass spectrum. For example, in the mass spectrum of β-lactoglobulin shown in Figure 3.35, peaks near $m/z = 18{,}388$ (corresponding to the +1 charged ion) and $m/z = 9{,}194$ (corresponding to the +2 charged ion) are visible (indicated by the blue arrows). Although multiple peaks for the same ion may appear to be a nuisance, they enable the spectrometrist to measure the mass of an analyte ion more than once in a single experiment, improving the overall precision of the calculated result.

Peptides can be sequenced by mass spectrometry

Earlier in this chapter, the Edman degradation was presented as a method for identifying the sequence of a peptide. Mass spectrometry of peptide fragments is an alternative to Edman degradation as a means of sequencing proteins. Ions of proteins that have been analyzed by a mass spectrometer, the *precursor ions,* can be broken into smaller peptide chains by bombardment with atoms of an inert gas such as helium or argon. These new fragments, or *product ions,* can be passed through a second mass analyzer for further mass characterization. The utilization of two mass analyzers arranged in this manner is referred to as *tandem mass spectrometry*.

Importantly, the product-ion fragments are formed in chemically predictable ways that can provide clues to the amino acid sequence of the precursor ion. For peptide analytes, product ions can be formed such that individual amino acid residues are cleaved from the precursor ion (Figure 3.36A). Hence, a family of ions is detected; each ion represents a fragment of the original peptide with one or more amino acids removed from one end.

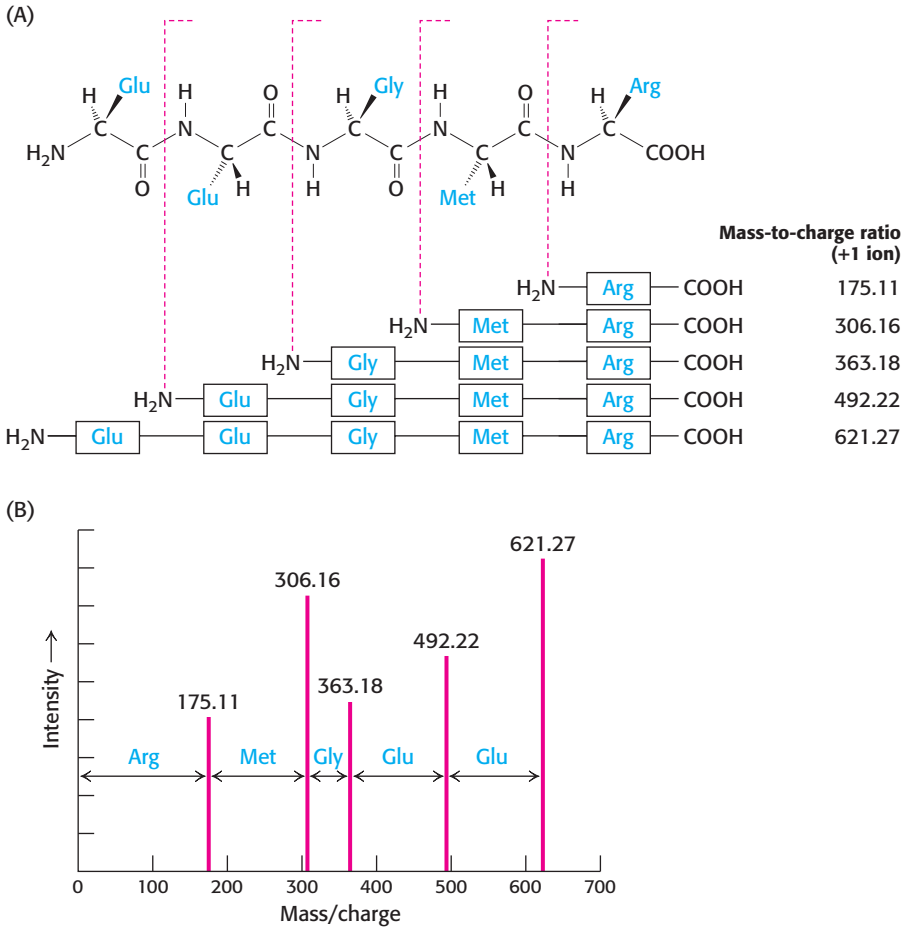


Figure 3.36 Peptide sequencing by tandem mass spectrometry. (A) Within the mass spectrometer, peptides can be fragmented by bombardment with inert gaseous ions to generate a family of product ions in which individual amino acids have been removed from one end. As drawn here, the carboxyl fragment of the cleaved peptide bond is ionized. (B) The product ions are detected in the second mass analyzer. The mass differences between the peaks indicate the sequence of amino acids in the precursor ion. [After H. Steen and M. Mann. *Nat. Rev. Mol. Cell Biol.* 5:699–711, 2004.]

Figure 3.36B depicts a representative mass spectrum from a fragmented peptide. The mass differences between the product ions indicate the amino acid sequence of the precursor peptide ion.

Individual proteins can be identified by mass spectrometry

The combination of the mass spectrometry with the chromatographic and peptide-cleavage techniques described earlier in this chapter enables highly sensitive protein identification in complex biological mixtures. When a protein is cleaved by chemical or enzymatic methods (see Table 3.3), a specific and predictable family of peptide fragments is formed. We learned in Chapter 2 that each protein has a unique, precisely defined amino acid sequence. Hence, the identity of the individual peptides formed from this cleavage reaction—and, importantly, their corresponding masses—is a distinctive signature for that particular protein. Protein cleavage, followed by chromatographic separation and mass spectrometry, enables rapid identification and quantitation of these signatures, even if they are present at very low concentrations.

As an example of the power of this proteomic approach, consider the analysis of the nuclear-pore complex from yeast, which facilitates the transport of large molecules into and out of the nucleus. This huge macromolecular

complex was purified from yeast cells by careful procedures. The purified complex was fractionated by HPLC followed by gel electrophoresis. Individual bands from the gel were isolated, cleaved with trypsin, and analyzed by MALDI-TOF mass spectrometry. The fragments produced were compared with amino acid sequences deduced from the DNA sequence of the yeast genome as shown in Figure 3.37. A total of 174 nuclear-pore proteins were identified in this manner. Many of these proteins had not previously been identified as being associated with the nuclear pore despite years of study. Furthermore, mass spectrometric methods are sensitive enough to detect essentially all components of the pore if they are present in the samples used. Thus, a complete list of the components constituting this macromolecular complex could be obtained in a straightforward manner. Proteomic analysis of this type is growing in power as mass spectrometric and biochemical fractionation methods are refined.

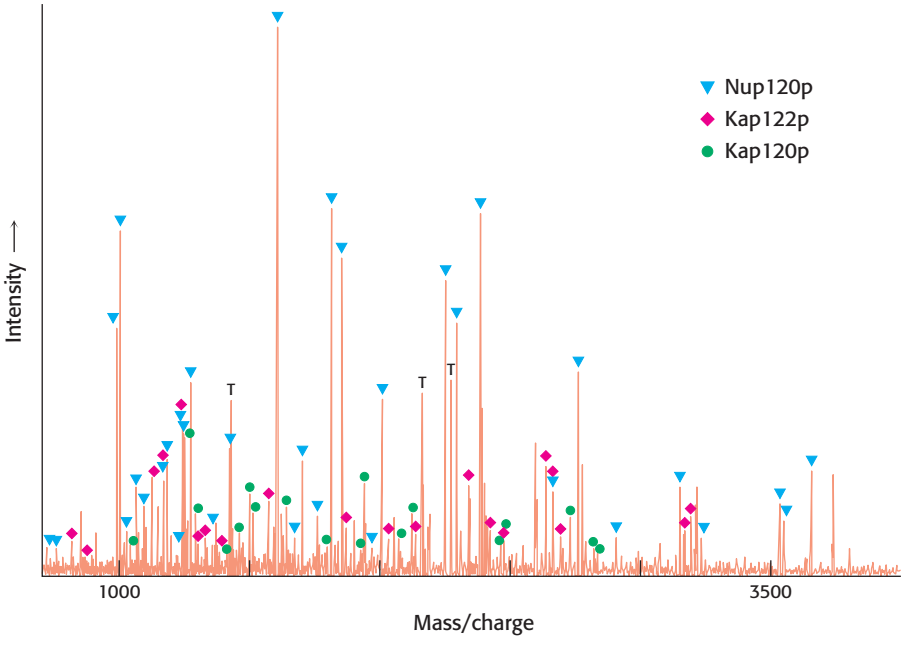


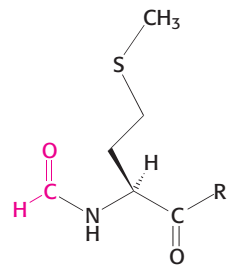
Figure 3.37 Proteomic analysis by mass spectrometry. This mass spectrum was obtained by analyzing a trypsin-treated band in a gel derived from a yeast nuclear-pore sample. Many of the peaks were found to match the masses predicted for peptide fragments from three proteins (Nup120p, Kap122p, and Kap120p) within the yeast genome. The band corresponded to an apparent molecular mass of 100 kd. [From M. P. Rout, J. D. Aitchison, A. Suprapto, K. Hjertaas, Y. Zhao, and B. T. Chait. *J. Cell Biol.* 148:635–651, 2000.]

3.5 Peptides Can Be Synthesized by Automated Solid-Phase Methods

Peptides of defined sequence can be synthesized to assist in biochemical analysis. These peptides are valuable tools for several purposes.

1. *Synthetic peptides can serve as antigens to stimulate the formation of specific antibodies*. Suppose we want to isolate the protein expressed by a specific gene. Peptides can be synthesized that match the translation of part of the gene's nucleic acid sequence, and antibodies can be generated that target these peptides. These antibodies can then be used to isolate the intact protein or localize it within the cell.

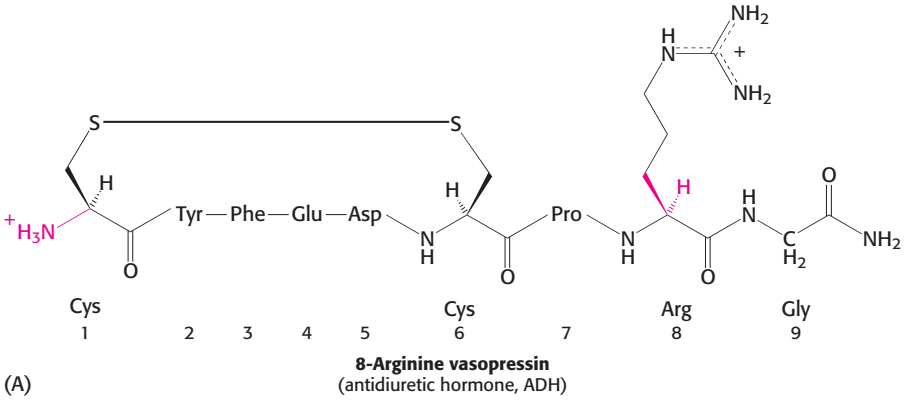
2. *Synthetic peptides can be used to isolate receptors for many hormones and other signal molecules*. For example, white blood cells are attracted to bacteria by formylmethionyl (fMet) peptides released in the breakdown of bacterial proteins. Synthetic formylmethionyl peptides have been useful in identifying the white blood cell's cell-surface receptor for this class of peptide.



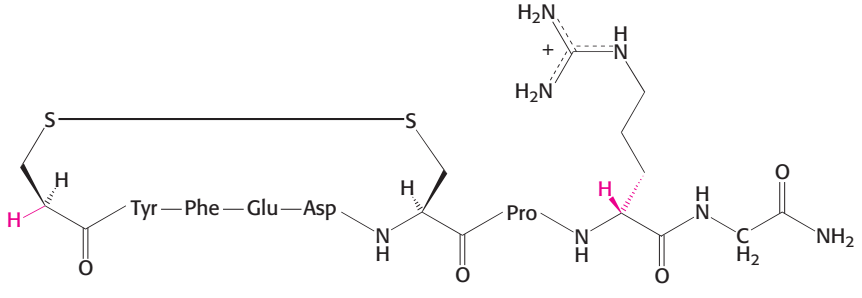
fMet peptide

Moreover, synthetic peptides can be attached to agarose beads to prepare affinity chromatography columns for the purification of receptor proteins that specifically recognize the peptides.

3. *Synthetic peptides can serve as drugs.* Vasopressin is a peptide hormone that stimulates the reabsorption of water in the distal tubules of the kidney, leading to the formation of more-concentrated urine. Patients with diabetes insipidus are deficient in vasopressin (also called *antidiuretic hormone*), and so they excrete large volumes of dilute urine (more than 5 liters per day) and are continually thirsty. This defect can be treated by administering 1-desamino-8-D-arginine vasopressin, a synthetic analog of the missing hormone (Figure 3.38). This synthetic peptide is degraded in vivo much more slowly than vasopressin and does not increase blood pressure.

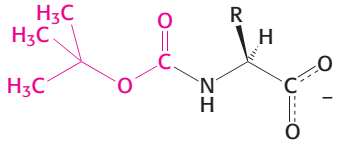


(A) **8-Arginine vasopressin** (antidiuretic hormone, ADH)

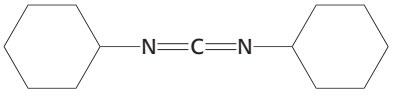


(B) **1-Desamino-8-D-arginine vasopressin**

Figure 3.38 Vasopressin and a synthetic vasopressin analog. Structural formulas of (A) vasopressin, a peptide hormone that stimulates water resorption, and (B) 1-desamino-8-D-arginine vasopressin, a more stable synthetic analog of this antidiuretic hormone.



***t*-Butyloxycarbonyl amino acid (*t*-Boc amino acid)**



Dicyclohexylcarbodiimide (DCC)

4. Finally, *studying synthetic peptides can help define the rules governing the three-dimensional structure of proteins.* We can ask whether a particular sequence by itself tends to fold into an α helix, a β strand, or a hairpin turn or behaves as a random coil. The peptides created for such studies can incorporate amino acids not normally found in proteins, allowing more variation in chemical structure than is possible with the use of only 20 amino acids.

How are these peptides constructed? The amino group of one amino acid is linked to the carboxyl group of another. However, a unique product is formed only if a single amino group and a single carboxyl group are available for reaction. Therefore, it is necessary to block some groups and to activate others to prevent unwanted reactions. First, the carboxyl-terminal amino acid is attached to an insoluble resin by its carboxyl group, effectively protecting it from further peptide-bond-forming reactions (Figure 3.39).

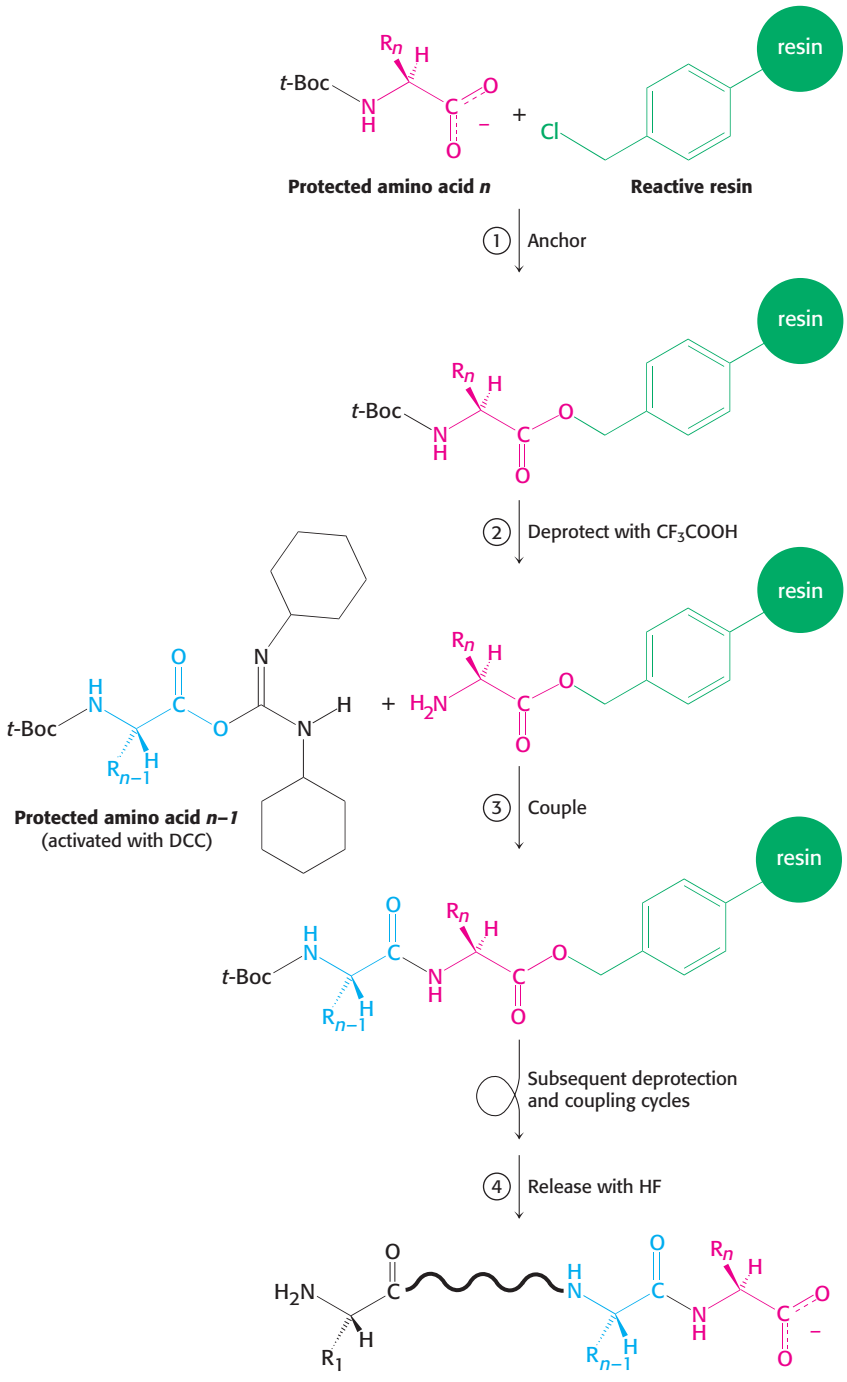


Figure 3.39 Solid-phase peptide synthesis. The sequence of steps in solid-phase synthesis is: (1) anchoring of the C-terminal amino acid to a solid resin, (2) deprotection of the amino terminus, and (3) coupling of the free amino terminus with the DCC-activated carboxyl group of the next amino acid. Steps 2 and 3 are repeated for each added amino acid. Finally, in step 4, the completed peptide is released from the resin.

The α-amino group of this amino acid is blocked with a protecting group such as a *tert*-butyloxycarbonyl (*t*-Boc) group. The *t*-Boc protecting group of this amino acid is then removed with trifluoroacetic acid.

The next amino acid (in the protected *t*-Boc form) and *dicyclohexylcarbodiimide* (DCC) are added together. At this stage, only the carboxyl group of the incoming amino acid and the amino group of the resin-bound amino acid are free to form a peptide bond. DCC reacts with the carboxyl group of the incoming amino acid, activating it for the peptide-bond-forming reaction. After the peptide bond has formed, excess reagents and dicyclohexylurea are washed away, leaving the desired dipeptide product attached to the beads. Additional amino acids are linked by the same sequence of reactions. At the end of the synthesis, the peptide is released from the beads by

the addition of hydrofluoric acid (HF), which cleaves the carboxyl ester anchor without disrupting peptide bonds. Protecting groups on potentially reactive side chains, such as that of lysine, also are removed at this time.

A major advantage of this *solid-phase method,* first developed by R. Bruce Merrifield, is that the desired product at each stage is bound to beads that can be rapidly filtered and washed, and so there is no need to purify intermediates. All reactions are carried out in a single vessel, eliminating losses caused by repeated transfers of products. This cycle of reactions can be readily automated, which makes it feasible to routinely synthesize peptides containing about 50 residues in good yield and purity. In fact, the solid-phase method has been used to synthesize interferons (155 residues) that have antiviral activity and ribonuclease (124 residues) that is catalytically active. The protecting groups and cleavage agents may be varied for increased flexibility or convenience.

Synthetic peptides can be linked to create even longer molecules. With the use of specially developed *peptide-ligation* methods, proteins of 100 amino acids or more can by synthesized in very pure form. These methods enable the construction of even sharper tools for examining protein structure and function.

3.6 Three-Dimensional Protein Structure Can Be Determined by X-ray Crystallography and NMR Spectroscopy

Elucidation of the three-dimensional structure of a protein is often the source of a tremendous amount of insight into its corresponding function, inasmuch as the specificity of active sites and binding sites is defined by the precise atomic arrangement within these regions. For example, knowledge of the structure of a protein enables the biochemist to predict its mechanism of action, the effects of mutations on its function, and the desired features of drugs that may inhibit or augment its activity. X-ray crystallography and nuclear magnetic resonance spectroscopy are the two most important techniques for elucidating the conformation of proteins.

X-ray crystallography reveals three-dimensional structure in atomic detail

X-ray crystallography was the first method developed to determine protein structure in atomic detail. This technique provides the clearest visualization of the precise three-dimensional positions of most atoms within a protein. Of all forms of radiation, x-rays provide the best resolution for the determination of molecular structures because their wavelength approximately corresponds to that of a covalent bond. The three components in an x-ray crystallographic analysis are a *protein crystal,* a *source of x-rays,* and a *detector* (Figure 3.40).

X-ray crystallography first requires the preparation of a protein or protein complex in crystal form, in which all protein molecules are oriented in a fixed, repeated arrangement with respect to one another. Slowly adding ammonium sulfate or another salt to a concentrated solution of protein to reduce its solubility favors the formation of highly ordered crystals—the process of salting out discussed on page 68. For example, myoglobin crystallizes in 3 M ammonium sulfate. Protein crystallization can be quite challenging: a concentrated solution of highly pure material is required and it is often difficult to predict which experimental conditions will yield the most-effective crystals. Methods for screening many different crystallization conditions using a small amount of protein sample have been developed. Typically, hundreds of conditions must be tested to obtain crystals fully suitable for crystallographic studies. Nevertheless, increasingly large

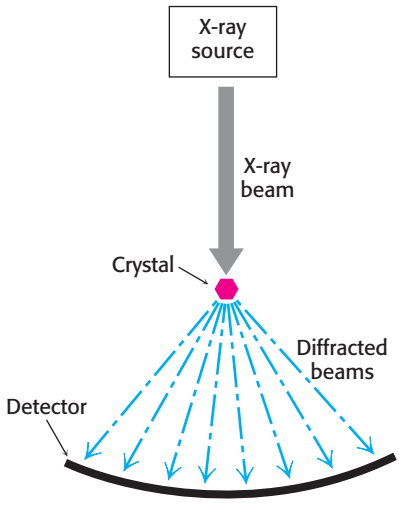


Figure 3.40 An x-ray crystallographic experiment. An x-ray source generates a beam, which is diffracted by a crystal. The resulting diffraction pattern is collected on a detector.

and complex proteins have been crystallized. For example, poliovirus, an 8500-kd assembly of 240 protein subunits surrounding an RNA core, has been crystallized and its structure solved by x-ray methods. Crucially, proteins frequently crystallize in their biologically active configuration. Enzyme crystals may display catalytic activity if the crystals are suffused with substrate.

After a suitably pure crystal of protein has been obtained, a source of x-rays is required. A beam of x-rays of wavelength 1.54 Å is produced by accelerating electrons against a copper target. Equipment suitable for generating x-rays in this manner is available in many laboratories. Alternatively, x-rays can be produced by *synchrotron radiation*, the acceleration of electrons in circular orbits at speeds close to the speed of light. Synchrotron-generated x-ray beams are much more intense than those generated by electrons hitting copper. Several facilities throughout the world generate synchrotron radiation, such as the Advanced Light Source at Argonne National Laboratory outside Chicago and the Photon Factory in Tsukuba City, Japan.

When a narrow beam of x-rays is directed at the protein crystal, most of the beam passes directly through the crystal while a small part is scattered in various directions. These scattered, or *diffracted*, x-rays can be detected by x-ray film or by a solid-state electronic detector. The scattering pattern provides abundant information about protein structure. The basic physical principles underlying the technique are:

1. *Electrons scatter x-rays*. The amplitude of the wave scattered by an atom is proportional to its number of electrons. Thus, a carbon atom scatters six times as strongly as a hydrogen atom does.

2. *The scattered waves recombine*. Each diffracted beam comprises waves scattered by each atom in the crystal. The scattered waves reinforce one another at the film or detector if they are in phase (in step) there, and they cancel one another if they are out of phase.

3. *The way in which the scattered waves recombine depends only on the atomic arrangement*.

The protein crystal is mounted and positioned in a precise orientation with respect to the x-ray beam and the film. The crystal is rotated so that the beam can strike the crystal from many directions. This rotational motion results in an x-ray photograph consisting of a regular array of spots called *reflections*. The x-ray photograph shown in Figure 3.41 is a two-dimensional section through a three-dimensional array of 25,000 reflections. The intensities and positions of these reflections are the basic experimental data of an x-ray crystallographic analysis. Each reflection is formed from a wave with an amplitude proportional to the square root of the observed intensity of the spot. Each wave also has a *phase*—that is, the timing of its crests and troughs relative to those of other waves. Additional experiments or calculations must be performed to determine the phases corresponding to each reflection.

The next step is to reconstruct an image of the protein from the observed reflections. In light microscopy or electron microscopy, the diffracted beams are focused by lenses to directly form an image. However, appropriate lenses for focusing x-rays do not exist. Instead, the image is formed by applying a mathematical relation called a *Fourier transform* to the measured amplitudes and calculated phases of every observed reflection. The image obtained is referred to as the *electron-density map*. It is a three-dimensional graphic representation of where the electrons are most densely localized and is used to determine the positions of the atoms in the crystallized molecule

Figure 3.41 An x-ray diffraction pattern. X-ray precession photograph from a crystal of myoglobin. [Mel Pollinger/Fran Heyl Associates.]

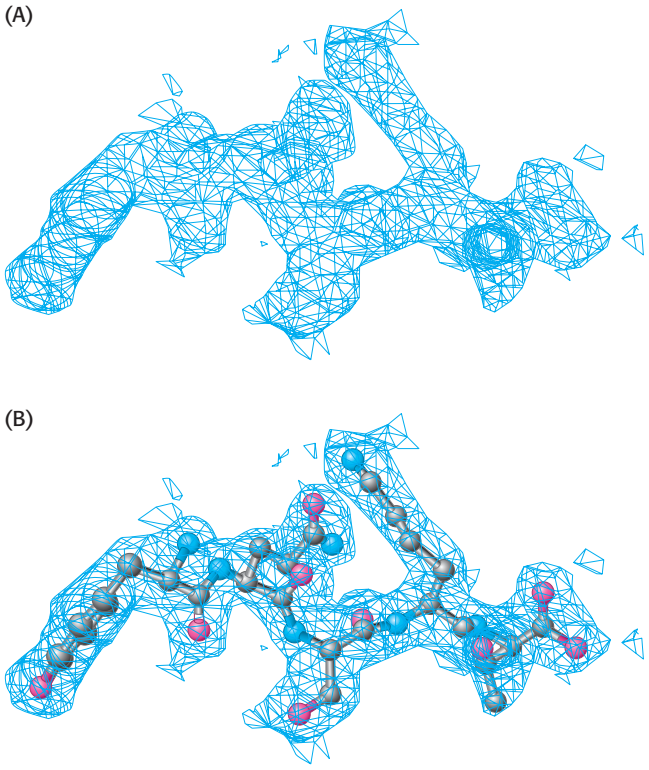


Figure 3.42 Interpretation of an electron-density map. (A) A segment of an electron-density map is drawn as a three-dimensional contour plot, in which the regions inside the "cage" represent the regions of highest electron density. (B) A model of the protein is built into this map so as to maximize the placement of atoms within this density. [Drawn from 1FCH.pdb.]

(Figure 3.42). Critical to the interpretation of the map is its *resolution,* which is determined by the number of scattered intensities used in the Fourier transform. The fidelity of the image depends on this resolution, as shown by the optical analogy in Figure 3.43. A resolution of 6 Å reveals the course of the polypeptide chain but few other structural details. The reason is that polypeptide chains pack together so that their centers are between 5 Å and 10 Å apart. Maps at higher resolution are needed to delineate groups of atoms, which lie between 2.8 Å and 4.0 Å apart, and individual atoms, which are between 1.0 Å and 1.5 Å apart. The ultimate resolution of an x-ray analysis is determined by the degree of perfection of the crystal. For proteins, this limiting resolution is often about 2 Å.

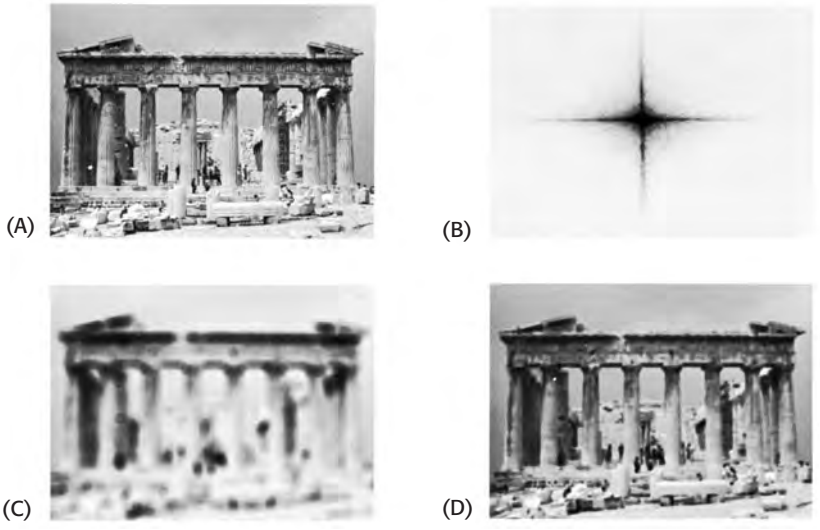


Figure 3.43 Resolution affects the quality of an image. The effect of resolution on the quality of a reconstructed image is shown by an optical analog of x-ray diffraction: (A) a photograph of the Parthenon; (B) an optical diffraction pattern of the Parthenon; (C and D) images reconstructed from the pattern in part B. More data were used to obtain image D than image C, which accounts for the higher quality of image D. [Courtesy of Dr. Thomas Steitz (part A) and Dr. David DeRosier (part B).]

Nuclear magnetic resonance spectroscopy can reveal the structures of proteins in solution

X-ray crystallography is the most powerful method for determining protein structures. However, some proteins do not readily crystallize. Furthermore, although structures present in crystallized proteins very closely represent those of proteins free of the constraints imposed by the crystalline environment, structures in solution can be sources of additional insights. *Nuclear magnetic resonance* (NMR) *spectroscopy* is unique in being able to reveal the atomic structure of macromolecules *in solution,* provided that highly concentrated solutions (~1 mM, or 15 mg ml^{-1} for a 15-kd protein) can be obtained. This technique depends on the fact that certain atomic nuclei are intrinsically magnetic. Only a limited number of isotopes display this property, called *spin,* and those most important to biochemistry are listed in Table 3.4. The simplest example is the hydrogen nucleus (^1H), which is a proton. The spinning of a proton generates a magnetic moment. This moment can take either of two orientations, or spin states (called α and β), when an external magnetic field is applied (Figure 3.44). The energy difference between these states is proportional to the strength of the imposed magnetic field. The α state has a slightly lower energy because it is aligned with this applied field. Hence, in a given population of nuclei, slightly more will occupy the α state (by a factor of the order of 1.00001 in a typical experiment). A spinning proton in an α state can be raised to an excited state (β state) by applying a pulse of electromagnetic radiation (a radio-frequency, or RF, pulse), provided that the frequency corresponds to the energy difference between the α and the β states. In these circumstances, the spin will change from α to β; in other words, *resonance* will be obtained.

These properties can be used to examine the chemical surroundings of the hydrogen nucleus. The flow of electrons around a magnetic nucleus generates a small local magnetic field that opposes the applied field. The degree of such shielding depends on the surrounding electron density. Consequently, nuclei in different environments will change states, or resonate, at slightly different field strengths or radiation frequencies. A resonance spectrum for a molecule is obtained by keeping the magnetic field constant and varying the frequency of the electromagnetic radiation. The nuclei of the perturbed sample absorb electromagnetic radiation at a frequency that can be measured. The different frequencies, termed *chemical shifts,* are expressed in fractional units δ (parts per million, or ppm) relative to the shifts of a standard compound, such as a water-soluble derivative of tetramethylsilane, that is added with the sample. For example, a $—CH_3$ proton typically exhibits a chemical shift (δ) of 1 ppm, compared with a chemical shift of 7 ppm for an aromatic proton. The chemical shifts of most protons in protein molecules fall between 0 and 9 ppm (Figure 3.45). Most protons in many proteins can be resolved by using this technique of *one-dimensional NMR*. With this information, we can then deduce changes to a particular chemical group under different conditions, such as the conformational change of a protein from a disordered structure to an α helix in response to a change in pH.

We can garner even more information by examining how the spins on different protons affect their neighbors. By inducing a transient magnetization in a sample through the application of a radio-frequency pulse, we can alter the spin on one nucleus and examine the effect on the spin of a neighboring nucleus. Especially revealing is a *two-dimensional spectrum obtained by nuclear Overhauser enhancement spectroscopy* (NOESY), *which graphically displays pairs of protons that are in close proximity,* even if they are not close together in the primary structure. The basis for this technique is the

Table 3.4 Biologically important nuclei giving NMR signals

Nucleus	Natural abundance (% by weight of the element)
^1H	99.984
^2H	0.016
^{13}C	1.108
^{14}N	99.635
^{15}N	0.365
^{17}O	0.037
^{23}Na	100.0
^{25}Mg	10.05
^{31}P	100.0
^{35}Cl	75.4
^{39}K	93.1

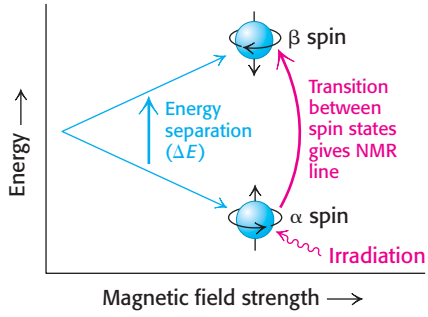


Figure 3.44 Basis of NMR spectroscopy. The energies of the two orientations of a nucleus of spin $^1/_2$ (such as ^{31}P and ^1H) depend on the strength of the applied magnetic field. Absorption of electromagnetic radiation of appropriate frequency induces a transition from the lower to the upper level.

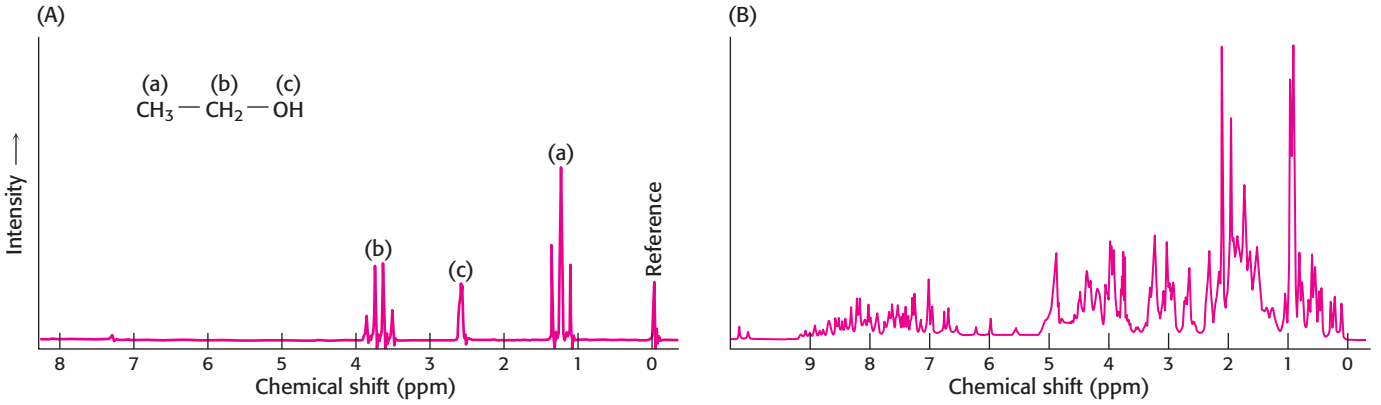


Figure 3.45 One-dimensional NMR spectra. (A) ^{1}H-NMR spectrum of ethanol (CH_3CH_2OH) shows that the chemical shifts for the hydrogen are clearly resolved. (B) ^{1}H-NMR spectrum of a 55 amino acid fragment of a protein having a role in RNA splicing shows a greater degree of complexity. A large number of peaks are present and many overlap. [(A) After C. Branden and J. Tooze, *Introduction to Protein Structure* (Garland, 1991), p. 280; (B) courtesy of Dr. Barbara Amann and Dr. Wesley McDermott.]

nuclear Overhauser effect (NOE), an interaction between nuclei that is proportional to the inverse sixth power of the distance between them. Magnetization is transferred from an excited nucleus to an unexcited one if the two nuclei are less than about 5 Å apart (Figure 3.46A). In other words, the effect provides a means of detecting the location of atoms relative to one another in the three-dimensional structure of the protein. The peaks that lie along the diagonal of a NOESY spectrum (shown in white in Figure 3.46B) correspond to those present in a one-dimensional NMR experiment. The peaks apart from the diagonal (shown in red in Figure 3.46B), referred to as *off-diagonal peaks* or *cross-peaks*, provide crucial new information: *they identify pairs of protons that are less than 5 Å apart*. A two-dimensional NOESY spectrum for a protein comprising 55 amino acids is shown in Figure 3.47. The large number of off-diagonal peaks reveals short proton–proton distances. The three-dimensional structure of a protein can be reconstructed with the use of such proximity relations. Structures are

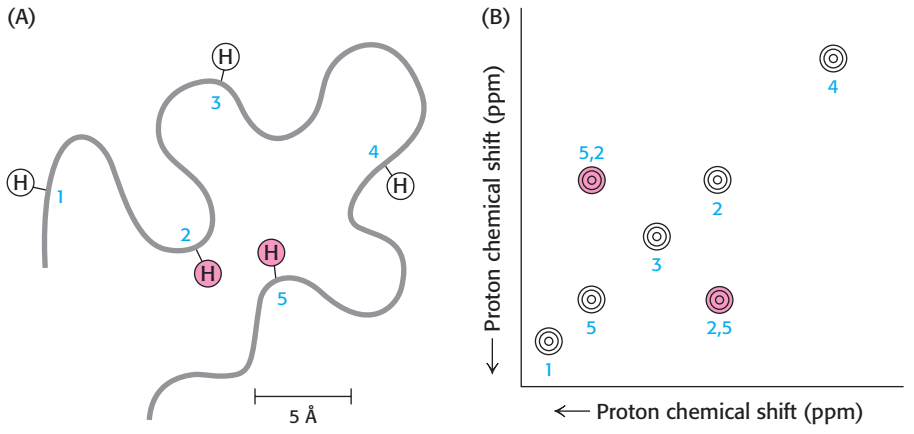


Figure 3.46 The nuclear Overhauser effect. The nuclear Overhauser effect (NOE) identifies pairs of protons that are in close proximity. (A) Schematic representation of a polypeptide chain highlighting five particular protons. Protons 2 and 5 are in close proximity (~4 Å apart), whereas other pairs are farther apart. (B) A highly simplified NOESY spectrum. The diagonal shows five peaks corresponding to the five protons in part A. The peak above the diagonal and the symmetrically related one below reveal that proton 2 is close to proton 5.

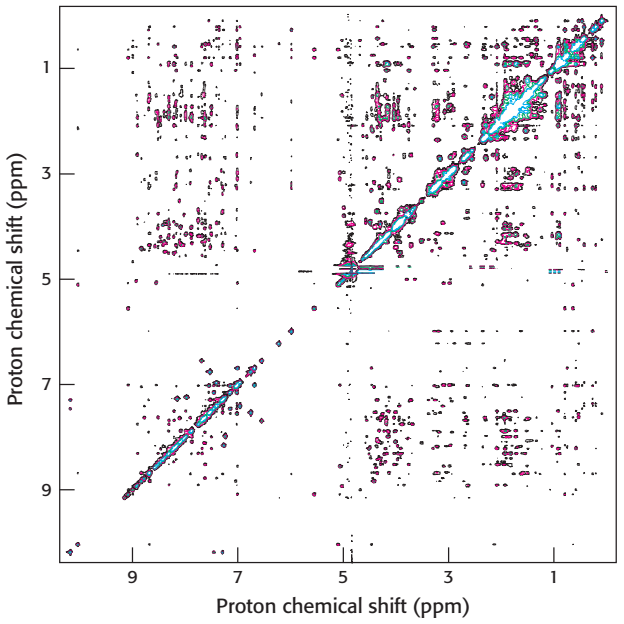


Figure 3.47 Detecting short proton–proton distances. A NOESY spectrum for a 55 amino acid domain from a protein having a role in RNA splicing. Each off-diagonal peak corresponds to a short proton–proton separation. This spectrum reveals hundreds of such short proton–proton distances, which can be used to determine the three-dimensional structure of this domain. [Courtesy of Dr. Barbara Amann and Dr. Wesley McDermott.]

calculated such that protons that must be separated by less than 5 Å on the basis of NOESY spectra are close to one another in the three-dimensional structure (Figure 3.48). If a sufficient number of distance constraints are applied, the three-dimensional structure can nearly be determined uniquely.

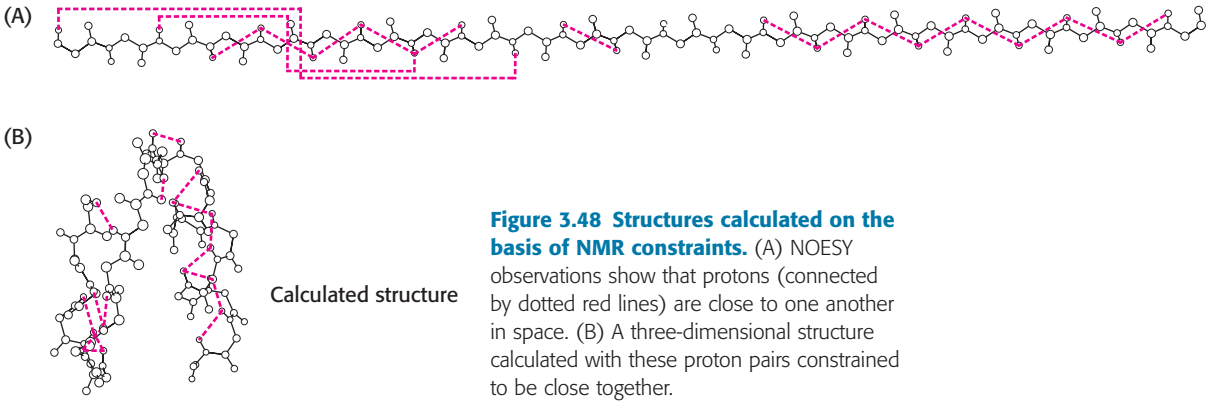


Figure 3.48 Structures calculated on the basis of NMR constraints. (A) NOESY observations show that protons (connected by dotted red lines) are close to one another in space. (B) A three-dimensional structure calculated with these proton pairs constrained to be close together.

In practice, a family of related structures is generated by NMR spectroscopy for three reasons (Figure 3.49). First, not enough constraints may be experimentally accessible to fully specify the structure. Second, the distances obtained from analysis of the NOESY spectrum are only approximate. Finally, the experimental observations are made not on single molecules but on a large number of molecules in solution that may have slightly different structures at any given moment. Thus, the family of structures generated from NMR structure analysis indicates the range of conformations for the protein in solution. At present, NMR spectroscopy can determine

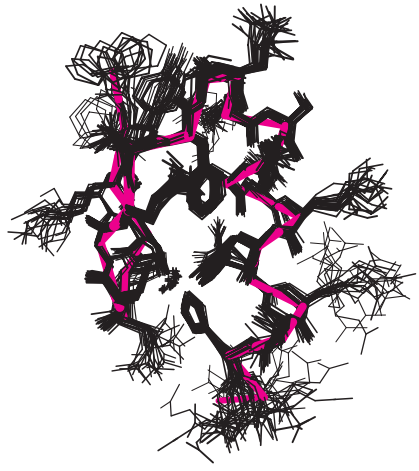


Figure 3.49 A family of structures. A set of 25 structures for a 28 amino acid domain from a zinc-finger-DNA-binding protein. The red line traces the average course of the protein backbone. Each of these structures is consistent with hundreds of constraints derived from NMR experiments. The differences between the individual structures are due to a combination of imperfections in the experimental data and the dynamic nature of proteins in solution. [Courtesy of Dr. Barbara Amann.]

the structures of only relatively small proteins (40 kd), but its resolving power is certain to increase. The power of NMR has been greatly enhanced by the ability of recombinant DNA technology to produce proteins labeled uniformly or at specific sites with ^{13}C, ^{15}N, and ^{2}H (Chapter 5).

The structures of nearly 60,000 proteins had been elucidated by x-ray crystallography and NMR spectroscopy by the end of 2009, and several new structures are now determined each day. The coordinates are collected at the Protein Data Bank (www.pdb.org), and the structures can be accessed for visualization and analysis. Knowledge of the detailed molecular architecture of proteins has been a source of insight into how proteins recognize and bind other molecules, how they function as enzymes, how they fold, and how they evolved. This extraordinarily rich harvest is continuing at a rapid pace and is greatly influencing the entire field of biochemistry as well as other biological and physical sciences.

Summary

The rapid progress in gene sequencing has advanced another goal of biochemistry—elucidation of the proteome. The proteome is the complete set of proteins expressed and includes information about how they are modified, how they function, and how they interact with other molecules.

3.1 The Purification of Proteins Is an Essential First Step in Understanding Their Function

Proteins can be separated from one another and from other molecules on the basis of such characteristics as solubility, size, charge, and binding affinity. SDS–polyacrylamide gel electrophoresis separates the polypeptide chains of proteins under denaturing conditions largely according to mass. Proteins can also be separated electrophoretically on the basis of net charge by isoelectric focusing in a pH gradient. Ultracentrifugation and gel-filtration chromatography resolve proteins according to size, whereas ion-exchange chromatography separates them mainly on the basis of net charge. The high affinity of many proteins for specific chemical groups is exploited in affinity chromatography, in which proteins bind to columns containing beads bearing covalently linked substrates, inhibitors, or other specifically recognized groups. The mass of a protein can be determined by sedimentation-equilibrium measurements.

3.2 Amino Acid Sequences of Proteins Can Be Determined Experimentally

Amino acid sequences are rich in information concerning the kinship of proteins, their evolutionary relationships, and diseases produced by mutations. Knowledge of a sequence provides valuable clues to conformation and function. The amino acid composition of a protein can be ascertained by hydrolyzing the protein into its constituent amino acids in 6 M HCl at 110°C. The amino acids can be separated by ion-exchange chromatography and quantitated by their reaction with ninhydrin or fluorescamine. Amino acid sequences can be determined by Edman degradation, which removes one amino acid at a time from the amino end of a peptide. Longer polypeptide chains are broken into shorter ones for analysis by specifically cleaving them with reagents such as cyanogen bromide, which splits peptide bonds on the carboxyl side of methionine residues, or the enzyme trypsin, which cleaves on the carboxyl side of lysine and arginine residues.

3.3 Immunology Provides Important Techniques with Which to Investigate Proteins

Proteins can be detected and quantitated by highly specific antibodies; monoclonal antibodies are especially useful because they are homogeneous. Enzyme-linked immunosorbent assays and western blots of SDS–polyacrylamide gels are used extensively. Proteins can also be localized within cells by immunofluorescence microscopy and immunoelectron microscopy.

3.4 Mass Spectrometry Is a Powerful Technique for the Identification of Peptides and Proteins

Techniques such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) allow the generation of ions of proteins and peptides in the gas phase. The mass of such protein ions can be determined with great accuracy and precision. Masses determined by these techniques act as protein name tags because the mass of a protein or peptide is precisely determined by its amino acid composition and, hence, by its sequence. Tandem mass spectrometry is an alternative to Edman degradation that enables the rapid and highly accurate sequencing of peptides. Mass spectrometric techniques are central to proteomics because they make it possible to analyze the constituents of large macromolecular assemblies or other collections of proteins.

3.5 Peptides Can Be Synthesized by Automated Solid-Phase Methods

Polypeptide chains can be synthesized by automated solid-phase methods in which the carboxyl end of the growing chain is linked to an insoluble support. The carboxyl group of the incoming amino acid is activated by dicyclohexylcarbodiimide and joined to the amino group of the growing chain. Synthetic peptides can serve as drugs and as antigens to stimulate the formation of specific antibodies. They can also be sources of insight into the relation between amino acid sequence and conformation.

3.6 Three-Dimensional Protein Structure Can Be Determined by X-ray Crystallography and NMR Spectroscopy

X-ray crystallography and nuclear magnetic resonance spectroscopy have greatly enriched our understanding of how proteins fold, recognize other molecules, and catalyze chemical reactions. X-ray crystallography is possible because electrons scatter x-rays. The diffraction pattern produced can be analyzed to reveal the arrangement of atoms in a protein. The three-dimensional structures of tens of thousands of proteins are now known in atomic detail. Nuclear magnetic resonance spectroscopy reveals the structure and dynamics of proteins in solution. The chemical shift of nuclei depends on their local environment. Furthermore, the spins of neighboring nuclei interact with each other in ways that provide definitive structural information. This information can be used to determine complete three-dimensional structures of proteins.

Key Terms

proteome (p. 66)
assay (p. 67)
specific activity (p. 67)
homogenate (p. 67)
salting out (p. 68)
dialysis (p. 69)
gel-filtration chromatography (p. 69)
ion-exchange chromatography (p. 70)
cation exchange (p. 70)

anion exchange (p. 70)
affinity chromatography (p. 70)
high-pressure liquid chromatography (HPLC) (p. 71)
gel electrophoresis (p. 71)
isoelectric point (p. 73)
isoelectric focusing (p. 73)
two-dimensional electrophoresis (p. 74)
sedimentation coefficient (Svedberg unit, S) (p. 76)
Edman degradation (p. 81)
phenyl isothiocyanate (p. 81)
overlap peptide (p. 82)
antibody (p. 84)
antigen (p. 85)
antigenic determinant (epitope) (p. 85)
polyclonal antibody (p. 86)
monoclonal antibody (p. 86)
enzyme-linked immunosorbent assay (ELISA) (p. 88)
western blotting (p. 89)
fluorescence microscopy (p. 90)
green fluorescent protein (GFP) (p. 90)
matrix-assisted laser desorption/ionization (MALDI) (p. 91)
electrospray ionization (ESI) (p. 91)
time-of-flight (TOF) *mass analyzer* (p. 92)
tandem mass spectrometry (p. 93)
solid-phase method (p. 98)
x-ray crystallography (p. 98)
Fourier transform (p. 99)
electron-density map (p. 99)
nuclear magnetic resonance (NMR) spectroscopy (p. 101)
chemical shift (p. 101)

Problems

1. *Valuable reagents.* The following reagents are often used in protein chemistry:

CNBr	Trypsin	Ninhydrin
Urea	Performic acid	Phenyl isothiocyanate
Mercaptoethanol	6 N HCl	Chymotrypsin

Which one is the best suited for accomplishing each of the following tasks?

(a) Determination of the amino acid sequence of a small peptide.

(b) Reversible denaturation of a protein devoid of disulfide bonds. Which additional reagent would you need if disulfide bonds were present?

(c) Hydrolysis of peptide bonds on the carboxyl side of aromatic residues.

(d) Cleavage of peptide bonds on the carboxyl side of methionines.

(e) Hydrolysis of peptide bonds on the carboxyl side of lysine and arginine residues.

2. *Finding an end.* Anhydrous hydrazine ($H_2N—NH_2$) has been used to cleave peptide bonds in proteins. What are the reaction products? How might this technique be used to identify the carboxyl-terminal amino acid?

3. *Crafting a new breakpoint.* Ethyleneimine reacts with cysteine side chains in proteins to form *S*-aminoethyl derivatives. The peptide bonds on the carboxyl side of these modified cysteine residues are susceptible to hydrolysis by trypsin. Why?

4. *Spectrometry.* The absorbance A of a solution is defined as

$$A = \log_{10}(I_0/I)$$

in which I_0 is the incident-light intensity and I is the transmitted-light intensity. The absorbance is related to the molar absorption coefficient (extinction coefficient) ε (in $M^{-1}\ cm^{-1}$), concentration c (in M), and path length l (in cm) by

$$A = \varepsilon lc$$

The absorption coefficient of myoglobin at 580 nm is 15,000 $M^{-1}\ cm^{-1}$. What is the absorbance of a 1 mg ml^{-1} solution across a 1-cm path? What percentage of the incident light is transmitted by this solution?

5. *It's in the bag.* Suppose that you precipitate a protein with 1 M $(NH_4)_2SO_4$ and that you wish to reduce the concentration of the $(NH_4)_2SO_4$.You take 1 ml of your sample and dialyze it in 1000 ml of buffer. At the end of dialysis, what is the concentration of $(NH_4)_2SO_4$ in your sample? How could you further lower the $(NH_4)_2SO_4$ concentration?

6. *Too much or not enough.* Why do proteins precipitate at high salt concentrations? Although many proteins precipitate at high salt concentrations, some proteins require salt to dissolve in water. Explain why some proteins require salt to dissolve.

7. *A slow mover.* Tropomyosin, a 70-kd muscle protein, sediments more slowly than does hemoglobin (65 kd). Their sedimentation coefficients are 2.6S and 4.31S, respectively. Which structural feature of tropomyosin accounts for its slow sedimentation?

8. *Sedimenting spheres.* What is the dependence of the sedimentation coefficient s of a spherical protein on its mass? How much more rapidly does an 80-kd protein sediment than does a 40-kd protein?

9. *Frequently used in shampoos.* The detergent sodium dodecyl sulfate (SDS) denatures proteins. Suggest how SDS destroys protein structure.

10. *Size estimate.* The relative electrophoretic mobilities of a 30-kd protein and a 92-kd protein used as standards on an SDS–polyacrylamide gel are 0.80 and 0.41, respectively. What is the apparent mass of a protein having a mobility of 0.62 on this gel?

11. *Unexpected migration.* Some proteins migrate anomalously in SDS-PAGE gels. For instance, the molecular weight determined from an SDS-PAGE gel is sometimes very different from the molecular weight determined from the amino acid sequence. Suggest an explanation for this discrepancy.

12. *Sorting cells.* Fluorescence-activated cell sorting (FACS) is a powerful technique for separating cells according to their content of particular molecules. For example, a fluorescence-labeled antibody specific for a cell-surface protein can be used to detect cells containing such a molecule. Suppose that you want to isolate cells that possess a receptor enabling them to detect bacterial degradation products. However, you do not yet have an antibody directed against this receptor. Which fluorescence-labeled molecule would you prepare to identify such cells?

13. *Column choice.* (a) The octapeptide AVGWRVKS was digested with the enzyme trypsin. Which method would be most appropriate for separating the products: ion-exchange or gel-filtration chromatography? Explain. (b) Suppose that the peptide was digested with chymotrypsin. What would be the optimal separation technique? Explain.

14. *Power(ful) tools.* Monoclonal antibodies can be conjugated to an insoluble support by chemical methods. Explain how these antibody-bound beads can be exploited for protein purification.

15. *Assay development.* You wish to isolate an enzyme from its native source and need a method for measuring its activity throughout the purification. However, neither the substrate nor the product of the enzyme-catalyzed reaction can be detected by spectroscopy. You discover that the product of the reaction is highly antigenic when injected into mice. Propose a strategy to develop a suitable assay for this enzyme.

16. *Making more enzyme?* In the course of purifying an enzyme, a researcher performs a purification step that results in an *increase* in the total activity to a value greater than that present in the original crude extract. Explain how the amount of total activity might increase.

17. *Divide and conquer.* The determination of the mass of a protein by mass spectrometry often does not allow its unique identification among possible proteins within a complete proteome, but determination of the masses of all fragments produced by digestion with trypsin almost always allows unique identification. Explain.

18. *Know your limits.* Which two amino acids are indistinguishable in peptide sequencing by the tandem mass spectrometry method described in this chapter and why?

19. *Protein purification problem.* Complete the following table.

Purification procedure	Total protein (mg)	Total activity (units)	Specific activity (units mg^{-1})	Purification level	Yield (%)
Crude extract	20,000	4,000,000		1	100
$(NH_4)_2SO_4$ precipitation	5,000	3,000,000			
DEAE-cellulose chromatography	1,500	1,000,000			
Gel-filtration chromatography	500	750,000			
Affinity chromatography	45	675,000			

20. *The challenge of flexibility.* Structures of proteins comprising domains separated by flexible linker regions can be quite difficult to solve by x-ray crystallographic methods. Why might this be the case? What are possible experimental approaches to circumvent this barrier?

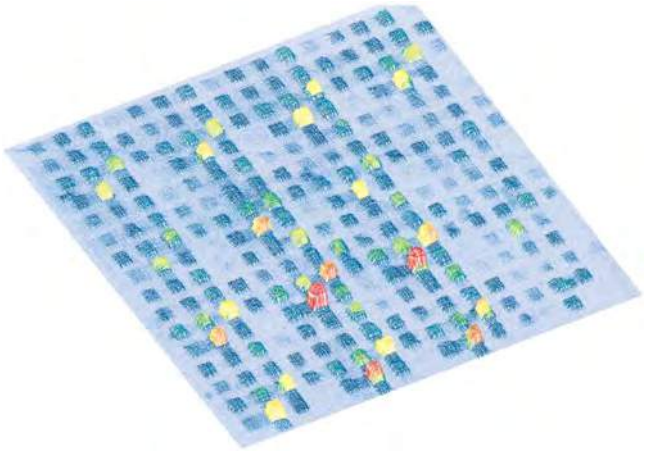
Chapter Integration Problems

21. *Quaternary structure.* A protein was purified to homogeneity. Determination of the mass by gel-filtration chromatography yields 60 kd. Chromatography in the presence of 6 M urea yields a 30-kd species. When the chromatography is repeated in the presence of 6 M urea and 10 mM β-mercaptoethanol, a single molecular species of 15 kd results. Describe the structure of the molecule.

22. *Helix–coil transitions.* (a) NMR measurements have shown that poly-L-lysine is a random coil at pH 7 but becomes α helical as the pH is raised above 10. Account for this pH-dependent conformational transition. (b) Predict the pH dependence of the helix–coil transition of poly-L-glutamate.

23. *Peptide mass determination.* You have isolated a protein from the bacterium *E. coli* and seek to confirm its identity by trypsin digestion and mass spectrometry. Determination of the masses of several peptide fragments has enabled you to deduce the identity of the protein. However, there is a discrepancy with one of the peptide fragments, which you believe should have the sequence MLNSFK and an $(M + H)^+$ value of 739.38. In your experiments, you repeatedly obtain an $(M + H)^+$ value of 767.38. What is the cause of this discrepancy and what does it tell you about the region of the protein from which this peptide is derived?

24. *Peptides on a chip.* Large numbers of different peptides can be synthesized in a small area on a solid support. This high-density array can then be probed with a fluorescence-labeled protein to find out which peptides are recognized. The binding of an antibody to an array of 1024 different peptides occupying a total area the size of a thumbnail is shown in the adjoining illustration. How would you synthesize such a peptide array? (Hint: Use light instead of acid to deprotect the terminal amino group in each round of synthesis.)



Fluorescence scan of an array of 1024 peptides in a 1.6-cm^2 area. Each synthesis site is a 400-μm square. A fluorescently labeled monoclonal antibody was added to the array to identify peptides that are recognized. The height and color of each square denote the fluorescence intensity. [After S. P. A. Fodor et al., *Science* 251(1991):767.]

25. *Exchange rate.* The amide hydrogen atoms of peptide bonds within proteins can exchange with protons in the solvent. In general, amide hydrogen atoms in buried regions of proteins and protein complexes exchange more slowly than those on the solvent-accessible surface do. Determination of these rates can be used to explore the protein-folding reaction, probe the tertiary structure of proteins, and identify the regions of protein–protein interfaces. These exchange reactions can be followed by studying the behavior of the protein in solvent that has been labeled with deuterium (^{2}H), a stable isotope of hydrogen. What two methods described in this chapter could be readily applied to the study of hydrogen–deuterium exchange rates in proteins?

Data Interpretation Problems

26. *Protein sequencing 1.* Determine the sequence of hexapeptide on the basis of the following data. Note: When the sequence is not known, a comma separates the amino acids (see Table 3.3).

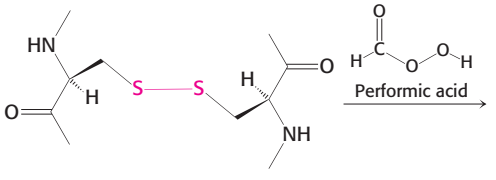
Amino acid composition: (2R,A,S,V,Y)
N-terminal analysis of the hexapeptide: A
Trypsin digestion: (R,A,V) and (R,S,Y)
Carboxypeptidase digestion: No digestion.
Chymotrypsin digestion: (A,R,V,Y) and (R,S)

27. *Protein sequencing 2.* Determine the sequence of a peptide consisting of 14 amino acids on the basis of the following data.

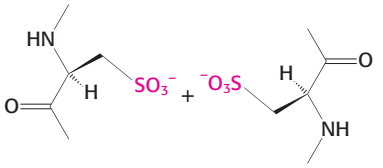
Amino acid composition: (4S,2L,F,G,I,K,M,T,W,Y)
N-terminal analysis: S
Carboxypeptidase digestion: L
Trypsin digestion: (3S,2L,F,I,M,T,W) (G,K,S,Y)
Chymotrypsin digestion: (F,I,S) (G,K,L) (L,S) (M,T) (S,W) (S,Y)
N-terminal analysis of (F,I,S) peptide: S
Cyanogen bromide treatment: (2S,F,G,I,K,L,M*,T,Y) (2S,L,W)
M*, methionine detected as homoserine

28. *Applications of two-dimensional electrophoresis.* Performic acid cleaves the disulfide linkage of cystine and converts the sulfhydryl groups into cysteic acid residues, which are then no longer capable of disulfide-bond formation.

Consider the following experiment: You suspect that a protein containing three cysteine residues has a single disulfide bond. You digest the protein with trypsin and subject the mixture to electrophoresis along one end of a sheet of paper. After treating the paper with performic acid, you subject the sheet to electrophoresis in the perpendicular direction and stain it with ninhydrin. How would the paper appear if the protein did not contain any disulfide bonds? If the protein contained a single disulfide bond? Propose an experiment to identify which cysteine residues form the disulfide bond.

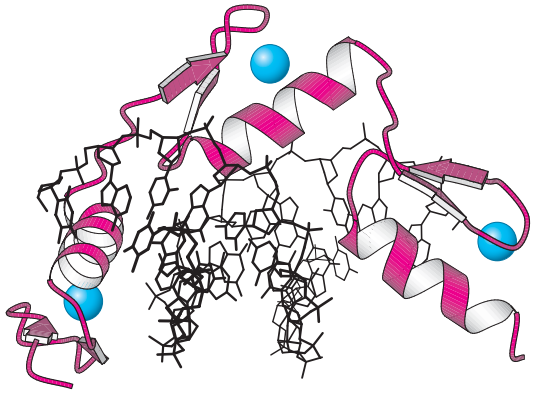


Cystine



Cysteic acid

DNA, RNA, and the Flow of Genetic Information



Having genes in common accounts for the resemblance of a mother to her daughters. Genes must be expressed to exert an effect, and proteins regulate such expression. One such regulatory protein, a zinc-finger protein (zinc ion is blue, protein is red), is shown bound to a control region of DNA (black). [(Left) Barnaby Hall/Photonica. (Right) Drawn from 1AAY.pdb.]

DNA and RNA are long linear polymers, called nucleic acids, that carry information in a form that can be passed from one generation to the next. These macromolecules consist of a large number of linked nucleotides, each composed of a sugar, a phosphate, and a base. Sugars linked by phosphates form a common backbone that plays a structural role, whereas *the sequence of bases along a nucleic acid chain carries genetic information.* The DNA molecule has the form of a *double helix*, a helical structure consisting of two complementary nucleic acid strands. *Each strand serves as the template for the other in DNA replication.* The genes of all cells and many viruses are made of DNA.

Genes specify the kinds of proteins that are made by cells, but DNA is not the direct template for protein synthesis. Rather, a DNA strand is copied into a class of RNA molecules called *messenger RNA* (mRNA), the information-carrying intermediates in protein synthesis. This process of *transcription* is followed by *translation*, the synthesis of proteins according to instructions given by mRNA templates. Thus, the flow of genetic information, or *gene expression*, in normal cells is

$$\text{DNA} \xrightarrow{\text{Transcription}} \text{RNA} \xrightarrow{\text{Translation}} \text{Protein}$$

This flow of information depends on the genetic code, which defines the relation between the sequence of bases in DNA (or its mRNA transcript) and the sequence of amino acids in a protein. The code is nearly the same in all organisms: a sequence of three bases, called a *codon*, specifies an amino

OUTLINE

acid. There is another step in the expression of most eukaryotic genes, which are mosaics of nucleic acid sequences called *introns* and *exons*. Both are transcribed, but before translation takes place, introns are cut out of newly synthesized RNA molecules, leaving mature RNA molecules with continuous exons. The existence of introns and exons has crucial implications for the evolution of proteins.

4.1 A Nucleic Acid Consists of Four Kinds of Bases Linked to a Sugar–Phosphate Backbone

The nucleic acids DNA and RNA are well suited to function as the carriers of genetic information by virtue of their covalent structures. These macromolecules are *linear polymers* built up from similar units connected end to end (Figure 4.1). Each monomer unit within the polymer is a *nucleotide*. A single nucleotide unit consists of three components: a sugar, a phosphate, and one of four bases. *The sequence of bases in the polymer uniquely characterizes a nucleic acid and constitutes a form of linear information*—information analogous to the letters that spell a person's name.

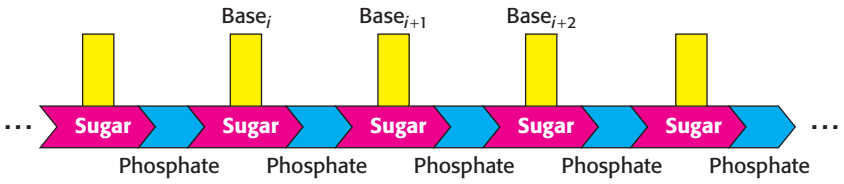


Figure 4.1 Polymeric structure of nucleic acids.

RNA and DNA differ in the sugar component and one of the bases

The sugar in *deoxyribonucleic acid* (DNA) is *deoxyribose*. The prefix deoxy indicates that the 2′-carbon atom of the sugar lacks the oxygen atom that is linked to the 2′-carbon atom of *ribose,* as shown in Figure 4.2. Note that sugar carbons are numbered with primes to differentiate them from atoms in the bases. The sugars in both nucleic acids are linked to one another by phosphodiester bridges. Specifically, the 3′-hydroxyl (3′-OH) group of the sugar moiety of one nucleotide is esterified to a phosphate group, which is, in turn, joined to the 5′-hydroxyl group of the adjacent sugar. The chain of sugars linked by phosphodiester bridges is referred to as the *backbone* of the nucleic acid (Figure 4.3). Whereas the backbone is constant in a nucleic acid, the bases vary from one monomer to the next. Two of the bases of

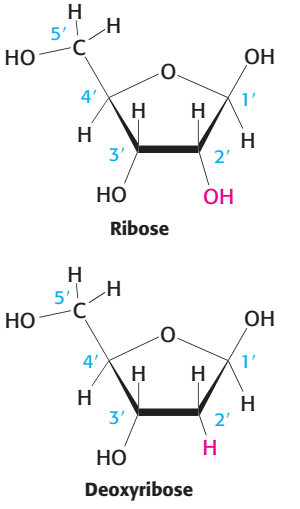


Figure 4.2 Ribose and deoxyribose. Atoms in sugar units are numbered with primes to distinguish them from atoms in bases (see Figure 4.4).

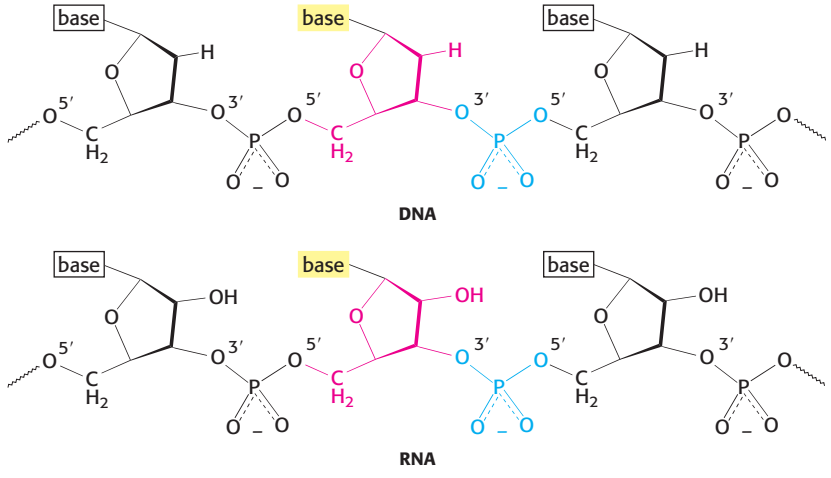


Figure 4.3 Backbones of DNA and RNA. The backbones of these nucleic acids are formed by 3′-to-5′ phosphodiester linkages. A sugar unit is highlighted in red and a phosphate group in blue.

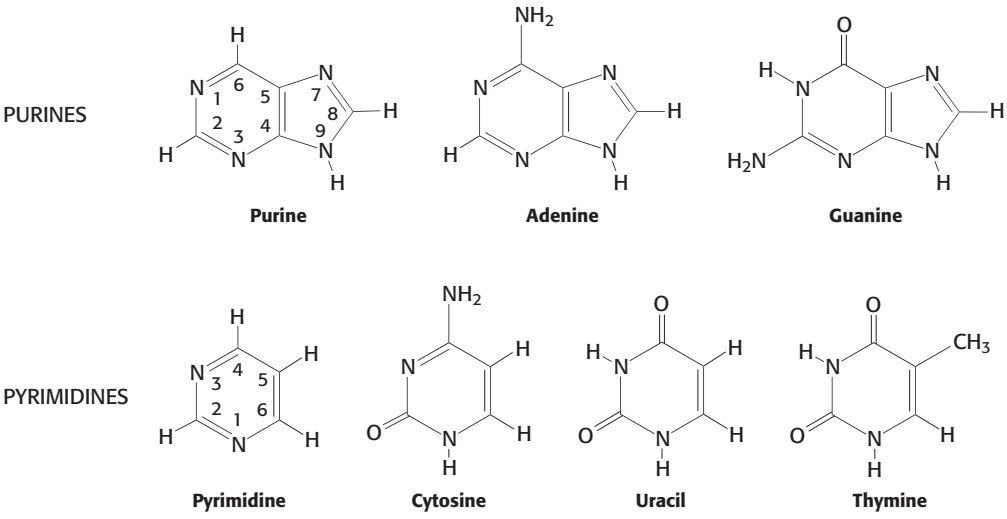


Figure 4.4 Purines and pyrimidines. Atoms within bases are numbered without primes. Uracil is present in RNA instead of thymine.

DNA are derivatives of *purine*—adenine (A) and guanine (G)—and two of *pyrimidine*—cytosine (C) and thymine (T), as shown in Figure 4.4.

Ribonucleic acid (RNA), like DNA, is a long unbranched polymer consisting of nucleotides joined by 3′-to-5′ phosphodiester linkages (see Figure 4.3). The covalent structure of RNA differs from that of DNA in two respects. First, the sugar units in RNA are riboses rather than deoxyriboses. Ribose contains a 2′-hydroxyl group not present in deoxyribose. Second, one of the four major bases in RNA is uracil (U) instead of thymine (T).

Note that each phosphodiester bridge has a negative charge. This negative charge repels nucleophilic species such as hydroxide ions; consequently, phosphodiester linkages are much less susceptible to hydrolytic attack than are other esters such as carboxylic acid esters. This resistance is crucial for maintaining the integrity of information stored in nucleic acids. The absence of the 2′-hydroxyl group in DNA further increases its resistance to hydrolysis. The greater stability of DNA probably accounts for its use rather than RNA as the hereditary material in all modern cells and in many viruses.

Nucleotides are the monomeric units of nucleic acids

The building blocks of nucleic acids and the precursors of these building blocks play many other roles throughout the cell—for instance, as energy currency and as molecular signals. Consequently, it is important to be familiar with the nomenclature of nucleotides and their precursors. A unit consisting of a base bonded to a sugar is referred to as a *nucleoside*. The four nucleoside units in RNA are called *adenosine, guanosine, cytidine*, and *uridine*, whereas those in DNA are called *deoxyadenosine, deoxyguanosine, deoxycytidine*, and *thymidine*. In each case, N-9 of a purine or N-1 of a pyrimidine is attached to C-1′ of the sugar by an *N*-glycosidic linkage (Figure 4.5). The base lies above the plane of sugar when the structure is written in the standard orientation; that is, the configuration of the *N*-glycosidic linkage is β (Section 11.1).

A *nucleotide* is a nucleoside joined to one or more phosphoryl groups by an ester linkage. *Nucleotide triphosphates, nucleosides joined to three phosphoryl groups, are the monomers—the building blocks—that are linked to form RNA and DNA.* The four nucleotide units that link to form DNA are nucleotide monophosphates called *deoxyadenylate, deoxyguanylate, deoxycytidylate,* and *thymidylate*. Note that thymidylate contains deoxyribose; by convention, the prefix deoxy is not added because thymine-containing

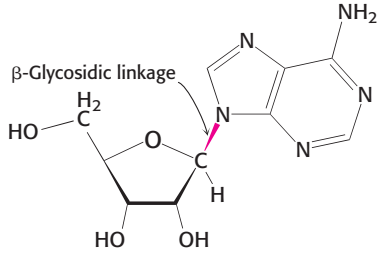


Figure 4.5 β-Glycosidic linkage in a nucleoside.

nucleotides are only rarely found in RNA. Similarly, the most common nucleotides that link to form RNA are nucleotide monophosphates adenylate, guanylate, cytidylate and uridylate.

Another means of denoting a nucleotide is the base name with the suffix "ate". This nomenclature does not describe the number of phosphoryl groups or the site of attachment to carbon of the ribose. A more precise nomenclature is also commonly used. A compound formed by the attachment of a phosphoryl group to C-5′ of a nucleoside sugar (the most common site of phosphate esterification) is called a *nucleoside 5′-phosphate* or a *5′-nucleotide*. In this naming system for nucleotides, the number of phosphoryl groups and the attachment site are designated. Look, for example at *adenosine 5′-triphosphate* (ATP; Figure 4.6). This nucleotide is tremendously important because, in addition to being a building block for RNA, it is the most commonly used energy currency. The energy released from cleavage of the triphosphate group is used to power many cellular processes (Chapter 15). Another nucleotide is deoxyguanosine 3′-monophosphate (3′-dGMP; see Figure 4.6). This nucleotide differs from ATP in that it contains guanine rather than adenine, contains deoxyribose rather than ribose (indicated by the prefix "d"), contains one rather than three phosphoryl groups, and has the phosphoryl group esterified to the hydroxyl group in the 3′ rather than the 5′ position.

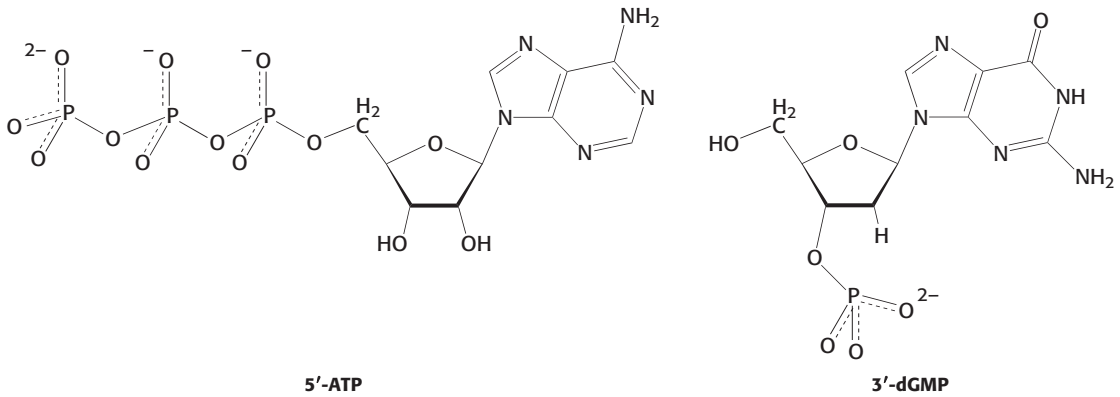


Figure 4.6 Nucleotides adenosine 5′-triphosphate (5′-ATP) and deoxyguanosine 3′-monophosphate (3′-dGMP).

Scientific communication frequently requires the sequence of a nucleic acid—in some cases, a sequence thousands of nucleotides in length—to be written like that on page 17. Rather than writing the cumbersome chemical structures, scientists have adopted the use of abbreviations. The abbreviated notations pApCpG or ACG denote a trinucleotide of DNA consisting of the building blocks deoxyadenylate monophosphate, deoxycytidylate monophosphate, and deoxyguanylate monophosphate linked by a phosphodiester bridge, where "p" denotes a phosphoryl group (Figure 4.7). The 5′ end will often have a phosphoryl group attached to the 5′-OH group. Note that, like a polypeptide (Section 2.2), *a DNA chain has directionality,* commonly called polarity. One end of the chain has a free 5′-OH group (or a 5′-OH group attached to a phosphoryl group) and the other end has a free 3′-OH group, neither of which is linked to another nucleotide. By convention, *the base sequence is written in the 5′-to-3′ direction*. Thus, ACG indicates that the unlinked 5′-OH group is on deoxyadenylate, whereas the unlinked 3′-OH group is on deoxyguanylate. Because of this polarity, ACG and GCA correspond to different compounds.

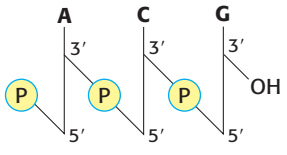


Figure 4.7 Structure of a DNA chain. The chain has a 5′ end, which is usually attached to a phosphoryl group, and a 3′ end, which is usually a free hydroxyl group.

DNA molecules are very long

A striking characteristic of naturally occurring DNA molecules is their length. A DNA molecule must comprise many nucleotides to carry the genetic information necessary for even the simplest organisms. For example, the DNA of a virus such as polyoma, which can cause cancer in certain organisms, consists of two intertwined strands of DNA, each 5100 nucleotides in length. The *E. coli* genome is a single DNA molecule consisting of two chains of 4.6 million nucleotides each (Figure 4.8).

The DNA molecules of higher organisms can be much larger. The human genome comprises approximately 3 billion nucleotides in each chain of DNA, divided among 24 distinct molecules of DNA called chromosomes (22 autosomal chromosomes plus the X and Y sex chromosomes) of different sizes. One of the largest known DNA molecules is found in the Indian muntjac, an Asiatic deer; its genome is nearly as large as the human genome but is distributed on only 3 chromosomes (Figure 4.9). The largest of these chromosomes has two chains of more than 1 billion nucleotides each. If such a DNA molecule could be fully extended, it would stretch more than 1 foot in length. Some plants contain even larger DNA molecules.

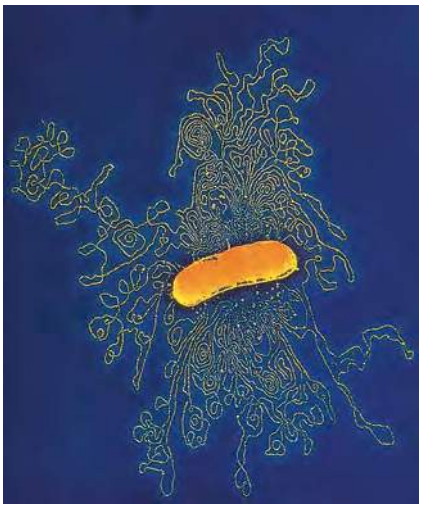


Figure 4.8 Electron micrograph of part of the *E. coli* genome. [Dr. Gopal Murti/ Science Photo Library/Photo Researchers.]

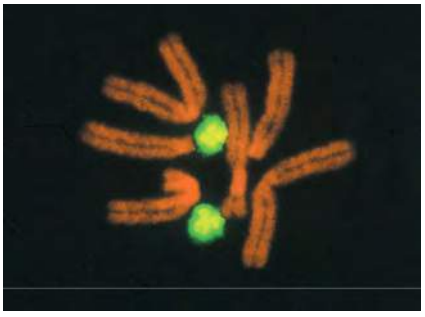


Figure 4.9 The Indian muntjac and its chromosomes. Cells from a female Indian muntjac (right) contain three pairs of very large chromosomes (stained orange). The cell shown is a hybrid containing a pair of human chromosomes (stained green) for comparison. [(Left) M. Birkhead, OSF/Animals Animals. (Right) J.–Y. Lee, M. Koi, E. J. Stanbridge, M. Oshimura, A. T. Kumamoto, and A. P. Feinberg. *Nat. Genet.* 7:30, 1994.]

4.2 A Pair of Nucleic Acid Chains with Complementary Sequences Can Form a Double-Helical Structure

As discussed in Chapter 1, the covalent structure of nucleic acids accounts for their ability to carry information in the form of a sequence of bases along a nucleic acid chain. The bases on the two separate nucleic acid strands form *specific base pairs* in such a way that a helical structure is formed. The double-helical structure of DNA facilitates the *replication* of the genetic material—that is, the generation of two copies of a nucleic acid from one.

The double helix is stabilized by hydrogen bonds and van der Waals interactions

The ability of nucleic acids to form specific base pairs was discovered in the course of studies directed at determining the three-dimensional structure of DNA. Maurice Wilkins and Rosalind Franklin obtained x-ray diffraction photographs of fibers of DNA (Figure 4.10). The characteristics of these diffraction patterns indicated that DNA is formed of two chains that wind in a regular helical structure. From these data and others, James Watson and

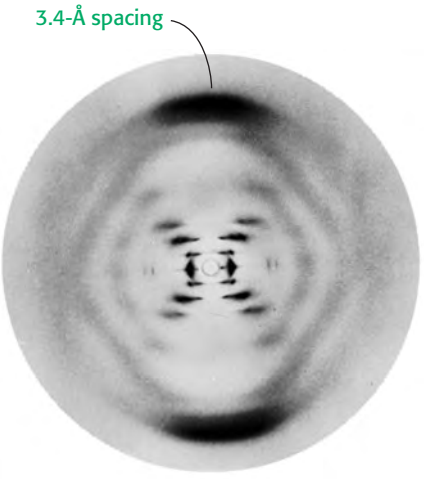


Figure 4.10 X-ray diffraction photograph of a hydrated DNA fiber. When crystals of a biomolecule are irradiated with x-rays, the x-rays are diffracted and these diffracted x-rays are seen as a series of spots, called reflections, on a screen behind the crystal. The structure of the molecule can be determined by the pattern of the reflections (Section 3.6). In regard to DNA crystals, the central cross is diagnostic of a helical structure. The strong arcs on the meridian arise from the stack of nucleotide bases, which are 3.4 Å apart. [Courtesy of Dr. Maurice Wilkins.]

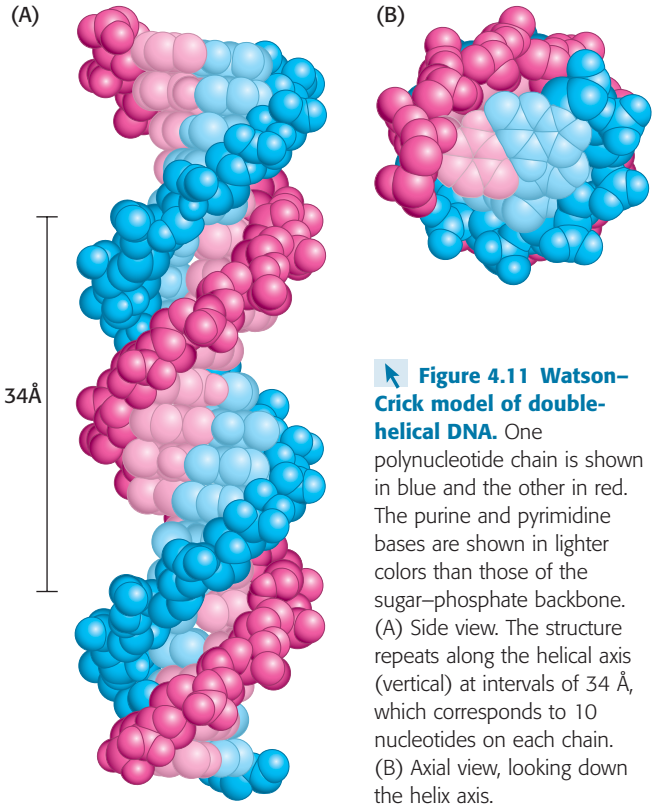


Figure 4.11 Watson–Crick model of double-helical DNA. One polynucleotide chain is shown in blue and the other in red. The purine and pyrimidine bases are shown in lighter colors than those of the sugar–phosphate backbone. (A) Side view. The structure repeats along the helical axis (vertical) at intervals of 34 Å, which corresponds to 10 nucleotides on each chain. (B) Axial view, looking down the helix axis.

Francis Crick deduced a structural model for DNA that accounted for the diffraction pattern and was the source of some remarkable insights into the functional properties of nucleic acids (Figure 4.11).

The features of the Watson–Crick model of DNA deduced from the diffraction patterns are:

1. Two helical polynucleotide chains are coiled around a common axis with a right-handed screw sense (p. 39). The chains are antiparallel, meaning that they have opposite polarity.

2. The sugar–phosphate backbones are on the outside and the purine and pyrimidine bases lie on the inside of the helix.

3. The bases are nearly perpendicular to the helix axis, and adjacent bases are separated by 3.4 Å. This spacing is readily apparent in the DNA diffraction pattern (see Figure 4.10). The helical structure repeats every 34 Å, and so there are 10 bases (= 34 Å per repeat/3.4 Å per base) per turn of helix. Each base is rotated 36 degrees from the one below it. (360 degrees per full turn/10 bases per turn).

4. The diameter of the helix is 20 Å.

How is such a regular structure able to accommodate an arbitrary sequence of bases, given the different sizes and shapes of the purines and pyrimidines? In attempting to answer this question, Watson and Crick discovered that guanine can be paired with cytosine and adenine with thymine to form base pairs that have essentially the same shape (Figure 4.12). These base pairs are held together by specific hydrogen bonds, which, although weak (4–21 kJ mol^{-1}, or 1–5 kcal mol^{-1}), stabilize the helix because of their large numbers in a DNA molecule. These *base-pairing rules* account for the observation, originally made by Erwin Chargaff in 1950, that the ratios of adenine to thymine and of guanine to cytosine are nearly the same in all species studied, whereas the adenine-to-guanine ratio varies considerably (Table 4.1).

Inside the helix, the bases are essentially stacked one on top of another (Figure 4.13). The stacking of base pairs contributes to the stability of the double helix in two ways. First, the double helix is stabilized by the hydrophobic effect (p. 9). The hydrophobic bases cluster in the interior of the helix away from the surrounding water, whereas the more polar surfaces are exposed to water. This arrangement is reminiscent of protein folding, where hydrophobic amino acids are in the protein's interior and the hydrophilic amino acids are on the exterior (Section 2.4). The hydrophobic effect stacks

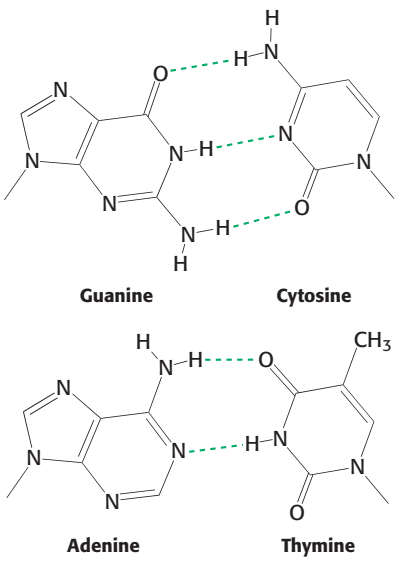


Figure 4.12 Structures of the base pairs proposed by Watson and Crick.

Table 4.1 Base compositions experimentally determined for a variety of organisms

Organism	A : T	G : C	A : G
Human being	1.00	1.00	1.56
Salmon	1.02	1.02	1.43
Wheat	1.00	0.97	1.22
Yeast	1.03	1.02	1.67
Escherichia coli	1.09	0.99	1.05
Serratia marcescens	0.95	0.86	0.70

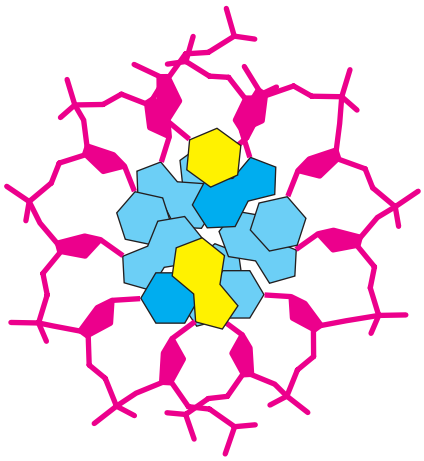


Figure 4.13 Axial view of DNA. Base pairs are stacked nearly one on top of another in the double helix.

the bases on top of one another. The stacked base pairs attract one another through van der Waals forces (p. 8), appropriately referred to as *stacking forces,* further contributing to stabilization of the helix. The energy associated with a single van der Waals interaction is quite small, typically from 2 to 4 kJ mol^{-1} (0.5–1.0 kcal mol^{-1}). In the double helix, however, a large number of atoms are in van der Waals contact, and the net effect, summed over these atom pairs, is substantial. In addition, base stacking in DNA is favored by the conformations of the somewhat rigid five-membered rings of the backbone sugars.

DNA can assume a variety of structural forms

Watson and Crick based their model (known as the *B-DNA helix*) on x-ray diffraction patterns of highly hydrated DNA fibers, which provided information about properties of the double helix that are averaged over its constituent residues. Under physiological conditions, most DNA is in the B form. X-ray diffraction studies of less-hydrated DNA fibers revealed a different form called *A-DNA*. Like B-DNA, A-DNA is a right-handed double helix made up of antiparallel strands held together by Watson–Crick base-pairing. The A-form helix is wider and shorter than the B-form helix, and its base pairs are tilted rather than perpendicular to the helix axis (Figure 4.14).

If the A-form helix were simply a property of dehydrated DNA, it would be of little significance. However, double-stranded regions of RNA and at least some RNA–DNA hybrids adopt a double-helical form very similar to that of A-DNA. What is the biochemical basis for differences between the two forms of DNA? Many of the structural differences between B-DNA and A-DNA arise from different puckerings of their ribose units (Figure 4.15). In A-DNA, C-3′ lies out of the plane (a conformation referred to as C-3′ endo) formed by the other four atoms of the ring; in B-DNA, C-2′ lies out of the plane (a conformation called C-2′

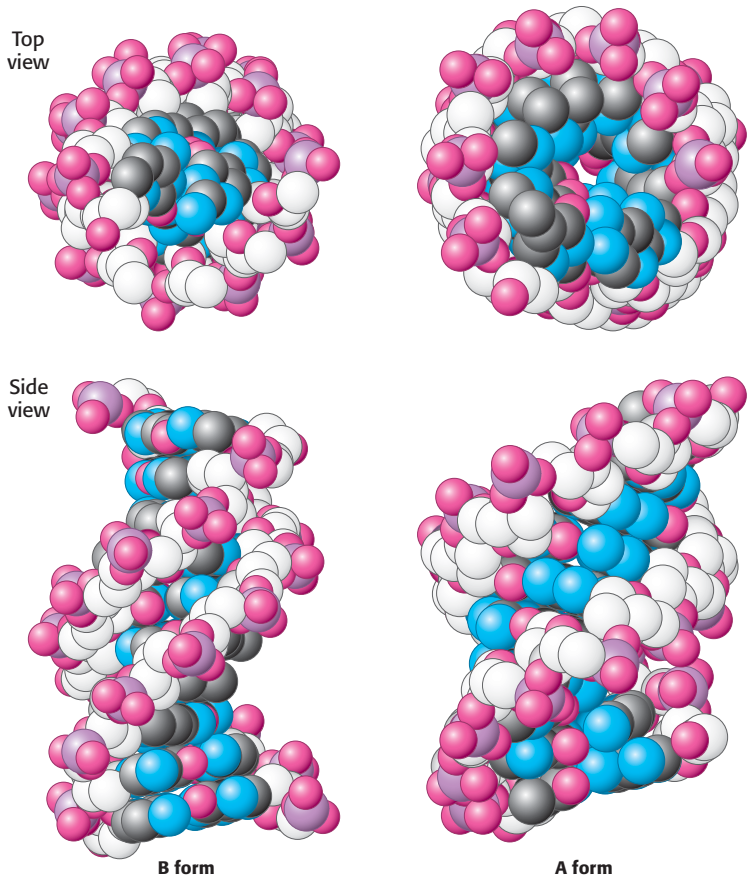
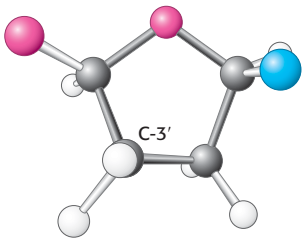
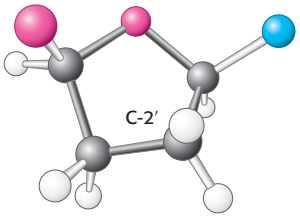


Figure 4.14 B-form and A-form DNA. Space-filling models of 10 base pairs of B-form and A-form DNA depict their right-handed helical structures. *Notice* that the B-form helix is longer and narrower than the A-form helix. The carbon atoms of the backbone are shown in white. [Drawn from 1BNA.pdb and 1DNZ.pdb.]



C-3′ endo (A form)



C-2′ endo (B form)

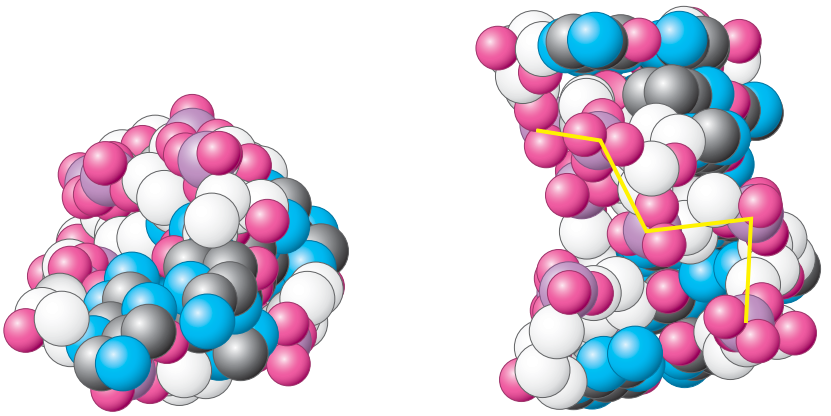
Figure 4.15 Sugar pucker. In A-form DNA, the C-3′ carbon atom lies above the approximate plane defined by the four other sugar nonhydrogen atoms (called C-3′ endo). In B-form DNA, each deoxyribose is in a C-2′-endo conformation, in which C-2′ lies out of the plane.

endo). The C-3′-endo puckering in A-DNA leads to an 11-degree tilting of the base pairs away from perpendicular to the helix. RNA helices are further induced to take the A-DNA form because of steric hindrance from the 2′-hydroxyl group: the 2′-oxygen atom would be too close to three atoms of the adjoining phosphoryl group and to one atom in the next base. In an A-form helix, in contrast, the 2′-oxygen atom projects outward, away from other atoms. The phosphoryl and other groups in the A-form helix bind fewer H_2O molecules than do those in B-DNA. Hence, dehydration favors the A form.

Z-DNA is a left-handed double helix in which backbone phosphates zigzag

Alexander Rich and his associates discovered a third type of DNA helix when they solved the structure of CGCGCG. They found that this hexanucleotide forms a duplex of antiparallel strands held together by Watson–Crick base-pairing, as expected. What was surprising, however, was that this double helix was *left-handed*, in contrast with the *right-handed* screw sense of the A-DNA and B-DNA helices. Furthermore, the phosphates in the backbone *zigzagged;* hence, they called this new form *Z-DNA* (Figure 4.16).

The existence of Z-DNA shows that DNA is a flexible, dynamic molecule. Although the biological role of Z-DNA is still under investigation, Z-DNA-binding proteins required for viral pathogenesis have been isolated from poxviruses, including variola, the agent of smallpox. The properties of A-, B-, and Z-DNA are compared in Table 4.2.



Top view Side view

Figure 4.16 Z-DNA. DNA oligomers such as CGCGCG adopt an alternative conformation under some conditions. This conformation is called Z-DNA because the phosphoryl groups zigzag along the backbone. [Drawn from 131D.pdb.]

Table 4.2 Comparison of A-, B-, and Z-DNA

	Helix type		
	A	B	Z
Shape	Broadest	Intermediate	Narrowest
Rise per base pair	2.3 Å	3.4 Å	3.8 Å
Helix diameter	25.5 Å	23.7 Å	18.4 Å
Screw sense	Right-handed	Right-handed	Left-handed
Glycosidic bond*	*anti*	*anti*	Alternating *anti* and *syn*
Base pairs per turn of helix	11	10.4	12
Pitch per turn of helix	25.3 Å	35.4 Å	45.6 Å
Tilt of base pairs from perpendicular to helix axis	19 degrees	1 degree	9 degrees

**Syn* and *anti* refer to the orientation of the *N*-glycosidic bond between the base and deoxyribose. In the *anti* orientation, the base extends away from the deoxyribose. In the *syn* orientation, the base is above the deoxyribose. Pyrimidine can be in *anti* orientations only, whereas purines can be *anti* or *syn*.

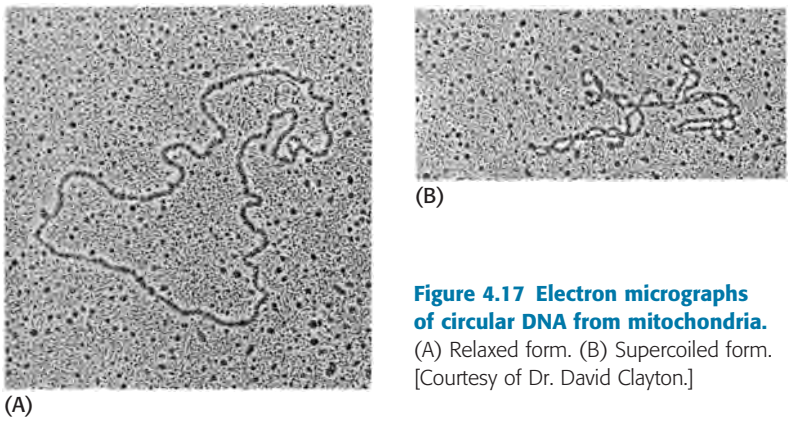


Figure 4.17 Electron micrographs of circular DNA from mitochondria. (A) Relaxed form. (B) Supercoiled form. [Courtesy of Dr. David Clayton.]

Some DNA molecules are circular and supercoiled

The DNA molecules in human chromosomes are linear. However, electron microscopic and other studies have shown that intact DNA molecules from bacteria and archaea are circular (Figure 4.17A). The term *circular* refers to the continuity of the DNA chains, not to their geometric form. DNA molecules inside cells necessarily have a very compact shape. Note that the *E. coli* chromosome, fully extended, would be about 1000 times as long as the greatest diameter of the bacterium.

A closed DNA molecule has a property unique to circular DNA. The axis of the double helix can itself be twisted or supercoiled into a *superhelix* (Figure 4.17B). A circular DNA molecule without any superhelical turns is known as a *relaxed molecule*. Supercoiling is biologically important for two reasons. First, *a supercoiled DNA molecule is more compact than its relaxed counterpart*. Second, *supercoiling may hinder or favor the capacity of the double helix to unwind and thereby affect the interactions between DNA and other molecules.* These topological features of DNA will be considered further in Chapter 28.

Single-stranded nucleic acids can adopt elaborate structures

Single-stranded nucleic acids often fold back on themselves to form well-defined structures. Such structures are especially prominent in RNA and RNA-containing complexes such as the ribosome—a large complex of RNAs and proteins on which proteins are synthesized.

The simplest and most-common structural motif formed is a *stem-loop,* created when two complementary sequences within a single strand come together to form double-helical structures (Figure 4.18). In many cases, these double helices are made up entirely of Watson–Crick base pairs. In other cases, however, the structures include mismatched base pairs or unmatched bases that bulge out from the helix. Such mismatches destabilize the local

Figure 4.18 Stem-loop structures. Stem-loop structures can be formed from single-stranded DNA and RNA molecules.

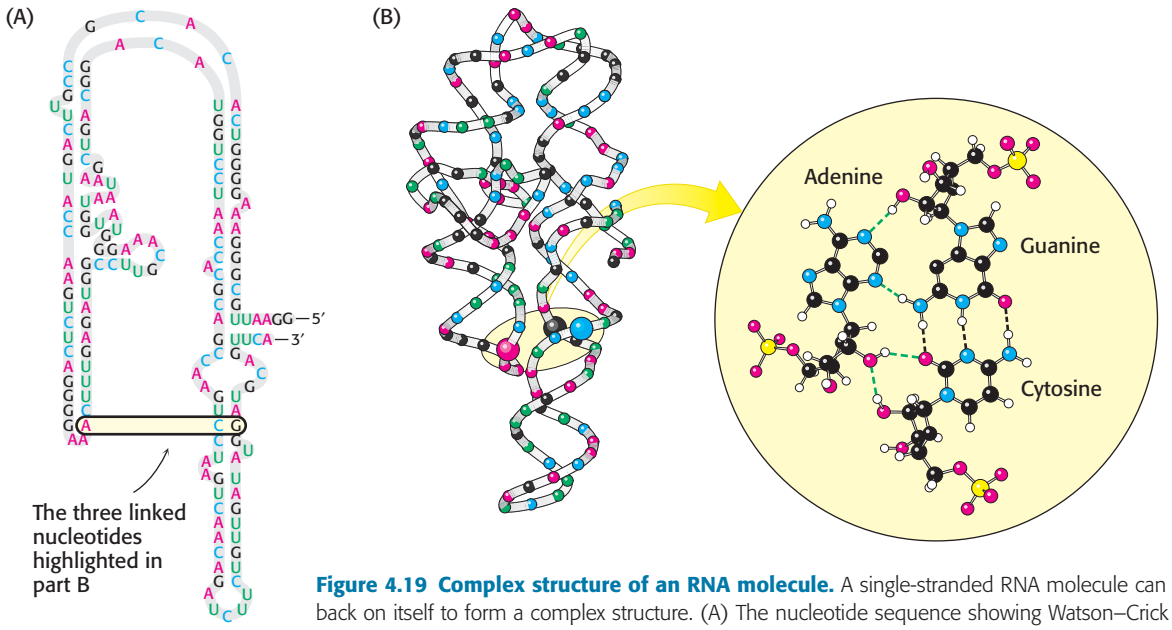


Figure 4.19 Complex structure of an RNA molecule. A single-stranded RNA molecule can fold back on itself to form a complex structure. (A) The nucleotide sequence showing Watson–Crick base pairs and other nonstandard base pairings in stem-loop structures. (B) The three-dimensional structure and one important long-range interaction between three bases. In the three-dimensional structure at the left, cytidine nucleotides are shown in blue, adenosine in red, guanosine in black, and uridine in green. In the detailed projection, hydrogen bonds within the Watson–Crick base pair are shown as dashed black lines; additional hydrogen bonds are shown as dashed green lines.

structure but introduce deviations from the standard double-helical structure that can be important for higher-order folding and for function (Figure 4.19).

Single-stranded nucleic acids can adopt structures that are more complex than simple stem-loops through the interaction of more widely separated bases. Often, three or more bases interact to stabilize these structures. In such cases, hydrogen-bond donors and acceptors that do not participate in Watson–Crick base pairs participate in hydrogen bonds to form nonstandard pairings. Metal ions such as magnesium ion (Mg^{2+}) often assist in the stabilization of these more elaborate structures. These complex structures allow RNA to perform a host of functions that the double-stranded DNA molecule cannot. Indeed, the complexity of some RNA molecules rivals that of proteins, and these RNA molecules perform a number of functions that had formerly been thought the private domain of proteins.

4.3 The Double Helix Facilitates the Accurate Transmission of Hereditary Information

The double-helical model of DNA and the presence of specific base pairs immediately suggested how the genetic material might replicate. *The sequence of bases of one strand of the double helix precisely determines the sequence of the other strand:* a guanine base on one strand is always paired with a cytosine base on the other strand, and so on. Thus, separation of a double helix into its two component chains would yield two single-stranded templates onto which new double helices could be constructed, each of which would have the same sequence of bases as the parent double helix. Consequently, as DNA is replicated, one of the chains of each daughter DNA molecule is newly synthesized, whereas the other is passed unchanged from the parent DNA molecule. This distribution of parental atoms is achieved by *semiconservative replication*.

Differences in DNA density established the validity of the semiconservative-replication hypothesis

Matthew Meselson and Franklin Stahl carried out a critical test of this hypothesis in 1958. They labeled the parent DNA with ^{15}N, a heavy isotope of nitrogen, to make it denser than ordinary DNA. The labeled DNA was generated by growing *E. coli* for many generations in a medium that contained $^{15}NH_4Cl$ as the sole nitrogen source. After the incorporation of heavy nitrogen was complete, the bacteria were abruptly transferred to a medium that contained ^{14}N, the ordinary isotope of nitrogen. The question asked was: What is the distribution of ^{14}N and ^{15}N in the DNA molecules after successive rounds of replication?

The distribution of ^{14}N and ^{15}N was revealed by the technique of *density-gradient equilibrium sedimentation*. A small amount of DNA was dissolved in a concentrated solution of cesium chloride having a density close to that of the DNA (1.7 g cm^{-3}). This solution was centrifuged until it was nearly at equilibrium. At that point, the opposing processes of sedimentation and diffusion created a gradient in the concentration of cesium chloride across the centrifuge cell. The result was a stable density gradient ranging from 1.66 to 1.76 g cm^{-3}. The DNA molecules in this density gradient were driven by centrifugal force into the region where the solution's density was equal to their own. The DNA yielded a narrow band that was detected by its absorption of ultraviolet light. A mixture of ^{14}N DNA and ^{15}N DNA molecules gave clearly separate bands because they differ in density by about 1% (Figure 4.20).

DNA was extracted from the bacteria at various times after they were transferred from a ^{15}N to a ^{14}N medium. Analysis of these samples by the density-gradient technique showed that there was a single band of DNA after one generation. The density of this band was precisely halfway between the densities of the ^{14}N DNA and ^{15}N DNA bands (Figure 4.21). *The*

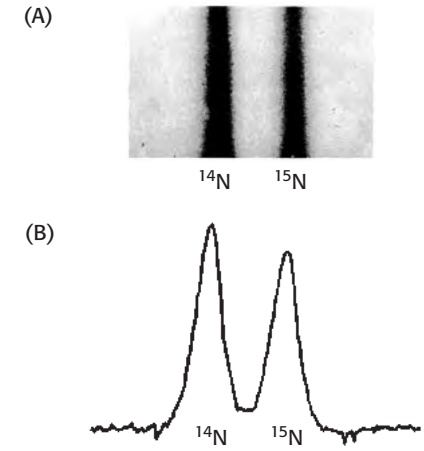


Figure 4.20 Resolution of ^{14}N DNA and ^{15}N DNA by density-gradient centrifugation. (A) Ultraviolet-absorption photograph of a centrifuged cell showing the two distinct bands of DNA. (B) Densitometric tracing of the absorption photograph. [From M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

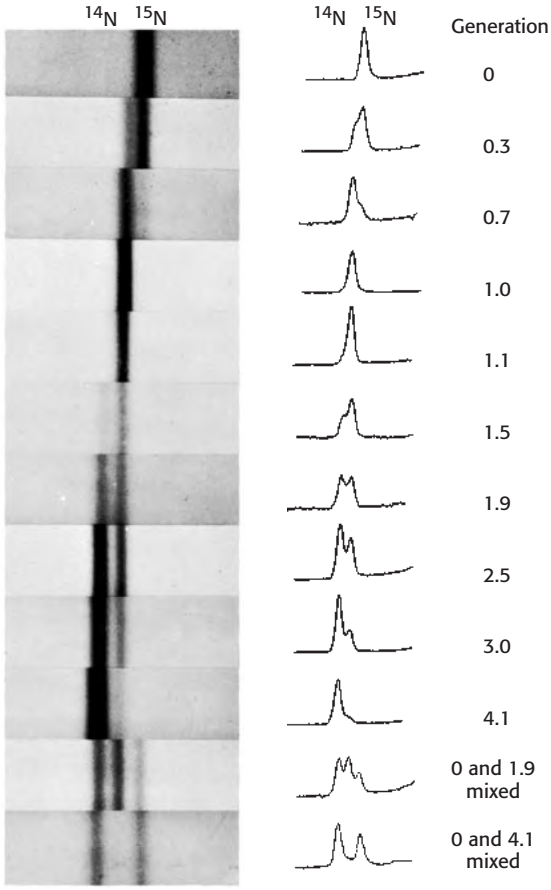


Figure 4.21 Detection of semiconservative replication of *E. coli* DNA by density-gradient centrifugation. The position of a band of DNA depends on its content of ^{14}N and ^{15}N. After 1.0 generation, all of the DNA molecules were hybrids containing equal amounts of ^{14}N and ^{15}N. [From M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

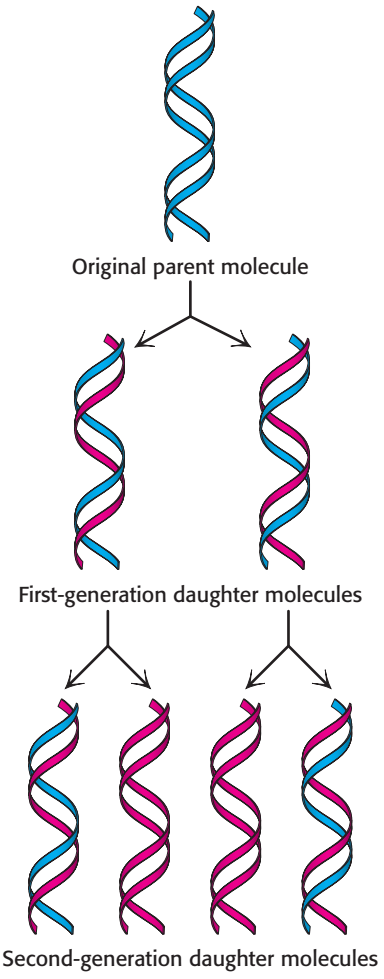


Figure 4.22 Diagram of semiconservative replication. Parental DNA is shown in blue and newly synthesized DNA in red. [After M. Meselson and F. W. Stahl. *Proc. Natl. Acad. Sci. U. S. A.* 44:671–682, 1958.]

absence of ^{15}N DNA indicated that parental DNA was not preserved as an intact unit after replication. The absence of ^{14}N DNA indicated that all the daughter DNA derived some of their atoms from the parent DNA. This proportion had to be half because the density of the hybrid DNA band was halfway between the densities of the ^{14}N DNA and ^{15}N DNA bands.

After two generations, there were equal amounts of two bands of DNA. One was hybrid DNA, and the other was ^{14}N DNA. Meselson and Stahl concluded from these incisive experiments that replication was semiconservative, and so each new double helix contains a parent strand and a newly synthesized strand. Their results agreed perfectly with the Watson–Crick model for DNA replication (Figure 4.22).

The double helix can be reversibly melted

In DNA replication and other processes, the two strands of the double helix must be separated from each other, at least in a local region. The two strands of a DNA helix readily come apart when the hydrogen bonds between base pairs are disrupted. In the laboratory, the double helix can be disrupted by heating a solution of DNA or by adding acid or alkali to ionize its bases. The dissociation of the double helix is called *melting* because it occurs abruptly at a certain temperature. The *melting temperature* (T_m) of DNA is defined as the temperature at which half the helical structure is lost. Inside cells, however, the double helix is not melted by the addition of heat. Instead, proteins called *helicases* use chemical energy (from ATP) to disrupt the helix (Chapter 28).

Stacked bases in nucleic acids absorb less ultraviolet light than do unstacked bases, an effect called *hypochromism*. Thus, the melting of nucleic acids is readily monitored by measuring their absorption of light, which is maximal at a wavelength of 260 nm (Figure 4.23).

Separated complementary strands of nucleic acids spontaneously reassociate to form a double helix when the temperature is lowered below T_m. This renaturation process is sometimes called *annealing*. The facility with which double helices can be melted and then reassociated is crucial for the biological functions of nucleic acids.

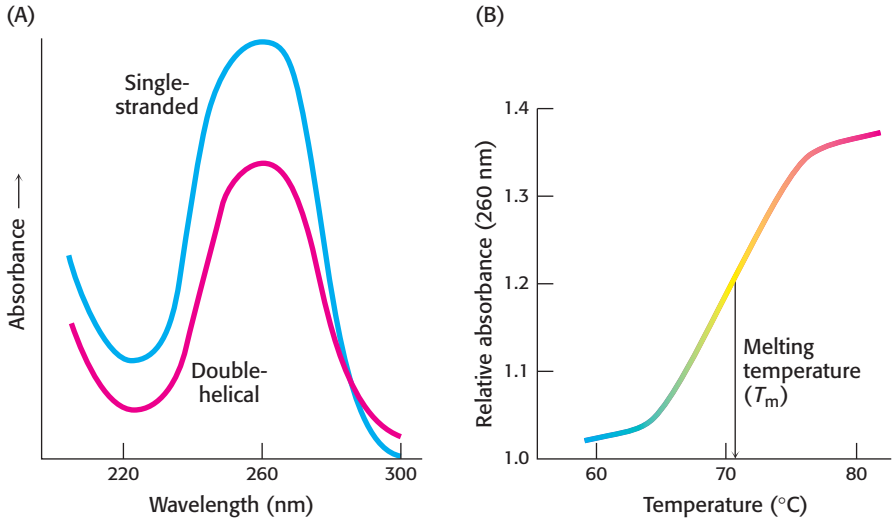


Figure 4.23 Hypochromism. (A) Single-stranded DNA absorbs light more effectively than does double-helical DNA. (B) The absorbance of a DNA solution at a wavelength of 260 nm increases when the double helix is melted into single strands.

The ability to melt and reanneal DNA reversibly in the laboratory provides a powerful tool for investigating sequence similarity. For instance, DNA molecules from two different organisms can be melted and allowed to reanneal, or *hybridize,* in the presence of each other. If the sequences are similar, hybrid DNA duplexes, with DNA from each organism contributing a strand of the double helix, can form. The degree of hybridization is an indication of the relatedness of the genomes and hence the organisms. Similar hybridization experiments with RNA and DNA can locate genes in a cell's DNA that correspond to a particular RNA. We will return to this important technique in Chapter 5.

4.4 DNA Is Replicated by Polymerases That Take Instructions from Templates

We now turn to the molecular mechanism of DNA replication. The full replication machinery in a cell comprises more than 20 proteins engaged in intricate and coordinated interplay. In 1958, Arthur Kornberg and his colleagues isolated from *E. coli* the first known of the enzymes, called *DNA polymerases,* that promote the formation of the bonds joining units of the DNA backbone. *E. coli* has a number of DNA polymerases, designated by roman numerals, that participate in DNA replication and repair (Chapter 28).

DNA polymerase catalyzes phosphodiester-bridge formation

DNA polymerases catalyze the step-by-step addition of deoxyribonucleotide units to a DNA chain (Figure 4.24). The reaction catalyzed, in its simplest form, is

$$(\text{DNA})_n + \text{dNTP} \rightleftharpoons (\text{DNA})_{n+1} + \text{PP}_\text{i}$$

where dNTP stands for any deoxyribonucleotide and PP_i is a pyrophosphate ion.

DNA synthesis has the following characteristics:

1. The reaction requires all four activated precursors—that is, *the deoxynucleoside 5′-triphosphates dATP, dGTP, dCTP, and TTP*—as well as Mg^{2+} ion.

2. *The new DNA chain is assembled directly on a preexisting DNA template.* DNA polymerases catalyze the formation of a phosphodiester linkage efficiently only if the base on the incoming nucleoside triphosphate is complementary to the base on the template strand. Thus, DNA polymerase is a *template-directed enzyme* that synthesizes a product with a base sequence complementary to that of the template.

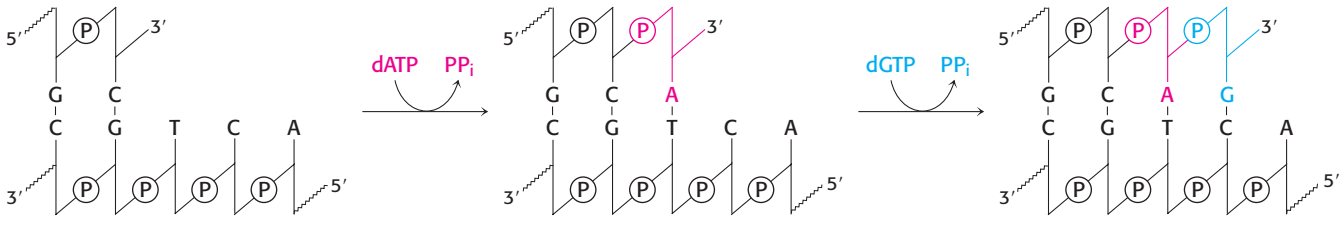


Figure 4.24 Polymerization reaction catalyzed by DNA polymerases.

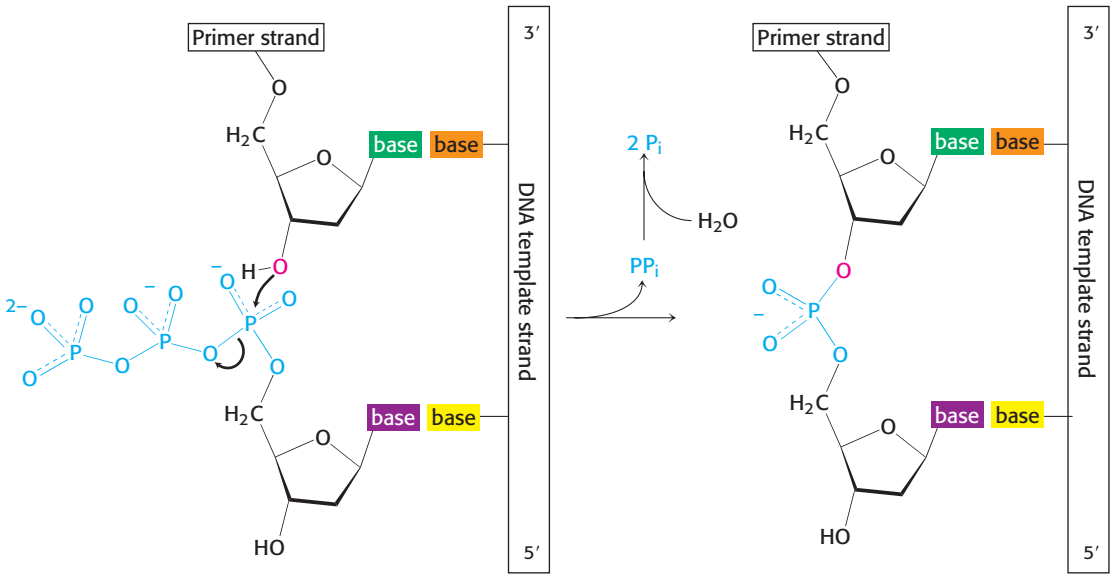


Figure 4.25 Chain-elongation reaction. DNA polymerases catalyze the formation of a phosphodiester bridge.

3. *DNA polymerases require a primer to begin synthesis.* A *primer* strand having a free 3′-OH group must be already bound to the template strand. The chain-elongation reaction catalyzed by DNA polymerases is a nucleophilic attack by the 3′-OH terminus of the growing chain on the innermost phosphorus atom of the deoxynucleoside triphosphate (Figure 4.25). A phosphodiester bridge is formed and pyrophosphate is released. The subsequent hydrolysis of pyrophosphate to yield two ions of orthophosphate (P_i) by *pyrophosphatase* helps drive the polymerization forward. *Elongation of the DNA chain proceeds in the 5′-to-3′ direction.*

4. *Many DNA polymerases are able to correct mistakes in DNA by removing mismatched nucleotides.* These polymerases have a distinct nuclease activity that allows them to excise incorrect bases by a separate reaction. This nuclease activity contributes to the remarkably high fidelity of DNA replication, which has an error rate of less than 10^{-8} per base pair.

The genes of some viruses are made of RNA

Genes in all cellular organisms are made of DNA. The same is true for some viruses but, for others, the genetic material is RNA. Viruses are genetic elements enclosed in protein coats that can move from one cell to another but are not capable of independent growth. A well-studied example of an RNA virus is the tobacco mosaic virus, which infects the leaves of tobacco plants. This virus consists of a single strand of RNA (6390 nucleotides) surrounded by a protein coat of 2130 identical subunits. An RNA polymerase that takes direction from an RNA template, called an *RNA-directed RNA polymerase*, copies the viral RNA. The infected cells die because of virus-instigated programmed cell death; in essence, the virus instructs the cell to commit suicide. Cell death results in discoloration in the tobacco leaf in a variegated pattern, hence the name mosaic virus.

Another important class of RNA virus comprises the *retroviruses,* so called because the genetic information flows from RNA to DNA rather than from DNA to RNA. This class includes human immunodeficiency virus 1 (HIV-1), the cause of AIDS, as well as a number of RNA viruses that produce tumors in susceptible animals. Retrovirus particles contain two copies of a single-stranded RNA molecule. On entering the cell, the RNA

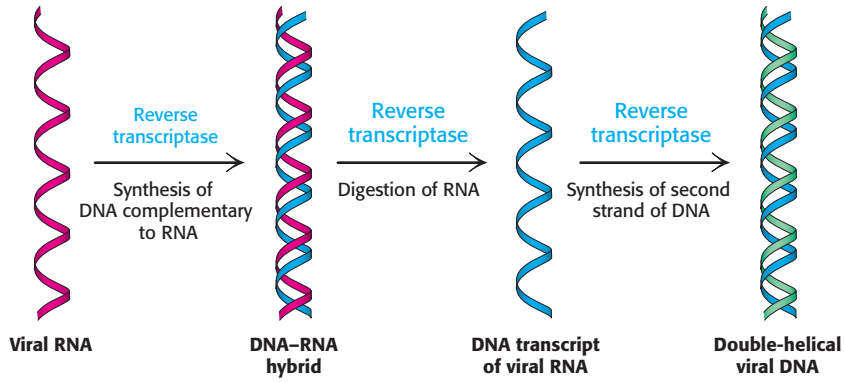


Figure 4.26 Flow of information from RNA to DNA in retroviruses. The RNA genome of a retrovirus is converted into DNA by reverse transcriptase, an enzyme brought into the cell by the infecting virus particle. Reverse transcriptase possesses several activities and catalyzes the synthesis of a complementary DNA strand, the digestion of the RNA, and the subsequent synthesis of the DNA strand.

is copied into DNA through the action of a viral enzyme called *reverse transcriptase* (Figure 4.26). The resulting double-helical DNA version of the viral genome can become incorporated into the chromosomal DNA of the host and is replicated along with the normal cellular DNA. At a later time, the integrated viral genome is expressed to form viral RNA and viral proteins, which assemble into new virus particles.

4.5 Gene Expression Is the Transformation of DNA Information into Functional Molecules

The information stored as DNA becomes useful when it is expressed in the production of RNA and proteins. This rich and complex topic is the subject of several chapters later in this book, but here we introduce the basics of gene expression. DNA can be thought of as archival information, stored and manipulated judiciously to minimize damage (mutations). It is expressed in two steps. First, an RNA copy is made that encodes directions for protein synthesis. This messenger RNA can be thought of as a photocopy of the original information: it can be made in multiple copies, used, and then disposed of. Second, the information in messenger RNA is translated to synthesize functional proteins. Other types of RNA molecules exist to facilitate this translation.

Several kinds of RNA play key roles in gene expression

Scientists used to believe that RNA played a passive role in gene expression, as a mere conveyor of information. However, recent investigations have shown that RNA plays a variety of roles, from catalysis to regulation. Cells contain several kinds of RNA (Table 4.3):

1. *Messenger RNA* (mRNA) is the template for protein synthesis, or *translation*. An mRNA molecule may be produced for each gene or group of genes that is to be expressed in *E. coli*, whereas a distinct mRNA is

Table 4.3 RNA molecules in *E. coli*

Type	Relative amount (%)	Sedimentation coefficient (*s*)	Mass (kd)	Number of nucleotides
Ribosomal RNA (rRNA)	80	23	1.2×10^3	3700
		16	0.55×10^3	1700
		5	3.6×10^1	120
Transfer RNA (tRNA)	15	4	2.5×10^1	75
Messenger RNA (mRNA)	5		Heterogeneous	

produced for each gene in eukaryotes. Consequently, mRNA is a heterogeneous class of molecules. In prokaryotes, the average length of an mRNA molecule is about 1.2 kilobases (kb). In eukaryotes, mRNA has structural features, such as stem-loop structures, that regulate the efficiency of translation and the lifetime of the mRNA.

Kilobase (kb)
A unit of length equal to 1000 base pairs of a double-stranded nucleic acid molecule (or 1000 bases of a single-stranded molecule).
One kilobase of double-stranded DNA has a length of 0.34 μm at its maximal extension (called the contour length) and a mass of about 660 kd.

2. *Transfer RNA* (tRNA) carries amino acids in an activated form to the ribosome for peptide-bond formation, in a sequence dictated by the mRNA template. There is at least one kind of tRNA for each of the 20 amino acids. Transfer RNA consists of about 75 nucleotides (having a mass of about 25 kd).

3. *Ribosomal RNA* (rRNA) is the major component of ribosomes (Chapter 30). In prokaryotes, there are three kinds of rRNA, called *23S, 16S,* and *5S RNA* because of their sedimentation behavior. One molecule of each of these species of rRNA is present in each ribosome. Ribosomal RNA was once believed to play only a structural role in ribosomes. We now know that rRNA is the actual catalyst for protein synthesis.

Ribosomal RNA is the most abundant of these three types of RNA. Transfer RNA comes next, followed by messenger RNA, which constitutes only 5% of the total RNA. Eukaryotic cells contain additional small RNA molecules.

4. *Small nuclear RNA* (snRNA) molecules participate in the splicing of RNA exons.

5. A small RNA molecule is an essential component of the *signal-recognition particle*, an RNA–protein complex in the cytoplasm that helps guide newly synthesized proteins to intracellular compartments and extracellular destinations.

6. *Micro RNA* (miRNA) is a class of small (about 21 nucleotides) noncoding RNAs that bind to complementary mRNA molecules and inhibit their translation.

7. *Small interfering RNA* (siRNA) is a class of small RNA molecules that bind to mRNA and facilitate its degradation. Micro RNA and small interfering RNA also provide scientists with powerful experimental tools for inhibiting the expression of specific genes in the cell.

8. RNA is a component of *telomerase,* an enzyme that maintains the telomeres (ends) of chromosomes during DNA replication.

In this chapter, we will consider rRNA, mRNA, and tRNA.

All cellular RNA is synthesized by RNA polymerases

The synthesis of RNA from a DNA template is called *transcription* and is catalyzed by the enzyme *RNA polymerase* (Figure 4.27). RNA polymerase catalyzes the initiation and elongation of RNA chains. The reaction catalyzed by this enzyme is

$$(\text{RNA})_n + \text{ribonucleoside triphosphate} \rightleftharpoons (\text{RNA})_{n+1} + \text{PP}_i$$

RNA polymerase requires the following components:

1. *A template.* The preferred template is *double-stranded DNA*. Single-stranded DNA also can serve as a template. RNA, whether single or double stranded, is not an effective template; nor are RNA–DNA hybrids.

2. *Activated precursors.* All four *ribonucleoside triphosphates*—ATP, GTP, UTP, and CTP—are required.

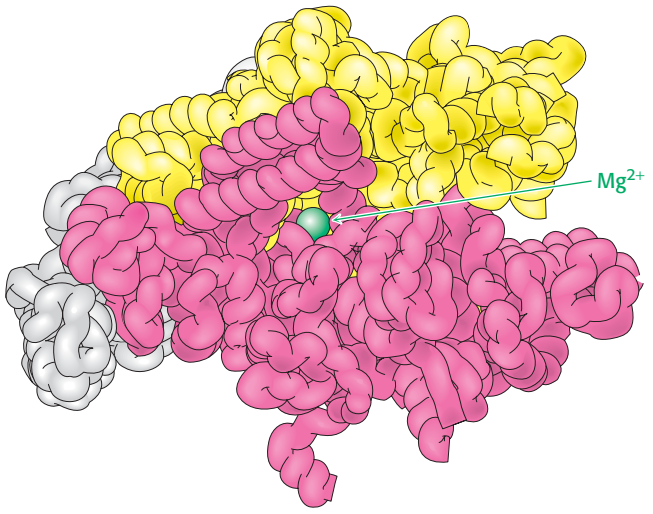


Figure 4.27 RNA Polymerase. This large enzyme comprises many subunits, including β (red) and β′ (yellow), which form a "claw" that holds the DNA to be transcribed. *Notice* that the active site includes a Mg^{2+} ion (green) at the center of the structure. The curved tubes making up the protein in the image represent the backbone of the polypeptide chain. [Drawn from IL9Z.pdb.]

3. *A divalent metal ion.* Either Mg^{2+} or Mn^{2+} is effective.

The synthesis of RNA is like that of DNA in several respects (Figure 4.28). First, the direction of synthesis is 5′ → 3′. Second, the mechanism of elongation is similar: the 3′-OH group at the terminus of the growing chain makes a nucleophilic attack on the innermost phosphoryl group of the incoming nucleoside triphosphate. Third, the synthesis is driven forward by the hydrolysis of pyrophosphate. In contrast with DNA polymerase, however, RNA polymerase does not require a primer. In addition, the ability of RNA polymerase to correct mistakes is not as extensive as that of DNA polymerase.

All three types of cellular RNA—mRNA, tRNA, and rRNA—are synthesized in *E. coli* by the same RNA polymerase according to

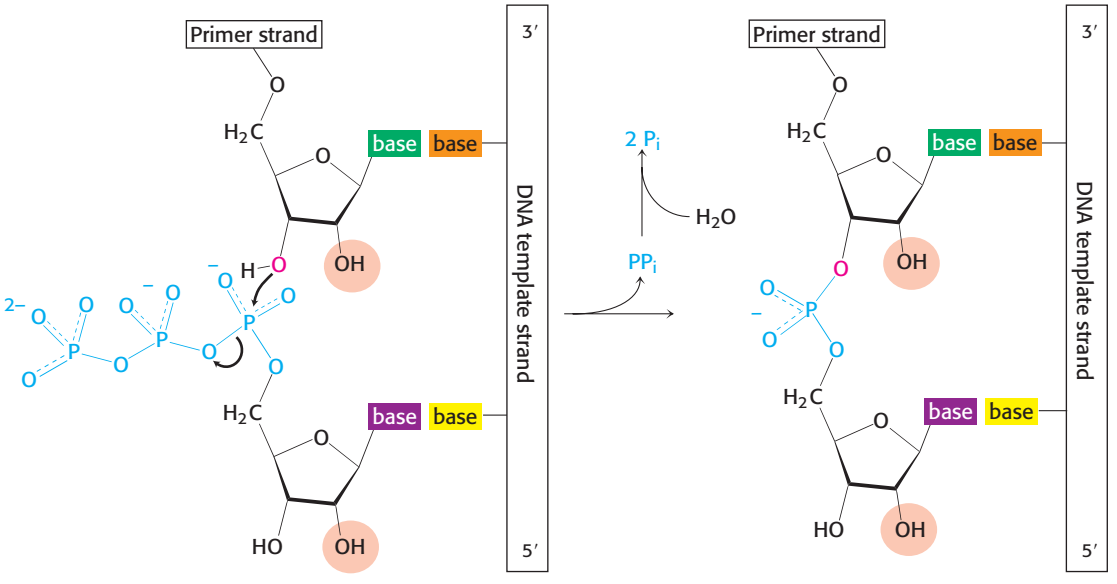


Figure 4.28 Transcription mechanism of the chain-elongation reaction catalyzed by RNA polymerase.

instructions given by a DNA template. In mammalian cells, there is a division of labor among several different kinds of RNA polymerases. We shall return to these RNA polymerases in Chapter 29.

RNA polymerases take instructions from DNA templates

RNA polymerase, like the DNA polymerases described earlier, takes instructions from a DNA template. The earliest evidence was the finding that the *base composition* of newly synthesized RNA is the complement of that of the DNA template strand, as exemplified by the RNA synthesized from a template of single-stranded DNA from the ϕX174 virus (Table 4.4). Hybridization experiments also revealed that RNA synthesized by RNA polymerase is complementary to its DNA template. In these experiments, DNA is melted and allowed to reassociate in the presence of mRNA. RNA–DNA hybrids will form if the RNA and DNA have complementary sequences. The strongest evidence for the fidelity of transcription came from base-sequence studies. For instance, the nucleotide sequence of a segment of the gene encoding the enzymes required for tryptophan synthesis was determined with the use of DNA-sequencing techniques (Section 5.1). Likewise, the sequence of the mRNA for the corresponding gene was determined. The results showed that the RNA sequence is the precise complement of the DNA template sequence (Figure 4.29).

Table 4.4 Base composition (percentage) of RNA synthesized from a viral DNA template

DNA template (plus, or coding, strand of ϕX174)		RNA product	
A	25	U	25
T	33	A	32
G	24	C	23
C	18	G	20

5′—GCGGCGACGCGCAGUUAAUCCCACAGCCGCCAGUUCCGCUGGCGGCAU—3′ **mRNA**
3′—CGCCGCTGCGCGTCAATTAGGGTGTCGGCGGTCAAGGCGACCGCCGTA—5′ **Template strand of DNA**
5′—GCGGCGACGCGCAGTTAATCCCACAGCCGCCAGTTCCGCTGGCGGCAT—3′ **Coding strand of DNA**

Figure 4.29 Complementarity between mRNA and DNA. The base sequence of mRNA (red) is the complement of that of the DNA template strand (blue). The sequence shown here is from the tryptophan operon, a segment of DNA containing the genes for five enzymes that catalyze the synthesis of tryptophan. The other strand of DNA (black) is called the coding strand because it has the same sequence as the RNA transcript except for thymine (T) in place of uracil (U).

Consensus sequence

Not all base sequences of promoter sites are identical. However, they do possess common features, which can be represented by an idealized consensus sequence. Each base in the consensus sequence TATAAT is found in most prokaryotic promoters. Nearly all promoter sequences differ from this consensus sequence at only one or two bases.

Transcription begins near promoter sites and ends at terminator sites

RNA polymerase must detect and transcribe discrete genes from within large stretches of DNA. What marks the beginning of the unit to be transcribed? DNA templates contain regions called *promoter sites* that specifically bind RNA polymerase and determine where transcription begins. In bacteria, two sequences on the 5′ (upstream) side of the first nucleotide to

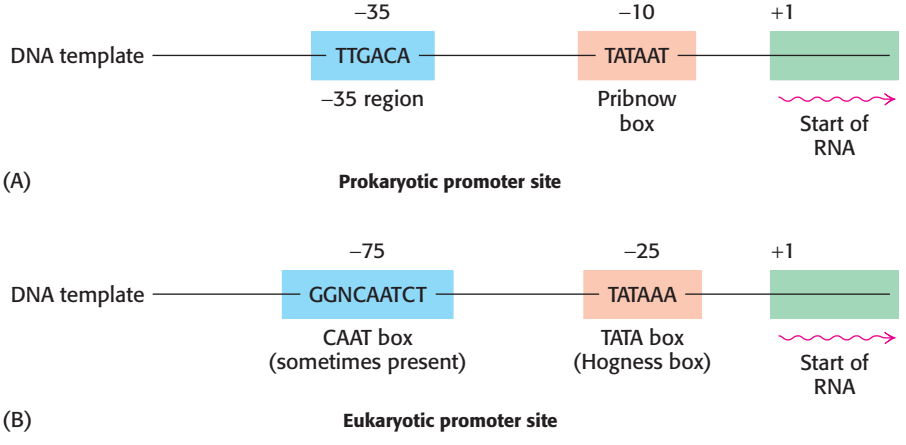


Figure 4.30 Promoter sites for transcription in (A) prokaryotes and (B) eukaryotes. Consensus sequences are shown. The first nucleotide to be transcribed is numbered +1. The adjacent nucleotide on the 5′ side is numbered −1. The sequences shown are those of the coding strand of DNA.

be transcribed function as promoter sites (Figure 4.30A). One of them, called the *Pribnow box*, has the consensus sequence TATAAT and is centered at −10 (10 nucleotides on the 5′ side of the first nucleotide transcribed, which is denoted by +1). The other, called the −35 *region*, has the consensus sequence TTGACA. The first nucleotide transcribed is usually a purine.

Eukaryotic genes encoding proteins have promoter sites with a TATAAA consensus sequence, called a *TATA box* or a *Hogness box*, centered at about −25 (Figure 4.30B). Many eukaryotic promoters also have a *CAAT box* with a GGN*CAAT*CT consensus sequence centered at about –75. The transcription of eukaryotic genes is further stimulated by *enhancer sequences*, which can be quite distant (as many as several kilobases) from the start site, on either its 5′ or its 3′ side.

RNA polymerase proceeds along the DNA template, transcribing one of its strands until it synthesizes a terminator sequence. This sequence encodes a termination signal, which in *E. coli* is a *base-paired hairpin* on the newly synthesized RNA molecule (Figure 4.31). This hairpin is formed by base-pairing of self-complementary sequences that are rich in G and C. Nascent RNA spontaneously dissociates from RNA polymerase when this hairpin is followed by a string of U residues. Alternatively, RNA synthesis can be terminated by the action of *rho*, a protein. Less is known about the termination of transcription in eukaryotes. A more detailed discussion of the initiation and termination of transcription will be given in Chapter 29. The important point now is that *discrete start and stop signals for transcription are encoded in the DNA template.*

In eukaryotes, the messenger RNA is modified after transcription (Figure 4.32). A "cap" structure is attached to the 5′ end, and a sequence of adenylates, the poly(A) tail, is added to the 3′ end. These modifications will be presented in detail in Chapter 29.

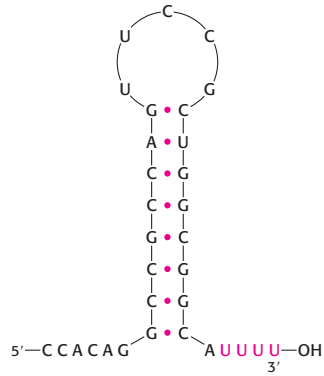


Figure 4.31 Base sequence of the 3′ end of an mRNA transcript in *E. coli*. A stable hairpin structure is followed by a sequence of uridine (U) residues.

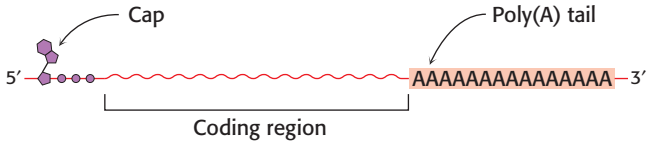


Figure 4.32 Modification of mRNA. Messenger RNA in eukaryotes is modified after transcription. A nucleotide "cap" structure is added to the 5′ end, and a poly(A) tail is added at the 3′ end.

Transfer RNAs are the adaptor molecules in protein synthesis

We have seen that mRNA is the template for protein synthesis. How then does it direct amino acids to become joined in the correct sequence to form a protein? In 1958, Francis Crick wrote:

> RNA presents mainly a sequence of sites where hydrogen bonding could occur. One would expect, therefore, that whatever went onto the template in a *specific* way did so by forming hydrogen bonds. It is therefore a natural hypothesis that the amino acid is carried to the template by an adaptor molecule, and that the adaptor is the part that actually fits onto the RNA. In its simplest form, one would require twenty adaptors, one for each amino acid.

This highly innovative hypothesis soon became established as fact. *The adaptors in protein synthesis are transfer RNAs*. The structure and reactions of these remarkable molecules will be considered in detail in Chapter 30. For the moment, it suffices to note that tRNAs contain an *amino acid-attachment site* and a *template-recognition site*. A tRNA molecule carries a specific amino acid in an activated form to the site of protein synthesis. The carboxyl group of this amino acid is esterified to the 3′- or 2′-hydroxyl group of the ribose unit at the 3′ end of the tRNA chain (Figure 4.33). The

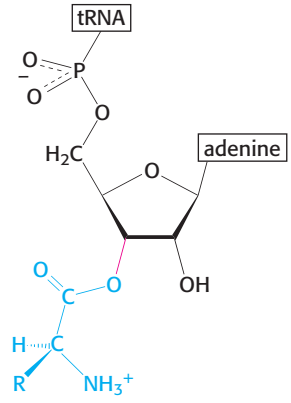


Figure 4.33 Attachment of an amino acid to a tRNA molecule. The amino acid (shown in blue) is esterified to the 3′-hydroxyl group of the terminal adenylate of tRNA.

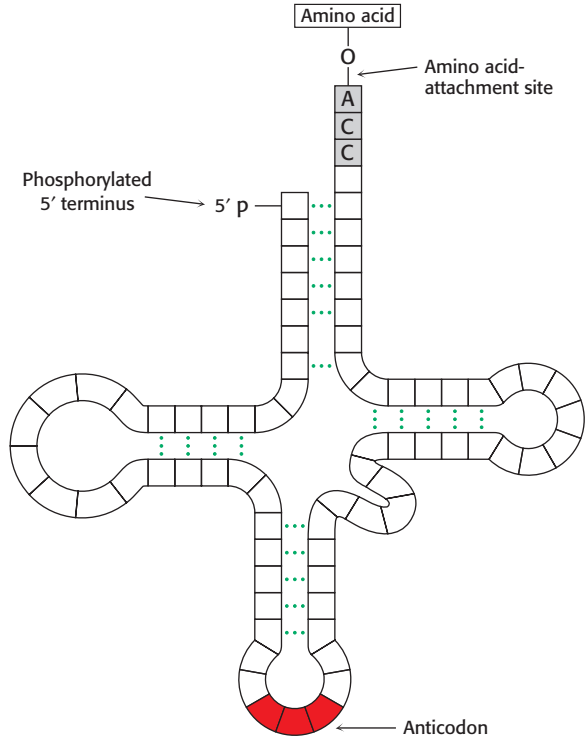


Figure 4.34 General structure of an aminoacyl-tRNA. The amino acid is attached at the 3′ end of the RNA. The anticodon is the template-recognition site. *Notice* that the tRNA has a cloverleaf structure with many hydrogen bonds (green dots) between bases.

joining of an amino acid to a tRNA molecule to form an *aminoacyl-tRNA* is catalyzed by a specific enzyme called an *aminoacyl-tRNA synthetase*. This esterification reaction is driven by ATP cleavage. There is at least one specific synthetase for each of the 20 amino acids. The template-recognition site on tRNA is a sequence of three bases called an *anticodon* (Figure 4.34). The anticodon on tRNA recognizes a complementary sequence of three bases, called a *codon*, on mRNA.

4.6 Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point

The *genetic code* is the relation between the sequence of bases in DNA (or its RNA transcripts) and the sequence of amino acids in proteins. Experiments by Marshall Nirenberg, Har Gobind Khorana, Francis Crick, Sydney Brenner, and others established the following features of the genetic code by 1961:

1. *Three nucleotides encode an amino acid*. Proteins are built from a basic set of 20 amino acids, but there are only four bases. Simple calculations show that a minimum of three bases is required to encode at least 20 amino acids. Genetic experiments showed that *an amino acid is in fact encoded by a group of three bases, or codon*.

2. *The code is nonoverlapping*. Consider a base sequence ABCDEF. In an overlapping code, ABC specifies the first amino acid, BCD the next, CDE the next, and so on. In a nonoverlapping code, ABC designates the first amino acid, DEF the second, and so forth. Genetic experiments again established the code to be nonoverlapping.

3. *The code has no punctuation*. In principle, one base (denoted as Q) might serve as a "comma" between groups of three bases.

. . . QABCQDEFQGHIQJKLQ . . .

However, it is not the case. Rather, *the sequence of bases is read sequentially from a fixed starting point*, without punctuation.

4. *The genetic code is degenerate*. Most amino acids are encoded by more than one codon. There are 64 possible base triplets and only 20 amino acids, and in fact 61 of the 64 possible triplets specify particular amino acids. Three triplets (called *stop codons*) designate the termination of translation. Thus, *for most amino acids, there is more than one code word.*

Major features of the genetic code

All 64 codons have been deciphered (Table 4.5). Because the code is highly degenerate, only tryptophan and methionine are encoded by just one triplet each. Each of the other 18 amino acids is encoded by two or more. Indeed, leucine, arginine, and serine are specified by six codons each. The number of codons for a particular amino acid correlates with its frequency of occurrence in proteins.

Codons that specify the same amino acid are called *synonyms*. For example, CAU and CAC are synonyms for histidine. Note that synonyms are not distributed haphazardly throughout the genetic code. In Table 4.5, an amino acid specified by two or more synonyms occupies a single box (unless it is specified by more than four synonyms). The amino acids in a box are specified by codons that have the same first two bases but differ in the third base, as exemplified by GUU, GUC, GUA, and GUG. Thus, *most synonyms differ only in the last base of the triplet*. Inspection of the code shows that XYC and XYU always encode the same amino acid, whereas XYG and XYA usually encode the same amino acid. The structural basis for these equivalences of codons will become evident

Table 4.5 The genetic code

First Position (5′ end)	Second Position U	C	A	G	Third Position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

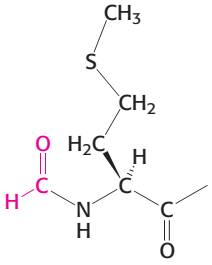
Note: This table identifies the amino acid encoded by each triplet. For example, the codon 5′- AUG-3′ on mRNA specifies methionine, whereas CAU specifies histidine. UAA, UAG, and UGA are termination signals. AUG is part of the initiation signal, in addition to coding for internal methionine residues.

when we consider the nature of the anticodons of tRNA molecules (Section 30.3).

What is the biological significance of the extensive degeneracy of the genetic code? If the code were not degenerate, 20 codons would designate amino acids and 44 would lead to chain termination. The probability of mutating to chain termination would therefore be much higher with a nondegenerate code. Chain-termination mutations usually lead to inactive proteins, whereas substitutions of one amino acid for another are usually rather harmless. Moreover, the code is constructed such that a change in any single nucleotide base of a codon results in a synonym or an amino acid with similar chemical properties. Thus, *degeneracy minimizes the deleterious effects of mutations*.

Messenger RNA contains start and stop signals for protein synthesis

Messenger RNA is translated into proteins on *ribosomes*—large molecular complexes assembled from proteins and ribosomal RNA. How is mRNA interpreted by the translation apparatus? The start signal for protein synthesis is complex in bacteria. Polypeptide chains in bacteria start with a modified amino acid—namely, formylmethionine (fMet). A specific tRNA, the initiator tRNA, carries fMet. This fMet-tRNA recognizes the codon AUG or, less frequently, GUG. However, AUG is also the codon for an internal methionine residue, and GUG is the codon for an internal valine residue. Hence, the signal for the first amino acid in a prokaryotic polypeptide chain must be more complex than that for all subsequent ones. *AUG (or GUG) is only part of the initiation signal* (Figure 4.35). In bacteria, the initiating AUG (or GUG) codon is preceded several nucleotides away by a purine-rich sequence, called the *Shine–Dalgarno sequence*, that base-pairs with a complementary sequence in a ribosomal RNA molecule (Section 30.3). In eukaryotes, the AUG closest to the 5′ end of an mRNA molecule is usually the start signal for protein synthesis. This particular AUG is read by an initiator tRNA conjugated to methionine. After the initiator AUG has been located, the *reading frame* is established—groups of three nonoverlapping nucleotides are defined, beginning with the initiator AUG codon.



fMet

As already mentioned, *UAA, UAG, and UGA designate chain termination*. These codons are read not by tRNA molecules but rather by specific proteins called *release factors* (Section 30.3). Binding of a release factor to the ribosome releases the newly synthesized protein.

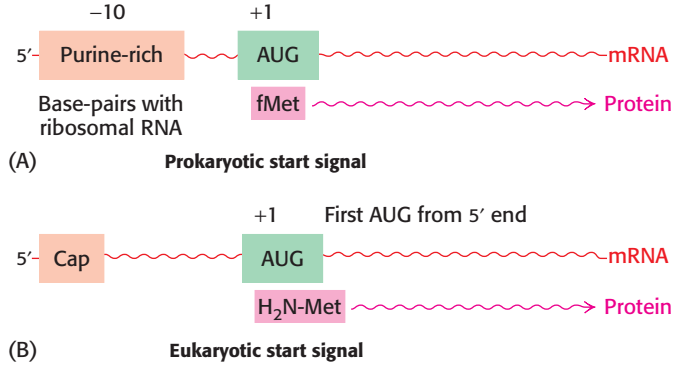


Figure 4.35 Initiation of protein synthesis. Start signals are required for the initiation of protein synthesis in (A) prokaryotes and (B) eukaryotes.

The genetic code is nearly universal

Is the genetic code the same in all organisms? This question was answered by examining the base sequences of many wild-type and mutant genes, as well as the amino acid sequences of their encoded proteins. For each mutant, the nucleotide change in the gene leads to a change in the amino acid as predicted by the genetic code. Furthermore, mRNAs can be correctly translated by the protein-synthesizing machinery of very different species. For example, human hemoglobin mRNA is correctly translated by a wheat-germ extract, and bacteria efficiently express recombinant DNA molecules encoding human proteins such as insulin. These experimental findings strongly suggested that the genetic code is universal.

A surprise was encountered when the sequence of human mitochondrial DNA became known. Ribosomes read UGA in human mitochondria as a codon for tryptophan rather than as a stop signal (Table 4.6). Furthermore, AGA and AGG are read as stop signals rather than as codons for arginine, and AUA is read as a codon for methionine instead of isoleucine. Mitochondria of other species, such as those of yeast, also have genetic codes that differ slightly from the standard one. The genetic code of mitochondria can differ from that of the rest of the cell because mitochondrial DNA encodes a distinct set of tRNAs. Do any cellular protein-synthesizing systems deviate from the standard genetic code? At least 16 organisms deviate from the standard genetic code. Ciliated protozoa differ from most organisms in reading UAA and UAG as codons for amino acids rather than as stop signals; UGA is their sole termination signal. Thus, *the genetic code is nearly but not absolutely universal*. Variations clearly exist in mitochondria and in species, such as ciliates, that branched off very early in eukaryotic evolution. It is interesting to note that two of the codon reassignments in human mitochondria diminish the information content of the third base of the triplet. For instance, in the common genetic code, AUG encodes methionine only while AUA is a codon for isoleucine. However, in human mitochondria both AUA and AUG specify methionine. Most variations from the standard genetic code are in the direction of a simpler code.

Why has the code remained nearly invariant through billions of years of evolution, from bacteria to human beings? A mutation that altered the reading of mRNA would change the amino acid sequence of most, if not all, proteins synthesized by that particular organism. Many of these changes would undoubtedly be deleterious, and so there would be strong selection against a mutation with such pervasive consequences.

Table 4.6 Distinctive codons of human mitochondria

Codon	Standard code	Mitochondrial code
UGA	Stop	Trp
UGG	Trp	Trp
AUA	Ile	Met
AUG	Met	Met
AGA	Arg	Stop
AGG	Arg	Stop

4.7 Most Eukaryotic Genes Are Mosaics of Introns and Exons

In bacteria, polypeptide chains are encoded by a continuous array of triplet codons in DNA. For many years, genes in higher organisms also were assumed to be continuous; the DNA sequence encoding a gene had a discrete beginning and ending with no other interrupting, noncoding DNA sequences. This view was unexpectedly shattered in 1977, when investigators, including Philip Sharp and Richard Roberts, discovered that several genes are *discontinuous*. The mosaic nature of eukaryotic genes was revealed by electron microscopic studies of hybrids formed between mRNA and a segment of DNA containing the corresponding gene (Figure 4.36). For example, the gene for the β chain of hemoglobin is interrupted within its amino acid-coding sequence by a long *intron* of 550 base pairs and a short one of 120 base pairs. Thus, the *β-globin gene is split into three coding*

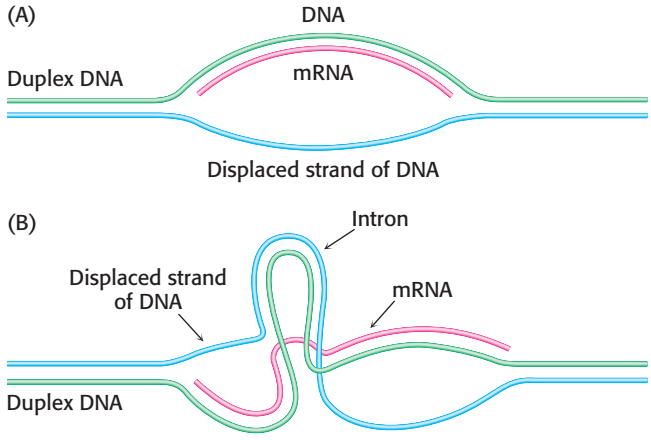


Figure 4.36 Detection of introns by electron microscopy. An mRNA molecule (shown in red) is hybridized to genomic DNA containing the corresponding gene. (A) A single loop of single-stranded DNA (shown in blue) is seen if the gene is continuous. (B) Two loops of single-stranded DNA (blue) and a loop of double-stranded DNA (blue and green) are seen if the gene contains an intron. Additional loops are evident if more than one intron is present.

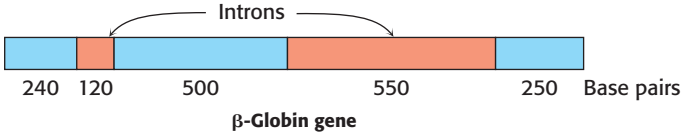


Figure 4.37 Structure of the β-globin gene.

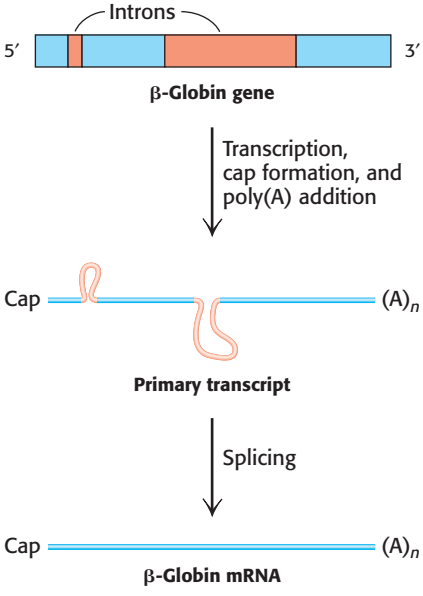


Figure 4.38 Transcription and processing of the β-globin gene. The gene is transcribed to yield the primary transcript, which is modified by cap and poly(A) addition. The introns in the primary RNA transcript are removed to form the mRNA.

sequences (Figure 4.37). The average human gene has 8 introns, and some have more than 100. The size ranges from 50 to 10,000 nucleotides.

RNA processing generates mature RNA

At what stage in gene expression are introns removed? Newly synthesized RNA chains (pre-mRNA or primary transcript) isolated from nuclei are much larger than the mRNA molecules derived from them; in regard to β-globin RNA, the former consists of approximately 1600 nucleotides and the latter approximately 900 nucleotides. In fact, the primary transcript of the β-globin gene contains two regions that are not present in the mRNA. *These regions in primary transcript are excised, and the coding sequences are simultaneously linked by a precise splicing enzyme to form the mature mRNA* (Figure 4.38). Regions that are removed from the primary transcript are called *introns* (for *int*ervening sequences), whereas those that are retained in the mature RNA are called *exons* (for *ex*pressed sequences). A common feature in the expression of discontinuous, or split, genes is that their exons are ordered in the same sequence in mRNA as in DNA. Thus, the codons in split genes, like continuous genes, are in the same linear order as the amino acids in the polypeptide products.

Splicing is a complex operation that is carried out by *spliceosomes,* which are assemblies of proteins and small RNA molecules. RNA plays the catalytic role (Section 29.3). This enzymatic machinery recognizes signals in the nascent RNA that specify the splice sites. *Introns nearly always begin*

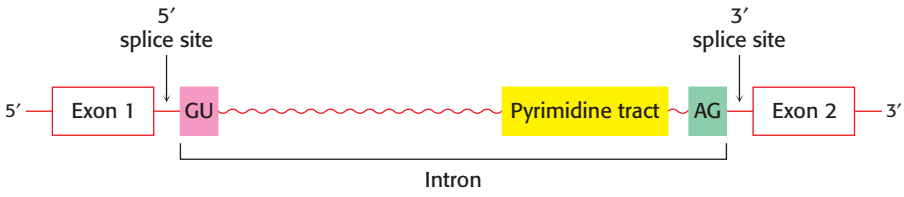


Figure 4.39 Consensus sequence for the splicing of mRNA precursors.

with GU and end with an AG that is preceded by a pyrimidine-rich tract (Figure 4.39). *This consensus sequence is part of the signal for splicing.*

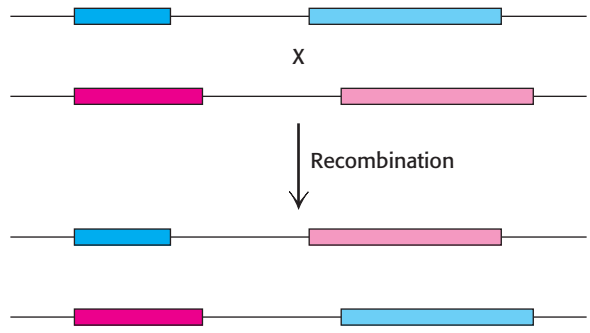


Figure 4.40 Exon shuffling. Exons can be readily shuffled by recombination of DNA to expand the genetic repertoire.

Many exons encode protein domains

Most genes of higher eukaryotes, such as birds and mammals, are split. Lower eukaryotes, such as yeast, have a much higher proportion of continuous genes. In prokaryotes, split genes are extremely rare. Have introns been inserted into genes in the evolution of higher organisms? Or have introns been removed from genes to form the streamlined genomes of prokaryotes and simple eukaryotes? Comparisons of the DNA sequences of genes encoding proteins that are highly conserved in evolution suggest that *introns were present in ancestral genes and were lost in the evolution of organisms that have become optimized for very rapid growth, such as prokaryotes.* The positions of introns in some genes are at least 1 billion years old. Furthermore, a common mechanism of splicing developed before the divergence of fungi, plants, and vertebrates, as shown by the finding that mammalian cell extracts can splice yeast RNA.

What advantages might split genes confer? *Many exons encode discrete structural and functional units of proteins.* An attractive hypothesis is that *new proteins arose in evolution by the rearrangement of exons encoding discrete structural elements, binding sites, and catalytic sites*, a process called *exon shuffling*. Because it preserves functional units but allows them to interact in new ways, exon shuffling is a rapid and efficient means of generating novel genes (Figure 4.40). Introns are extensive regions in which DNA can break and recombine with no deleterious effect on encoded proteins. In contrast, the exchange of sequences between different exons usually leads to loss of function.

Another advantage conferred by split genes is the potential for generating a series of related proteins by splicing a nascent RNA transcript in different ways. For example, a precursor of an antibody-producing cell forms an antibody that is anchored in the cell's plasma membrane (Figure 4.41). The attached antibody recognizes a specific foreign antigen, an event that leads to cell differentiation and proliferation. The activated antibody-producing cells then splice their nascent RNA transcript in an alternative manner to form soluble antibody molecules that are secreted rather than retained on the cell surface. We see here a clear-cut example of a benefit conferred by the complex arrangement of introns and exons in higher organisms. *Alternative splicing is a facile means of forming a set of proteins that are variations of a basic motif according to a developmental program without requiring a gene for each protein.*

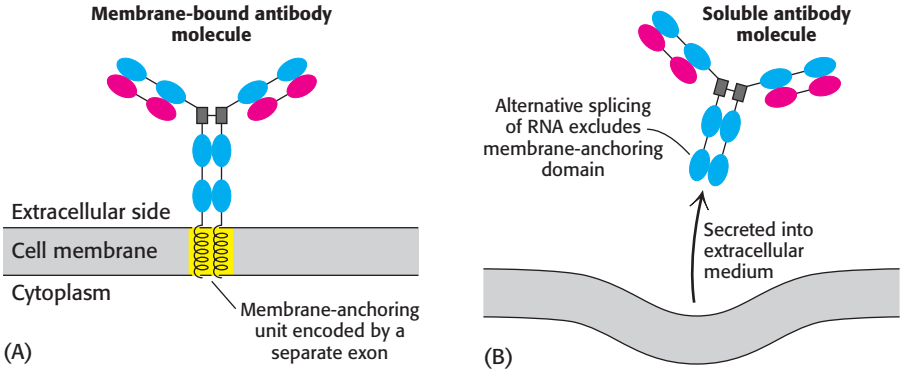


Figure 4.41 Alternative splicing. Alternative splicing generates mRNAs that are templates for different forms of a protein: (A) a membrane-bound antibody on the surface of a lymphocyte and (B) its soluble counterpart, exported from the cell. The membrane-bound antibody is anchored to the plasma membrane by a helical segment (highlighted in yellow) that is encoded by its own exon.

Summary

4.1 A Nucleic Acid Consists of Four Kinds of Bases Linked to a Sugar–Phosphate Backbone

DNA and RNA are linear polymers of a limited number of monomers. In DNA, the repeating units are nucleotides, with the sugar being a deoxyribose and the bases being adenine (A), thymine (T), guanine (G), and cytosine (C). In RNA, the sugar is a ribose and the base uracil (U) is used in place of thymine. DNA is the molecule of heredity in all prokaryotic and eukaryotic organisms. In viruses, the genetic material is either DNA or RNA.

4.2 A Pair of Nucleic Acid Chains with Complementary Sequences Can Form a Double-Helical Structure

All cellular DNA consists of two very long, helical polynucleotide chains coiled around a common axis. The sugar–phosphate backbone of each strand is on the outside of the double helix, whereas the purine and pyrimidine bases are on the inside. The two chains are held together by hydrogen bonds between pairs of bases: adenine is always paired with thymine, and guanine is always paired with cytosine. Hence, one strand of a double helix is the complement of the other. The two strands of the double helix run in opposite directions. Genetic information is encoded in the precise sequence of bases along a strand.

DNA is a structurally dynamic molecule that can exist in a variety of helical forms: A-DNA, B-DNA (the classic Watson–Crick helix), and Z-DNA. In A-, B-, and Z-DNA, two antiparallel chains are held together by Watson–Crick base pairs and stacking interactions between bases in the same strand. A- and B-DNA are right-handed helices. In B-DNA, the base pairs are nearly perpendicular to the helix axis. Z-DNA is a left-handed helix. Most of the DNA in a cell is in the B-form.

Double-stranded DNA can also wrap around itself to form a supercoiled structure. The supercoiling of DNA has two important consequences. Supercoiling compacts the DNA and, because supercoiled DNA is partly unwound, it is more accessible for interactions with other biomolecules.

Single-stranded nucleic acids, most notably RNA, can form complicated three-dimensional structures that may contain extensive double-helical regions that arise from the folding of the chain into hairpins.

4.3 The Double Helix Facilitates the Accurate Transmission of Hereditary Information

The structural nature of the double helix readily accounts for the accurate replication of genetic material because the sequence of bases in one strand determines the sequence of bases in the other strand. In replication, the strands of the helix separate and a new strand complementary to each of the original strands is synthesized. Thus, two new double helices are generated, each composed of one strand from the original molecule and one newly synthesized strand. This mode of replication is called semiconservative replication because each new helix retains one of the original strands.

In order for replication to take place, the strands of the double helix must be separated. In vitro, heating a solution of double-helical DNA separates the strands, a process called melting. On cooling, the strands reanneal and re-form the double helix. In the cell, special proteins temporarily separate the strands in replication.

4.4 DNA Is Replicated by Polymerases That Take Instructions from Templates

In the replication of DNA, the two strands of a double helix unwind and separate as new chains are synthesized. Each parent strand acts as a template for the formation of a new complementary strand. The replication of DNA is a complex process carried out by many proteins, including several DNA polymerases. The activated precursors in the synthesis of DNA are the four deoxyribonucleoside 5′-triphosphates. The new strand is synthesized in the 5′ → 3′ direction by a nucleophilic attack by the 3′-hydroxyl terminus of the primer strand on the innermost phosphorus atom of the incoming deoxyribonucleoside triphosphate. Most important, DNA polymerases catalyze the formation of a phosphodiester linkage only if the base on the incoming nucleotide is complementary to the base on the template strand. In other words, DNA polymerases are template-directed enzymes. The genes of some viruses, such as tobacco mosaic virus, are made of single-stranded RNA. An RNA-directed RNA polymerase mediates the replication of this viral RNA. Retroviruses, exemplified by HIV-1, have a single-stranded RNA genome that undergoes reverse transcription into double-stranded DNA by reverse transcriptase, an RNA-directed DNA polymerase.

4.5 Gene Expression Is the Transformation of DNA Information into Functional Molecules

The flow of genetic information in normal cells is from DNA to RNA to protein. The synthesis of RNA from a DNA template is called transcription, whereas the synthesis of a protein from an RNA template is termed translation. Cells contain several kinds of RNA, among which are messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA), which vary in size from 75 to more than 5000 nucleotides. All cellular RNA is synthesized by RNA polymerases according to instructions given by DNA templates. The activated intermediates are ribonucleoside triphosphates and the direction of synthesis, like that of DNA, is 5′ → 3′. RNA polymerase differs from DNA polymerase in not requiring a primer.

4.6 Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point

The genetic code is the relation between the sequence of bases in DNA (or its RNA transcript) and the sequence of amino acids in proteins. Amino acids are encoded by groups of three bases (called codons) starting from a fixed point. Sixty-one of the 64 codons specify particular amino acids, whereas the other 3 codons (UAA, UAG, and UGA) are signals for chain termination. Thus, for most amino acids, there is more than one code word. In other words, the code is degenerate. The genetic code is nearly the same in all organisms. Natural mRNAs contain start and stop signals for translation, just as genes do for directing where transcription begins and ends.

4.7 Most Eukaryotic Genes Are Mosaics of Introns and Exons

Most genes in higher eukaryotes are discontinuous. Coding sequences in these split genes, called exons, are separated by noncoding sequences, called introns, which are removed in the conversion of the primary transcript into mRNA and other functional mature RNA molecules. Split genes, like continuous genes, are colinear with their polypeptide products. A striking feature of many exons is that they encode functional domains in proteins. New proteins probably arose in the course of

evolution by the shuffling of exons. Introns may have been present in primordial genes but were lost in the evolution of such fast-growing organisms as bacteria and yeast.

Key Terms

double helix (p. 109)
deoxyribonucleic acid (DNA) (p. 110)
deoxyribose (p. 110)
ribose (p. 110)
purine (p. 111)
pyrimidine (p. 111)
ribonucleic acid (RNA) (p. 111)
nucleoside (p. 111)
nucleotide (p. 111)
B-DNA (p. 115)
A-DNA (p. 115)
Z-DNA (p. 116)
semiconservative replication (p. 118)
DNA polymerase (p. 121)
template (p. 121)
primer (p. 122)
reverse transcriptase (p. 122)
messenger RNA (mRNA) (p. 123)
translation (p. 123)
transfer RNA (tRNA) (p. 124)
ribosomal RNA (rRNA) (p. 124)
small nuclear RNA (snRNA) (p. 124)
micro RNA (miRNA) (p. 124)
small interfering RNA (siRNA) (p. 124)
transcription (p. 124)
RNA polymerase (p. 124)
promoter site (p. 126)
anticodon (p. 128)
codon (p. 128)
genetic code (p. 128)
ribosome (p. 130)
Shine–Dalgarno sequence (p. 130)
intron (p. 132)
exon (p. 132)
splicing (p. 132)
spliceosomes (p. 132)
exon shuffling (p. 133)
alternative splicing (p. 133)

Problems

1. *A t instead of an s?* Differentiate between a nucleoside and a nucleotide.

2. *A lovely pair*. What is a Watson–Crick base pair?

3. *Chargaff rules!* Biochemist Erwin Chargaff was the first to note that, in DNA, [A] = [T] and [G] = [C], equalities now called Chargraff's rule. Using this rule, determine the percentages of all the bases in DNA that is 20% thymine.

4. *But not always*. A single strand of RNA is 20% U. What can you predict about the percentages of the remaining bases?

5. *Complements*. Write the complementary sequence (in the standard $5' \longrightarrow 3'$ notation) for (a) GATCAA, (b) TCGAAC, (c) ACGCGT, and (d) TACCAT.

6. *Compositional constraint*. The composition (in mole-fraction units) of one of the strands of a double-helical DNA molecule is [A] = 0.30 and [G] = 0.24. (a) What can you say about [T] and [C] for the same strand? (b) What can you say about [A], [G], [T], and [C] of the complementary strand?

7. *Size matters*. Why are GC and AT the only base pairs permissible in the double helix?

8. *Strong, but not strong enough*. Why does heat denature, or melt, DNA in solution?

9. *Uniqueness*. The human genome contains 3 billion nucleotides arranged in a vast array of sequences. What is the minimum length of a DNA sequence that will, in all probability, appear only once in the human genome? You need consider only one strand and may assume that all four nucleotides have the same probability of appearance.

10. *Coming and going*. What does it mean to say that the DNA chains in a double helix have opposite polarity?

11. *All for one*. If the forces—hydrogen bonds and stacking forces—holding a helix together are weak, why is it difficult to disrupt a double helix?

12. *Overcharged*. DNA in the form of a double helix must be associated with cations, usually Mg^{2+}. Why is this requirement the case?

13. *Not quite from A to Z*. Describe the three forms that a double helix can assume.

14. *Lost DNA*. The DNA of a deletion mutant of λ bacteriophage has a length of 15 μm instead of 17 μm. How many base pairs are missing from this mutant?

15. *An unseen pattern*. What result would Meselson and Stahl have obtained if the replication of DNA were conservative (i.e., the parental double helix stayed together)? Give the expected distribution of DNA molecules after 1.0 and 2.0 generations for conservative replication.

16. *Tagging DNA*. (a) Suppose that you want to radioactively label DNA but not RNA in dividing and growing bacterial cells. Which radioactive molecule would you add to the culture medium? (b) Suppose that you want to prepare DNA in which the backbone phosphorus atoms are uniformly labeled with ^{32}P. Which precursors should be added to a solution containing DNA polymerase and primed template DNA? Specify the position of radioactive atoms in these precursors.

17. *Finding a template*. A solution contains DNA polymerase and the Mg^{2+} salts of dATP, dGTP, dCTP, and

TTP. The following DNA molecules are added to aliquots of this solution. Which of them would lead to DNA synthesis? (a) A single-stranded closed circle containing 1000 nucleotide units. (b) A double-stranded closed circle containing 1000 nucleotide pairs. (c) A single-stranded closed circle of 1000 nucleotides base-paired to a linear strand of 500 nucleotides with a free 3′-OH terminus. (d) A double-stranded linear molecule of 1000 nucleotide pairs with a free 3′-OH group at each end.

18. *Retrograde*. What is a retrovirus and how does information flow for a retrovirus differ from that for the infected cell?

19. *The right start*. Suppose that you want to assay reverse transcriptase activity. If polyriboadenylate is the template in the assay, what should you use as the primer? Which radioactive nucleotide should you use to follow chain elongation?

20. *Essential degradation*. Reverse transcriptase has ribonuclease activity as well as polymerase activity. What is the role of its ribonuclease activity?

21. *Virus hunting*. You have purified a virus that infects turnip leaves. Treatment of a sample with phenol removes viral proteins. Application of the residual material to scraped leaves results in the formation of progeny virus particles. You infer that the infectious substance is a nucleic acid. Propose a simple and highly sensitive means of determining whether the infectious nucleic acid is DNA or RNA.

22. *Mutagenic consequences*. Spontaneous deamination of cytosine bases in DNA takes place at low but measurable frequency. Cytosine is converted into uracil by loss of its amino group. After this conversion, which base pair occupies this position in each of the daughter strands resulting from one round of replication? Two rounds of replication?

23. *Information content*. (a) How many different 8-mer sequences of DNA are there? (Hint: There are 16 possible dinucleotides and 64 possible trinucleotides.) We can quantify the information-carrying capacity of nucleic acids in the following way. Each position can be one of four bases, corresponding to two bits of information ($2^2 = 4$). Thus, a chain of 5100 nucleotides corresponds to $2 \times 5100 = 10{,}200$ bits, or 1275 bytes (1 byte = 8 bits). (b) How many bits of information are stored in an 8-mer DNA sequence? In the *E. coli* genome? In the human genome? (c) Compare each of these values with the amount of information that can be stored on a computer compact disc, or CD (about 700 megabytes).

24. *Key polymerases*. Compare DNA polymerase and RNA polymerase from *E. coli* in regard to each of the following features: (a) activated precursors, (b) direction of chain elongation, (c) conservation of the template, and (d) need for a primer.

25. *Family resemblance*. Differentiate among mRNA, rRNA and tRNA.

26. *Encoded sequences*. (a) Write the sequence of the mRNA molecule synthesized from a DNA template strand having the following sequence.

5′–ATCGTACCGTTA–3′

(b) What amino acid sequence is encoded by the following base sequence of an mRNA molecule? Assume that the reading frame starts at the 5′ end.

5′–UUGCCUAGUGAUUGGAUG–3′

(c) What is the sequence of the polypeptide formed on addition of poly(UUAC) to a cell-free protein-synthesizing system?

27. *A tougher chain*. RNA is readily hydrolyzed by alkali, whereas DNA is not. Why?

28. *A picture is worth a thousand words*. Write a reaction sequence showing why RNA is more susceptible to nucleophilic attack than DNA.

29. *Flowing information*. What is meant by the phrase gene expression?

30. *We can all agree on that*. What is a consensus sequence?

31. *A potent blocker*. How does cordycepin (3′-deoxyadenosine) block the synthesis of RNA?

32. *Silent RNA*. The code word GGG cannot be deciphered in the same way as can UUU, CCC, and AAA, because poly(G) does not act as a template. Poly(G) forms a triple-stranded helical structure. Why is it an ineffective template?

33. *Sometimes it is not so bad*. What is meant by the degeneracy of the genetic code?

34. *In fact, it can be good*. What is the biological benefit of a degenerate genetic code?

35. *To bring together as associates*. Match the components in the right-hand column with the appropriate process in the left-hand column.

(a) Replication ____
(b) Transcription ____
(c) Translation ____

1. RNA polymerase
2. DNA polymerase
3. Ribosome
4. dNTP
5. tRNA
6. NTP
7. mRNA
8. primer
9. rRNA
10. promoter

36. *A lively contest.* Match the components in the right-hand column with the appropriate process in the left-hand column.

(a) fMet	1. continuous message
(b) Shine–Dalgarno	2. removed
(c) intron	3. the first of many
(d) exon	4. uniter
(e) pre-mRNA	5. joined
(f) mRNA	6. locate the start
(g) spliceosome	7. discontinuous message

37. *Two from one.* Synthetic RNA molecules of defined sequence were instrumental in deciphering the genetic code. Their synthesis first required the synthesis of DNA molecules to serve as templates. H. Gobind Khorana synthesized, by organic-chemical methods, two complementary deoxyribonucleotides, each with nine residues: $d(TAC)_3$ and $d(GTA)_3$. Partly overlapping duplexes that formed on mixing these oligonucleotides then served as templates for the synthesis by DNA polymerase of long, repeating double-helical DNA chains. The next step was to obtain long polyribonucleotide chains with a sequence complementary to only one of the two DNA strands. How did Khorana obtain only poly(UAC)? Only poly(GUA)?

38. *Triple entendre.* The RNA transcript of a region of T4 phage DNA contains the sequence 5′-AAAUGAGGA-3′. This sequence encodes three different polypeptides. What are they?

39. *Valuable synonyms.* Proteins generally have low contents of Met and Trp, intermediate contents of His and Cys, and high contents of Leu and Ser. What is the relation between the number of an amino acid's codons and the frequency with which the amino acid is present in proteins? What might be the selective advantage of this relation?

40. *A new translation.* A transfer RNA with a UGU anticodon is enzymatically conjugated to ^{14}C-labeled cysteine. The cysteine unit is then chemically modified to alanine (with the use of Raney nickel, which removes the sulfur atom of cysteine). The altered aminoacyl-tRNA is added to a protein-synthesizing system containing normal components except for this tRNA. The mRNA added to this mixture contains the following sequence:

5′–UUUUGCCAUGUUUGUGCU–3′

What is the sequence of the corresponding radiolabeled peptide?

41. *A tricky exchange.* Define exon shuffling and explain why its occurrence might be an evolutionary advantage.

42. *The unity of life.* What is the significance of the fact that human mRNA can be accurately translated in *E. coli*?

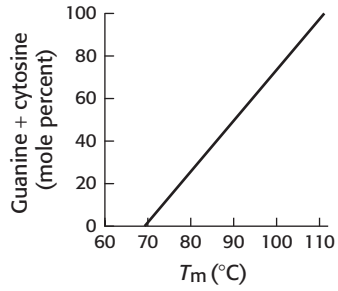
Chapter Integration Problems

43. *Back to the bench.* A protein chemist told a molecular geneticist that he had found a new mutant hemoglobin in which aspartate replaced lysine. The molecular geneticist expressed surprise and sent his friend scurrying back to the laboratory. (a) Why did the molecular geneticist doubt the reported amino acid substitution? (b) Which amino acid substitutions would have been more palatable to the molecular geneticist?

44. *Eons apart.* The amino acid sequences of a yeast protein and a human protein having the same function are found to be 60% identical. However, the corresponding DNA sequences are only 45% identical. Account for this differing degree of identity.

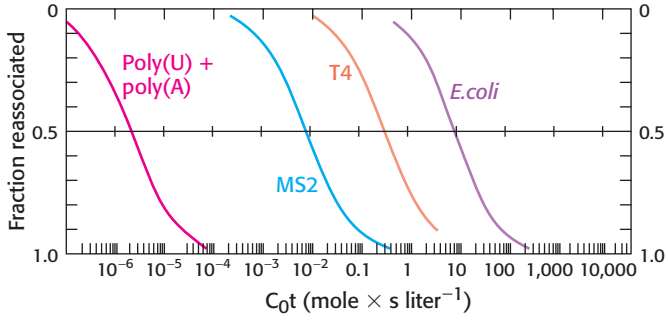
Data Interpretation Problems

45. *3 is greater than 2.* The adjoining illustration graphs the relation between the percentage of GC base pairs in DNA and the melting temperature. Account for these results.



[After R. J. Britten and D. E. Kohne, *Science* 161:529–540, 1968.]

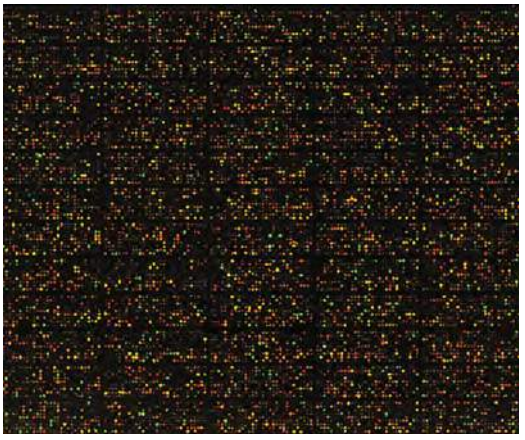
46. *Blast from the past.* The illustration below is a graph called a C_0t curve (pronounced "cot"). The *y*-axis shows the percentage of DNA that is double stranded. The *x*-axis is the product of the concentration of DNA and the time required for the double-stranded molecules to form. Explain why the mixture of poly(A) and poly(U) and the three DNAs shown vary in the C_0t value required to completely anneal. MS2 and T4 are bacterial viruses (bacteriophages) with genome sizes of 3569 and 168,903 bp, respectively. The *E. coli* genome is 4.6×10^6 bp.



[After J. Marmur and P. Doty, *J. Mol. Biol.* 5:120, 1962.]

CHAPTER 5

Exploring Genes and Genomes



Processes such as the development from a caterpillar into a butterfly entail dramatic changes in patterns of gene expression. The expression levels of thousands of genes can be monitored through the use of DNA arrays. At the right, a DNA microarray reveals the expression levels of more than 12,000 human genes; the brightness of each spot indicates the expression level of the corresponding gene. [(Left) Cathy Keifer/istockphoto.com. (Right) Agilent Technologies.]

Since its emergence in the 1970s, recombinant DNA technology has revolutionized biochemistry. The genetic endowment of organisms can now be precisely changed in designed ways. Recombinant DNA technology is the fruit of several decades of basic research on DNA, RNA, and viruses. It depends, first, on having enzymes that can cut, join, and replicate DNA and those that can reverse transcribe RNA. Restriction enzymes cut very long DNA molecules into specific fragments that can be manipulated; DNA ligases join the fragments together. Many kinds of restriction enzymes are available. By applying this assortment cleverly, researchers can treat DNA sequences as modules that can be moved at will from one DNA molecule to another. Thus, recombinant DNA technology is based on the use of enzymes that act on nucleic acids as substrates.

A second foundation is the base-pairing language that allows complementary sequences to recognize and bind to each other. Hybridization with complementary DNA (cDNA) or RNA probes is a sensitive means of detecting specific nucleotide sequences. In recombinant DNA technology, base-pairing is used to construct new combinations of DNA as well as to detect and amplify particular sequences.

Third, powerful methods have been developed for determining the sequence of nucleotides in DNA. These methods have been harnessed to

OUTLINE

sequence complete genomes: first, small genomes from viruses; then, larger genomes from bacteria; and, finally, eukaryotic genomes, including the 3-billion-base-pair human genome. Scientists are just beginning to exploit the enormous information content of these genome sequences.

Finally, recombinant DNA technology critically depends on our ability to deliver foreign DNA into host organisms. For example, DNA fragments can be inserted into plasmids, where they can be replicated within a short period of time in their bacterial hosts. In addition, viruses efficiently deliver their own DNA (or RNA) into hosts, subverting them either to replicate the viral genome and produce viral proteins or to incorporate viral DNA into the host genome.

These new methods have wide-ranging benefits across a broad spectrum of disciplines, including biotechnology, agriculture, and medicine. Among these benefits is the dramatic expansion of our understanding of human disease. Throughout this chapter, a specific disorder, amyotrophic lateral sclerosis (ALS), will be used to illustrate the effect that recombinant DNA technology has had on our knowledge of disease mechanisms. ALS was first described clinically in 1869 by the French neurologist Jean-Martin Charcot as a fatal neurodegenerative disease of progressive weakening and atrophy of voluntary muscles. ALS is commonly referred to as Lou Gehrig's Disease, for the baseball legend whose career and life were prematurely cut short as a result of this devastating disease. For many years, little progress had been made in the study of the mechanisms underlying ALS. As we shall see, significant advances have been made with the use of research tools facilitated by recombinant DNA technology.

5.1 The Exploration of Genes Relies on Key Tools

The rapid progress in biotechnology—indeed its very existence—is a result of a few key techniques.

1. *Restriction-Enzyme Analysis.* Restriction enzymes are precise molecular scalpels that allow an investigator to manipulate DNA segments.

2. *Blotting Techniques.* Southern and northern blots are used to separate and characterize DNA and RNA, respectively. The western blot, which uses antibodies to characterize proteins, was described in Chapter 3.

3. *DNA Sequencing.* The precise nucleotide sequence of a molecule of DNA can be determined. Sequencing has yielded a wealth of information concerning gene architecture, the control of gene expression, and protein structure.

4. *Solid-Phase Synthesis of Nucleic Acids.* Precise sequences of nucleic acids can be synthesized de novo and used to identify or amplify other nucleic acids.

5. *The Polymerase Chain Reaction* (PCR). The polymerase chain reaction leads to a billionfold amplification of a segment of DNA. One molecule of DNA can be amplified to quantities that permit characterization and manipulation. This powerful technique can be used to detect pathogens and genetic diseases, determine the source of a hair left at the scene of a crime, and resurrect genes from the fossils of extinct organisms.

A final set of techniques relies on the computer, without which, it would be impossible to catalog, access, and characterize the abundant information

generated by the techniques just outlined. Such uses of the computer will be presented in Chapter 6.

Restriction enzymes split DNA into specific fragments

Restriction enzymes, also called *restriction endonucleases,* recognize specific base sequences in double-helical DNA and cleave, at specific places, both strands of that duplex. To biochemists, these exquisitely precise scalpels are marvelous gifts of nature. They are indispensable for analyzing chromosome structure, sequencing very long DNA molecules, isolating genes, and creating new DNA molecules that can be cloned. Werner Arber and Hamilton Smith discovered restriction enzymes, and Daniel Nathans pioneered their use in the late 1960s.

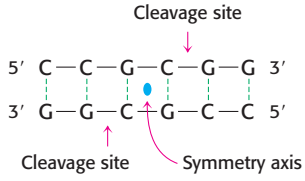
Restriction enzymes are found in a wide variety of prokaryotes. Their biological role is to cleave foreign DNA molecules. Many restriction enzymes recognize specific sequences of four to eight base pairs and hydrolyze a phosphodiester bond in each strand in this region. A striking characteristic of these cleavage sites is that they almost always possess *twofold rotational symmetry.* In other words, the recognized sequence is *palindromic,* or an inverted repeat, and the cleavage sites are symmetrically positioned. For example, the sequence recognized by a restriction enzyme from *Streptomyces achromogenes* is

Palindrome

A word, sentence, or verse that reads the same from right to left as it does from left to right.

Radar
Senile felines
Do geese see God?
Roma tibi subito motibus ibit amor

Derived from the Greek *palindromos,* "running back again."



In each strand, the enzyme cleaves the C–G phosphodiester bond on the 3′ side of the symmetry axis. As we shall see in Chapter 9, this symmetry corresponds to that of the structures of the restriction enzymes themselves.

Several hundred restriction enzymes have been purified and characterized. Their names consist of a three-letter abbreviation for the host organism (e.g., *Eco* for *Escherichia coli, Hin* for *Haemophilus influenzae, Hae* for *Haemophilus aegyptius*) followed by a strain designation (if needed) and a roman numeral (if more than one restriction enzyme from the same strain has been identified). The specificities of several of these enzymes are shown in Figure 5.1.

Restriction enzymes are used to cleave DNA molecules into specific fragments that are more readily analyzed and manipulated than the entire parent molecule. For example, the 5.1-kb circular duplex DNA of the tumor-producing SV40 virus is cleaved at one site by *Eco*RI, at four sites by *Hpa*I, and at 11 sites by *Hin*dIII. A piece of DNA, called a restriction fragment, produced by the action of one restriction enzyme can be specifically cleaved into smaller fragments by another restriction enzyme. The pattern of such fragments can serve as a *fingerprint* of a DNA molecule, as will be considered shortly. Indeed, complex chromosomes containing hundreds of millions of base pairs can be mapped by using a series of restriction enzymes.

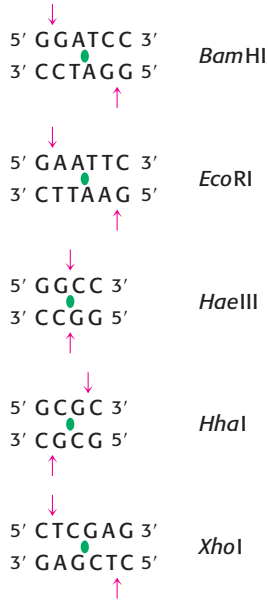


Figure 5.1 Specificities of some restriction endonucleases. The sequences that are recognized by these enzymes contain a twofold axis of symmetry. The two strands in these regions are related by a 180-degree rotation about the axis marked by the green symbol. The cleavage sites are denoted by red arrows. The abbreviated name of each restriction enzyme is given at the right of the sequence that it recognizes. Note that the cuts may be staggered or even.

Restriction fragments can be separated by gel electrophoresis and visualized

Small differences between related DNA molecules can be readily detected because their restriction fragments can be separated and displayed by gel electrophoresis. In Chapter 3, we considered the use of gel electrophoresis

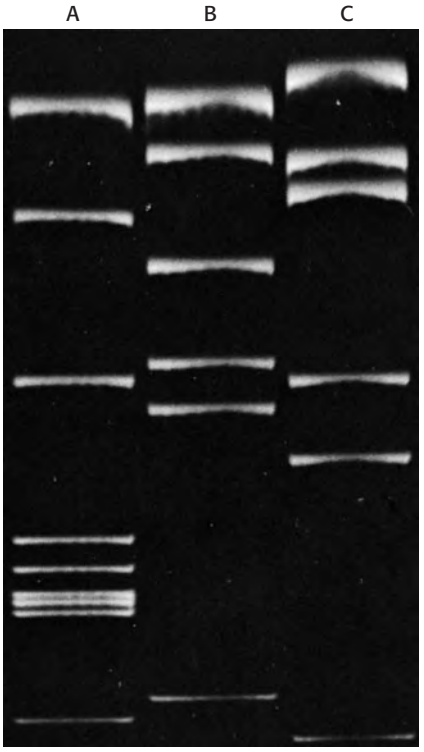


Figure 5.2 Gel-electrophoresis pattern of a restriction digest. This gel shows the fragments produced by cleaving SV40 DNA with each of three restriction enzymes. These fragments were made fluorescent by staining the gel with ethidium bromide. [Courtesy of Dr. Jeffrey Sklar.]

to separate protein molecules (Section 3.1). Because the phosphodiester backbone of DNA is highly negatively charged, this technique is also suitable for the separation of nucleic acid fragments. For most gels, the shorter the DNA fragment, the farther the migration. Polyacrylamide gels are used to separate, by size, fragments containing as many as 1000 base pairs, whereas more-porous agarose gels are used to resolve mixtures of larger fragments (as large as 20 kb). An important feature of these gels is their high resolving power. In certain kinds of gels, fragments differing in length by just one nucleotide of several hundred can be distinguished. Bands or spots of radioactive DNA in gels can be visualized by autoradiography. Alternatively, a gel can be stained with ethidium bromide, which fluoresces an intense orange when bound to a double-helical DNA molecule (Figure 5.2). A band containing only 50 ng of DNA can be readily seen.

A restriction fragment containing a specific base sequence can be identified by hybridizing it with a labeled complementary DNA strand (Figure 5.3). A mixture of restriction fragments is separated by electrophoresis through an agarose gel, denatured to form single-stranded DNA, and transferred to a nitrocellulose sheet. The positions of the DNA fragments in the gel are preserved on the nitrocellulose sheet, where they are exposed to a ^{32}P-labeled single-stranded *DNA probe*. The probe hybridizes with a restriction fragment having a complementary sequence, and autoradiography then reveals the position of the restriction-fragment–probe duplex. A particular fragment amid a million others can be readily identified in this way. This powerful technique is named *Southern blotting,* for its inventor Edwin Southern.

Similarly, RNA molecules can be separated by gel electrophoresis, and specific sequences can be identified by hybridization subsequent to their transfer to nitrocellulose. This analogous technique for the analysis of RNA has been whimsically termed *northern blotting*. A further play on words accounts for the term *western blotting,* which refers to a technique for detecting a particular protein by staining with specific antibody (Section 3.3). Southern, northern, and western blots are also known respectively as *DNA, RNA,* and *protein blots*.

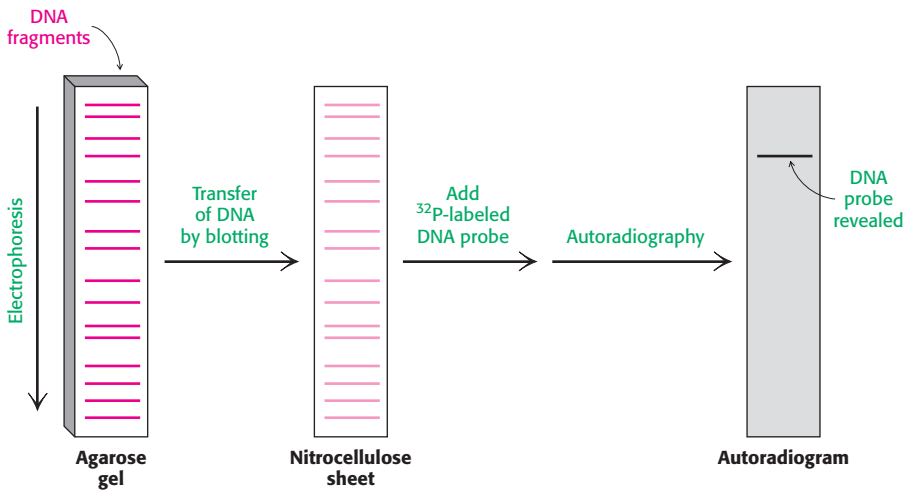
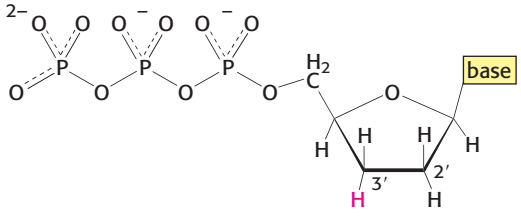


Figure 5.3 Southern blotting. A DNA fragment containing a specific sequence can be identified by separating a mixture of fragments by electrophoresis, transferring them to nitrocellulose, and hybridizing with a ^{32}P-labeled probe complementary to the sequence. The fragment containing the sequence is then visualized by autoradiography.

DNA can be sequenced by controlled termination of replication

The analysis of DNA structure and its role in gene expression also have been markedly facilitated by the development of powerful techniques for the *sequencing* of DNA molecules. The key to DNA sequencing is the generation of DNA fragments whose length depends on the last base in the sequence. Collections of such fragments can be generated through the *controlled termination of replication* (Sanger dideoxy method), a method developed by Frederick Sanger and coworkers. This technique has superseded alternative methods because of its simplicity. The same procedure is performed on four reaction mixtures at the same time. In all these mixtures, a DNA polymerase is used to make the complement of a particular sequence within a single-stranded DNA molecule. The synthesis is primed by a chemically synthesized fragment that is complementary to a part of the sequence known from other studies. In addition to the four deoxyribonucleoside triphosphates (radioactively labeled), each reaction mixture contains a small amount of the 2′,3′-*dideoxy analog* of one of the nucleotides, a different nucleotide for each reaction mixture.



2′, 3′-Dideoxy analog

The incorporation of this analog blocks further growth of the new chain because it lacks the 3′-hydroxyl terminus needed to form the next phosphodiester bond. The concentration of the dideoxy analog is low enough that chain termination will take place only occasionally. The polymerase will insert the correct nucleotide sometimes and the dideoxy analog other times, stopping the reaction. For instance, if the dideoxy analog of dATP is present, fragments of various lengths are produced, but all will be terminated by the dideoxy analog (Figure 5.4). Importantly, this dideoxy analog of dATP will be inserted only where a T was located in the DNA being sequenced. Thus, the fragments of different length will correspond to the positions of T. Four such sets of *chain-terminated fragments* (one for each dideoxy analog) then undergo electrophoresis, and the base sequence of the new DNA is read from the autoradiogram of the four lanes.

Fluorescence detection is a highly effective alternative to autoradiography because it eliminates the use of radioactive reagents and can be readily automated. A fluorescent tag is incorporated into each dideoxy analog—a differently colored one for each of the four chain terminators (e.g., a blue emitter for termination at A and a red one for termination at C). With the use of a mixture of terminators, a single reaction can be performed and the resulting fragments are separated by a technique known as *capillary electrophoresis,* in which the mixture is passed through a very narrow tube at high voltage to achieve efficient separation within a short time. As the DNA fragments emerge from the capillary, they are detected by their fluorescence; the sequence of their colors directly gives the base sequence (Figure 5.5). Sequences of as many as 500 bases can be determined in

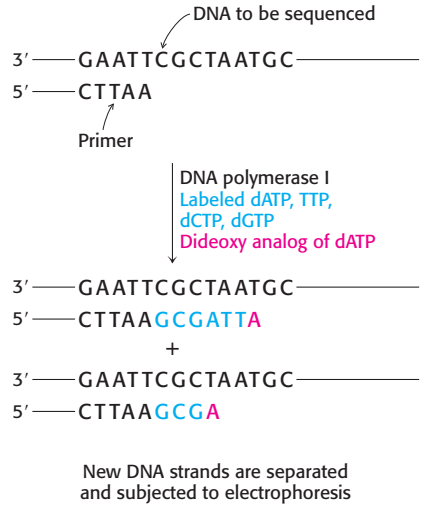


Figure 5.4 Strategy of the chain-termination method for sequencing DNA. Fragments are produced by adding the 2′,3′-dideoxy analog of a dNTP to each of four polymerization mixtures. For example, the addition of the dideoxy analog of dATP (shown in red) results in fragments ending in A. The strand cannot be extended past the dideoxy analog.

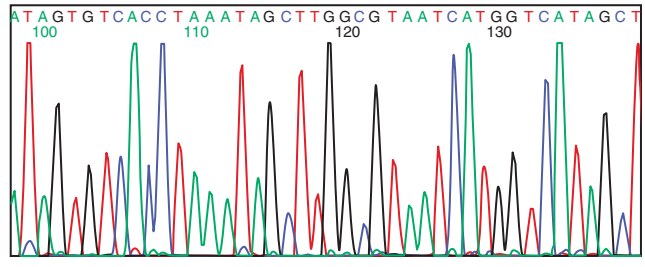
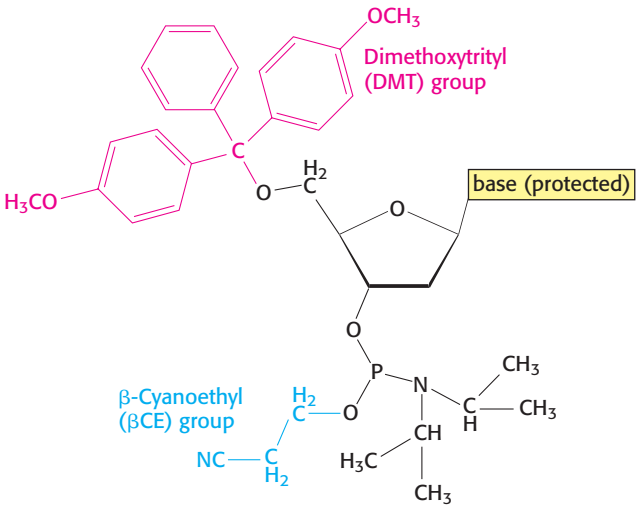


Figure 5.5 Fluorescence detection of oligonucleotide fragments produced by the dideoxy method. A sequencing reaction is performed with four chain-terminating dideoxy nucleotides, each labeled with a tag that fluoresces at a different wavelength (e.g., red for T). Each of the four colors represents a different base in a chromatographic trace produced by fluorescence measurements at four wavelengths. [After A. J. F. Griffiths et al., *An Introduction to Genetic Analysis,* 8th ed. (W. H. Freeman and Company, 2005).]

this way. Indeed, modern DNA-sequencing instruments can sequence more than 1 million bases per day with the use of this method.

DNA probes and genes can be synthesized by automated solid-phase methods



A deoxyribonucleoside 3′-phosphoramidite with DMT and βCE attached

DNA strands, like polypeptides (Section 3.4), can be synthesized by the sequential addition of activated monomers to a growing chain that is linked to an insoluble support. The activated monomers are protected *deoxyribonucleoside 3′-phosphoramidites*. In step 1, the 3′-phosphorus atom of this incoming unit becomes joined to the 5′-oxygen atom of the growing chain to form a *phosphite triester* (Figure 5.6). The 5′-OH group of the activated monomer is unreactive because it is blocked by a dimethoxytrityl (DMT) protecting group, and the 3′-phosphoryl group is rendered unreactive by attachment of the β-cyanoethyl (βCE) group. Likewise, amino groups on the purine and pyrimidine bases are blocked.

Coupling is carried out under anhydrous conditions because water reacts with phosphoramidites. In step 2, the phosphite triester (in which P is trivalent) is oxidized by iodine to form a *phosphotriester* (in which P is pentavalent). In step 3, the DMT protecting group on the 5′-OH group of the growing chain is removed by the addition of dichloroacetic acid, which leaves other protecting groups intact. The DNA chain is now elongated by one unit and ready for another cycle of addition. Each cycle takes only about 10 minutes and usually elongates more than 99% of the chains.

This solid-phase approach is ideal for the synthesis of DNA, as it is for polypeptides, because the desired product stays on the insoluble support

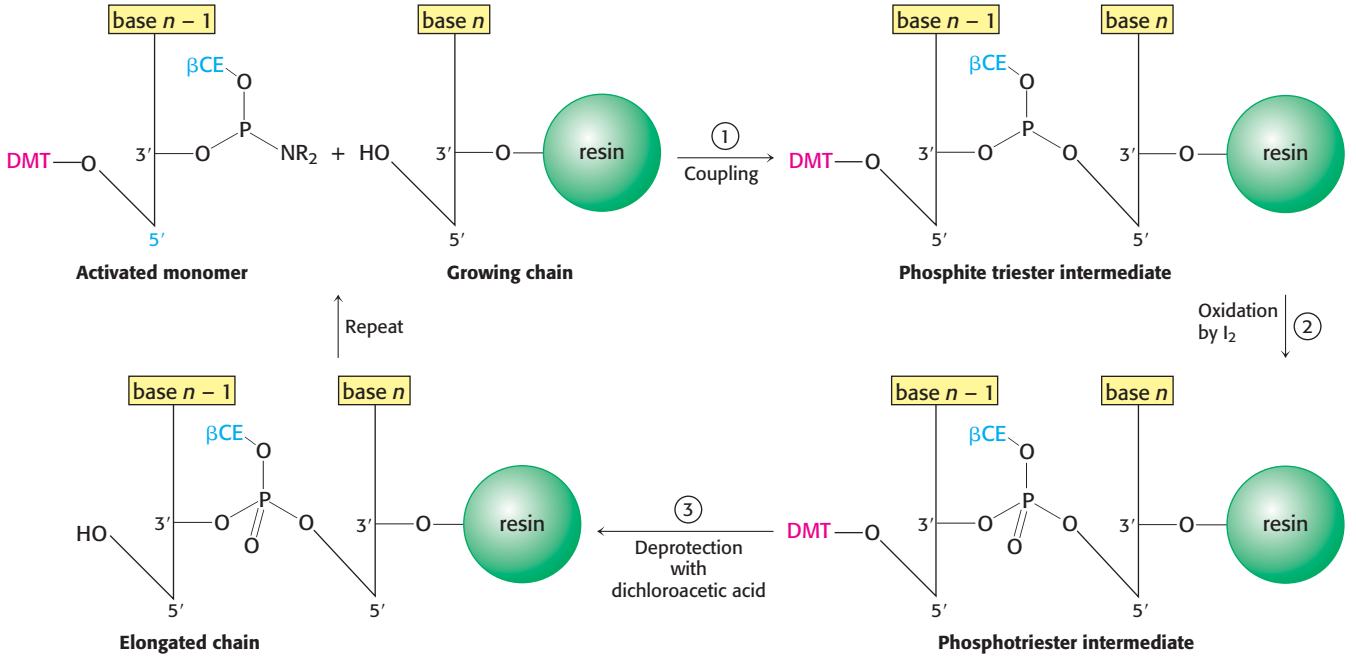


Figure 5.6 Solid-phase synthesis of a DNA chain by the phosphite triester method. The activated monomer added to the growing chain is a deoxyribonucleoside 3′-phosphoramidite containing a dimethoxytrityl (DMT) protecting group on its 5′-oxygen atom, a β-cyanoethyl (βCE) protecting group on its 3′-phosphoryl oxygen atom, and a protecting group on the base.

until the final release step. All the reactions take place in a single vessel, and excess soluble reagents can be added to drive reactions to completion. At the end of each step, soluble reagents and by-products are washed away from the resin that bears the growing chains. At the end of the synthesis, NH_3 is added to remove all protecting groups and release the oligonucleotide from the solid support. Because elongation is never 100% complete, the new DNA chains are of diverse lengths—the desired chain is the longest one. The sample can be purified by high-pressure liquid chromatography or by electrophoresis on polyacrylamide gels. DNA chains of as many as 100 nucleotides can be readily synthesized by this automated method.

The ability to rapidly synthesize DNA chains of any selected sequence opens many experimental avenues. For example, a synthesized oligonucleotide labeled at one end with ^{32}P or a fluorescent tag can be used to search for a complementary sequence in a very long DNA molecule or even in a genome consisting of many chromosomes. The use of labeled oligonucleotides as DNA probes is powerful and general. For example, a DNA probe that can base-pair to a known complementary sequence in a chromosome can serve as the starting point of an exploration of adjacent uncharted DNA. Such a probe can be used as a *primer* to initiate the replication of neighboring DNA by DNA polymerase. An exciting application of the solid-phase approach is the synthesis of new tailor-made genes. New proteins with novel properties can now be produced in abundance by the expression of synthetic genes. Finally, the synthetic scheme heretofore described can be slightly modified for the solid-phase synthesis of RNA oligonucleotides, which can be very powerful reagents for the degradation of specific mRNA molecules in living cells by a technique known as RNA interference (Section 5.4).

Selected DNA sequences can be greatly amplified by the polymerase chain reaction

In 1984, Kary Mullis devised an ingenious method called the *polymerase chain reaction* (PCR) for amplifying specific DNA sequences. Consider a DNA duplex consisting of a target sequence surrounded by nontarget DNA. Millions of copies of the target sequences can be readily obtained by PCR if the flanking sequences of the target are known. PCR is carried out by adding the following components to a solution containing the target sequence: (1) a pair of primers that hybridize with the flanking sequences of the target, (2) all four deoxyribonucleoside triphosphates (dNTPs), and (3) a heat-stable DNA polymerase. A PCR cycle consists of three steps (Figure 5.7).

1. *Strand Separation.* The two strands of the parent DNA molecule are separated by heating the solution to 95°C for 15 s.

2. *Hybridization of Primers.* The solution is then abruptly cooled to 54°C to allow each primer to hybridize to a DNA strand. One primer hybridizes to the 3′ end of the target on one strand, and the other primer hybridizes to the 3′ end on the complementary target strand. Parent DNA duplexes do not form, because the primers are present in large excess. Primers are typically from 20 to 30 nucleotides long.

3. *DNA Synthesis.* The solution is then heated to 72°C, the optimal temperature for heat-stable polymerases. One such enzyme is *Taq* DNA polymerase, which is derived from *Thermus aquaticus,* a thermophilic bacterium that lives in hot springs. The polymerase elongates both primers in the direction of the target sequence because DNA synthesis is in the 5′-to-3′

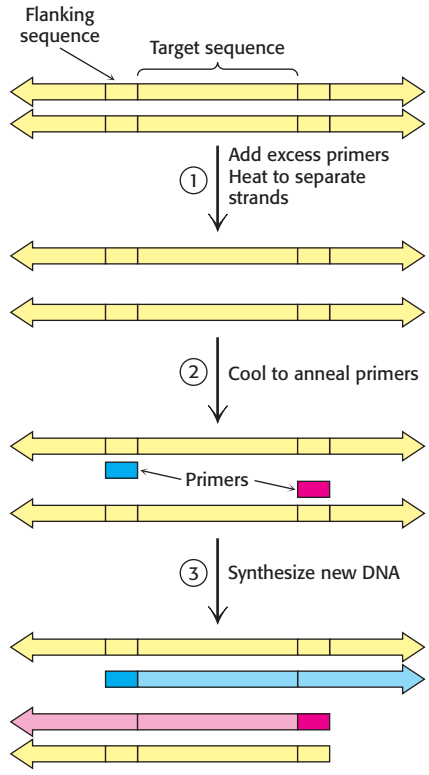


Figure 5.7 The first cycle in the polymerase chain reaction (PCR). A cycle consists of three steps: strand separation, the hybridization of primers, and the extension of primers by DNA synthesis.

direction. DNA synthesis takes place on both strands but extends beyond the target sequence.

These three steps—strand separation, hybridization of primers, and DNA synthesis—constitute one cycle of the PCR amplification and can be carried out repetitively just by changing the temperature of the reaction mixture. The thermostability of the polymerase makes it feasible to carry out PCR in a closed container; no reagents are added after the first cycle. At the completion of the second cycle, four duplexes containing the targeting sequence have been generated (Figure 5.8). Of the eight DNA strands comprising these duplexes, two short strands constitute only the target sequence—the sequence including and bounded by the primers. Subsequent cycles will amplify the target sequence exponentially. Ideally, after n cycles, the desired sequence is amplified 2^n-fold. The amplification is a millionfold after 20 cycles and a billionfold after 30 cycles, which can be carried out in less than an hour.

Several features of this remarkable method for amplifying DNA are noteworthy. First, the sequence of the target need not be known. All that is required is knowledge of the flanking sequences so that complementary primers can be synthesized. Second, the target can be much larger than the primers. Targets larger than 10 kb have been amplified by PCR. Third, primers do not have to be perfectly matched to flanking sequences to amplify targets. With the use of primers derived from a gene of known sequence, it is possible to search for variations on the theme. In this way, families of genes are being discovered by PCR. Fourth, PCR is highly specific because of the stringency of hybridization at relatively high temperature. *Stringency* is the required closeness of the match between primer and target, which can be controlled by temperature and salt. At high temperatures, only the DNA between hybridized primers is amplified. A gene constituting less than a millionth of the total DNA of a higher organism is accessible by PCR. Fifth, PCR is exquisitely sensitive. A single DNA molecule can be amplified and detected.

PCR is a powerful technique in medical diagnostics, forensics, and studies of molecular evolution

PCR can provide valuable diagnostic information in medicine. Bacteria and viruses can be readily detected with the use of specific primers. For example, PCR can reveal the presence of small amounts of DNA from the human immunodeficiency virus (HIV) in persons who have not yet mounted an immune response to this pathogen. In these patients, assays designed to detect antibodies against the virus would yield a false negative test result. Finding *Mycobacterium tuberculosis* bacilli in tissue specimens is slow and laborious. With PCR, as few as 10 tubercle bacilli per million human cells can be readily detected. PCR is a promising method for the early detection of certain cancers. This technique can identify mutations of certain growth-control genes, such as the *ras*

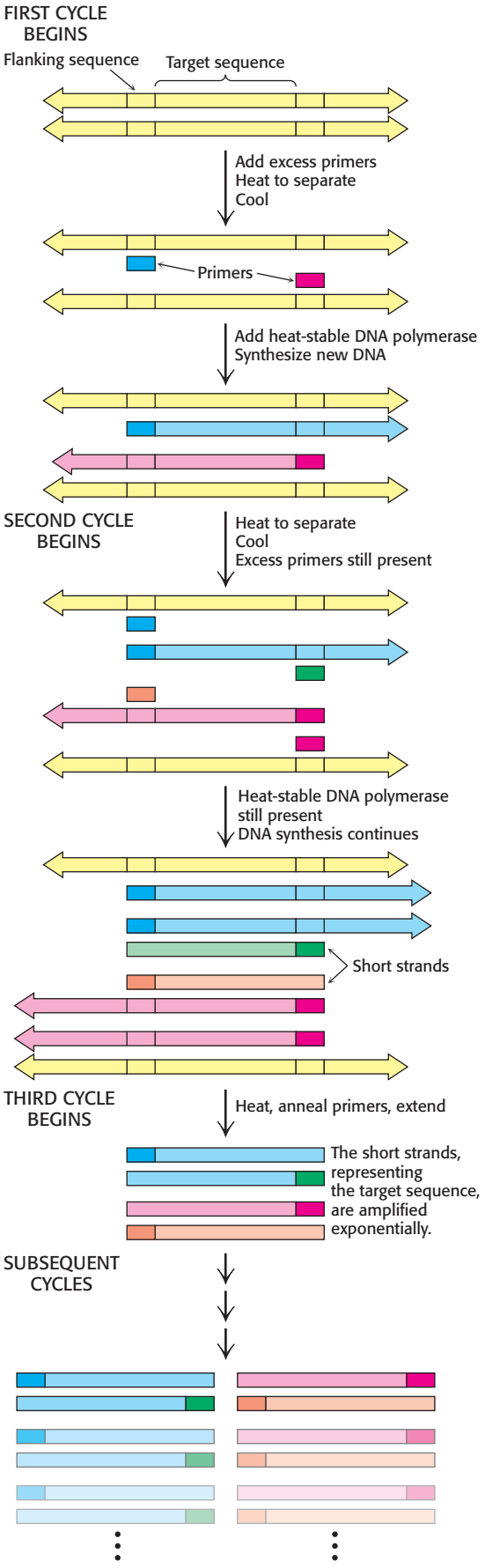


Figure 5.8 Multiple cycles of the polymerase chain reaction. The two short strands produced at the end of the third cycle (along with longer stands not shown) represent the target sequence. Subsequent cycles will amplify the target sequence exponentially and the parent sequence arithmetically.

genes (Chapter 14). The capacity to greatly amplify selected regions of DNA can also be highly informative in monitoring cancer chemotherapy. Tests using PCR can detect when cancerous cells have been eliminated and treatment can be stopped; they can also detect a relapse and the need to immediately resume treatment. PCR is ideal for detecting leukemias caused by chromosomal rearrangements.

PCR is also having an effect in forensics and legal medicine. An individual DNA profile is highly distinctive because many genetic loci are highly variable within a population. For example, variations at one specific location determines a person's HLA type (human leukocyte antigen type; Section 34.5); organ transplants are rejected when the HLA types of the donor and recipient are not sufficiently matched. PCR amplification of multiple genes is being used to establish biological parentage in disputed paternity and immigration cases. Analyses of blood stains and semen samples by PCR have implicated guilt or innocence in numerous assault and rape cases. The root of a single shed hair found at a crime scene contains enough DNA for typing by PCR (Figure 5.9).

DNA is a remarkably stable molecule, particularly when shielded from air, light, and water. Under such circumstances, large fragments of DNA can remain intact for thousands of years or longer. PCR provides an ideal method for amplifying such ancient DNA molecules so that they can be detected and characterized (Section 6.5). PCR can also be used to amplify DNA from microorganisms that have not yet been isolated and cultured. As will be discussed in Chapter 6, sequences from these PCR products can be sources of considerable insight into evolutionary relationships between organisms.

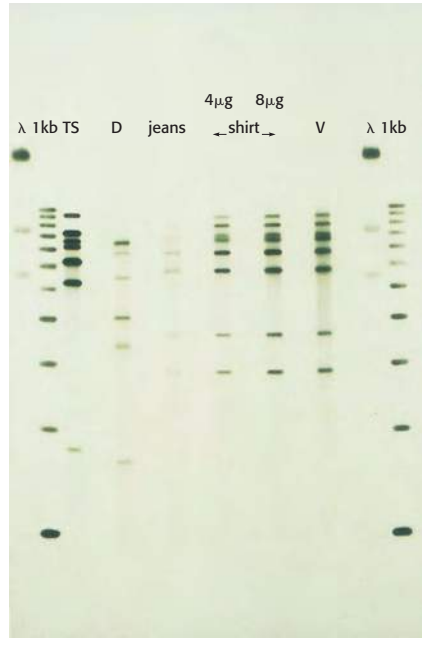


Figure 5.9 DNA and forensics. DNA isolated from bloodstains on the pants and shirt of a defendant was amplified by PCR, then compared with DNA from the victim as well as the defendant by using gel electrophoresis and autoradiography. DNA from the bloodstains on the defendant's clothing matched the pattern of the victim but not that of the defendant. The frequency of a coincidental match of the DNA pattern on the clothing and the victim is approximately 1 in 33 billion. Lanes λ, 1kb, and TS refer to control DNA samples; lane D, DNA from the defendant; jeans and shirt, DNA isolated from bloodstains on defendant's pants and shirt (two different amounts analyzed); V, DNA sample from victim's blood. [Courtesy of Cellmark Diagnostics, Germantown, Maryland.]

The tools for recombinant DNA technology have been used to identify disease-causing mutations

Let us consider how the techniques just described have been utilized in concert to study ALS, introduced at the beginning of this chapter. Five percent of all patients suffering from ALS have family members who also have been diagnosed with the disease. A heritable disease pattern is indicative of a strong genetic component of disease causation. To identify these disease-causing genetic alterations, researchers identify *polymorphisms* (instances of genetic variation) within an affected family that correlate with the emergence of disease. Polymorphisms may themselves cause disease or be closely linked to another genetic alteration that does. One class of polymorphisms are *restriction-fragment-length polymorphisms* (RFLPs), which are mutations within restriction sites that change the sizes of DNA fragments produced by the appropriate restriction enzyme. Using restriction digests and Southern blots of the DNA from members of ALS-affected families, researchers identified RFLPs that were found preferentially in those family members with a positive diagnosis. For some of these families, strong evidence was obtained for the disease-causing mutation within a specific region of chromosome 21.

After the probable location of one disease-causing gene had been identified, this same research group compared the locations of the ALS-associated RFLPs with the known sequence of chromosome 21. They noted that this chromosomal locus contains the *SOD1* gene, which encodes the Cu/Zn superoxide dismutase protein SOD1, an enzyme important for the protection of cells against oxidative damage (Section 18.3). PCR amplification of regions of the *SOD1* gene from the DNA of affected family members, followed by Sanger dideoxy sequencing of the targeted fragment, enabled the identification of 11 disease-causing mutations from 13 different families.

This work was pivotal for focusing further inquiry into the roles that superoxide dismutase and its corresponding mutant forms play in the pathology of ALS.

5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology

The pioneering work of Paul Berg, Herbert Boyer, and Stanley Cohen in the early 1970s led to the development of recombinant DNA technology, which has taken biology from an exclusively analytical science to a synthetic one. New combinations of unrelated genes can be constructed in the laboratory by applying recombinant DNA techniques. These novel combinations can be cloned—amplified many-fold—by introducing them into suitable cells, where they are replicated by the DNA-synthesizing machinery of the host. The inserted genes are often transcribed and translated in their new setting. What is most striking is that the genetic endowment of the host can be permanently altered in a designed way.

Restriction enzymes and DNA ligase are key tools in forming recombinant DNA molecules

Let us begin by seeing how novel DNA molecules can be constructed in the laboratory. An essential tool for the manipulation of recombinant DNA is a *vector,* a DNA molecule that can replicate autonomously in an appropriate host organism. Vectors are designed to enable the rapid, covalent insertion of DNA fragments of interest. *Plasmids* (naturally occurring circles of DNA that act as accessory chromosomes in bacteria) and bacteriophage lambda (λ phage), a virus, are choice vectors for cloning in *E. coli.* The vector can be prepared for accepting a new DNA fragment by cleaving it at a single specific site with a restriction enzyme. For example, the plasmid pSC101, a 9.9-kb double-helical circular DNA molecule, is split at a unique site by the *Eco*RI restriction enzyme. The staggered cuts made by this enzyme produce *complementary single-stranded ends,* which have specific affinity for each other and hence are known as *cohesive* or *sticky ends*. Any DNA fragment can be inserted into this plasmid if it has the same cohesive ends. Such a fragment can be prepared from a larger piece of DNA by using the same restriction enzyme as was used to open the plasmid DNA (Figure 5.10).

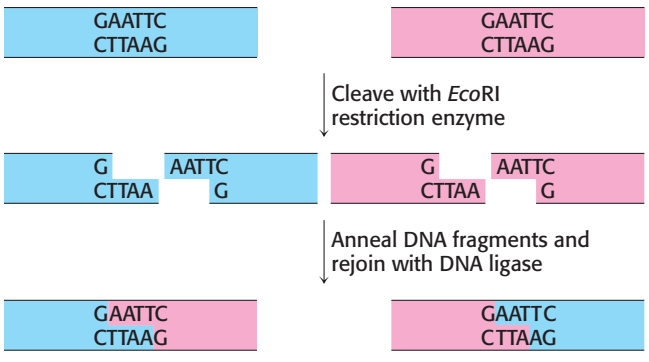


Figure 5.10 Joining of DNA molecules by the cohesive-end method. Two DNA molecules, cleaved with a common restriction enzyme such as *Eco*RI, can be ligated to form recombinant molecules.

The single-stranded ends of the fragment are then complementary to those of the cut plasmid. The DNA fragment and the cut plasmid can be annealed and then joined by *DNA ligase,* which catalyzes the formation of a phosphodiester bond at a break in a DNA chain. DNA ligase requires a free 3′-hydroxyl group and a 5′-phosphoryl group. Furthermore, the chains joined by ligase must be in a double helix. An energy source such as ATP or NAD^+ is required for the joining reaction, as will be discussed in Chapter 28.

What if the target DNA is not naturally flanked by the appropriate restriction sites? How is the fragment cut and annealed to the vector? The cohesive-end method for joining DNA molecules can still be used in these cases by adding a *short, chemically synthesized DNA linker* that can be cleaved by restriction enzymes. First, the linker is covalently joined to the ends of a DNA fragment. For example,

the 5′ ends of a decameric linker and a DNA molecule are phosphorylated by polynucleotide kinase and then joined by the ligase from T4 phage (Figure 5.11). This ligase can form a covalent bond between blunt-ended (flush-ended) double-helical DNA molecules. Cohesive ends are produced when these terminal extensions are cut by an appropriate restriction enzyme. Thus, *cohesive ends corresponding to a particular restriction enzyme can be added to virtually any DNA molecule*. We see here the fruits of combining enzymatic and synthetic chemical approaches in crafting new DNA molecules.

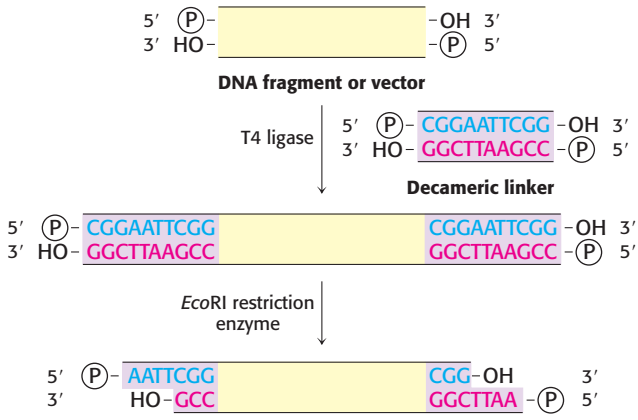


Figure 5.11 Formation of cohesive ends. Cohesive ends can be formed by the addition and cleavage of a chemically synthesized linker.

Plasmids and lambda phage are choice vectors for DNA cloning in bacteria

Many plasmids and bacteriophages have been ingeniously modified by researchers to enhance the delivery of recombinant DNA molecules into bacteria and to facilitate the selection of bacteria harboring these vectors. As already mentioned, plasmids are circular double-stranded DNA molecules that occur naturally in some bacteria. They range in size from two to several hundred kilobases. Plasmids carry genes for the inactivation of antibiotics, the production of toxins, and the breakdown of natural products. These accessory chromosomes can replicate independently of the host chromosome. In contrast with the host genome, they are dispensable under certain conditions. A bacterial cell may have no plasmids at all or it may house as many as 20 copies of a plasmid.

Many plasmids have been optimized for a particular experimental task. For example, one class of plasmids, known as *cloning vectors,* is particularly suitable for the rapid insertion and replication of a collection of DNA fragments. The creative placement of antibiotic-resistance genes or reporter genes or both within these plasmids enables the rapid identification of those vectors that harbor the desired DNA insert. For example, in *pBR322,* one of the first plasmids used for this purpose, insertion of DNA at the *Sal*I or *Bam*HI restriction site (Figure 5.12) inactivates the gene for tetracycline resistance, an effect called *insertional inactivation*. Cells containing pBR322 with a DNA insert at one of these restriction sites are resistant to ampicillin but sensitive to tetracycline, and so they can be readily selected. Another class of plasmids have been optimized for use as *expression vectors* for the production of large amounts of protein. In addition to antibiotic-resistance genes, they contain promoter sequences designed to drive the transcription of large amounts of a protein-coding DNA sequence. Often, these vectors contain sequences flanking the cloning site that simplify the addition of

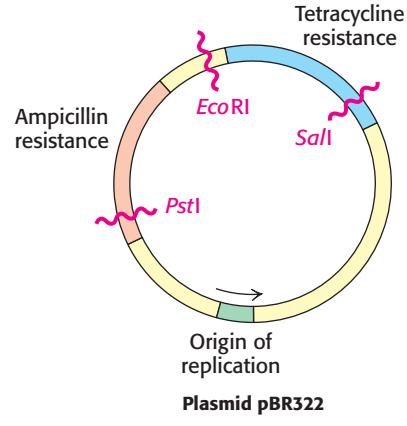


Figure 5.12 Genetic map of the plasmid pBR322. This plasmid carries two genes for antibiotic resistance. Like all other plasmids, it is a circular duplex DNA.

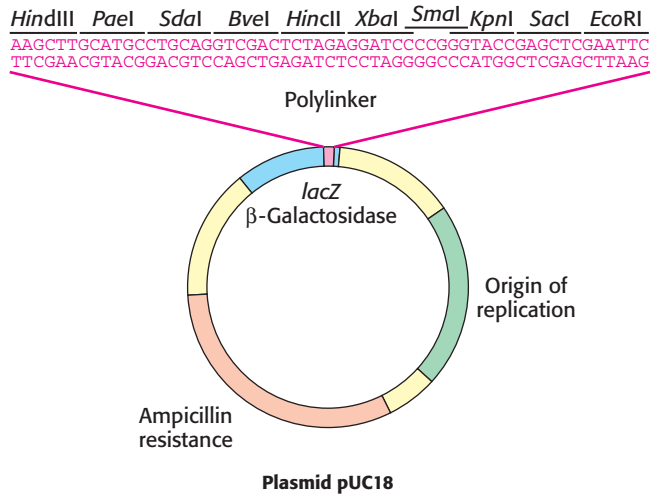


Figure 5.13 A polylinker in the plasmid pUC18. The plasmid pUC18 includes a polylinker within a gene for β-galactosidase (often called the *lacZ* gene). Insertion of a DNA fragment into one of the many restriction sites within this polylinker can be detected by the absence of β-galactosidase activity.

fusion tags to the protein of interest (Section 3.1), greatly facilitating the purification of the overexpressed protein. Both types of plasmid vectors often feature a *polylinker* region that includes many unique restriction sites within its sequence (Figure 5.13). This polylinker can be cleaved with many different restriction enzymes or combinations of enzymes, providing great versatility in the DNA fragments that can be inserted.

Another widely used vector, *λ phage,* enjoys a choice of life styles: this bacteriophage can destroy its host or it can become part of its host (Figure 5.14). In the *lytic pathway,* viral functions are fully expressed: viral DNA and proteins are quickly produced and packaged into virus particles, leading to the lysis (destruction) of the host cell and the sudden appearance of about 100 progeny virus particles, or *virions*. In the *lysogenic pathway,* the phage DNA becomes inserted into the host-cell genome and can be replicated together with host-cell DNA for many generations, remaining inactive. Certain environmental changes can trigger the expression of this dormant viral DNA, which leads to the formation of progeny viruses and lysis of the host. Large segments of the 48-kb DNA of λ phage are not essential for productive infection and can be replaced by foreign DNA, thus making λ phage an ideal vector.

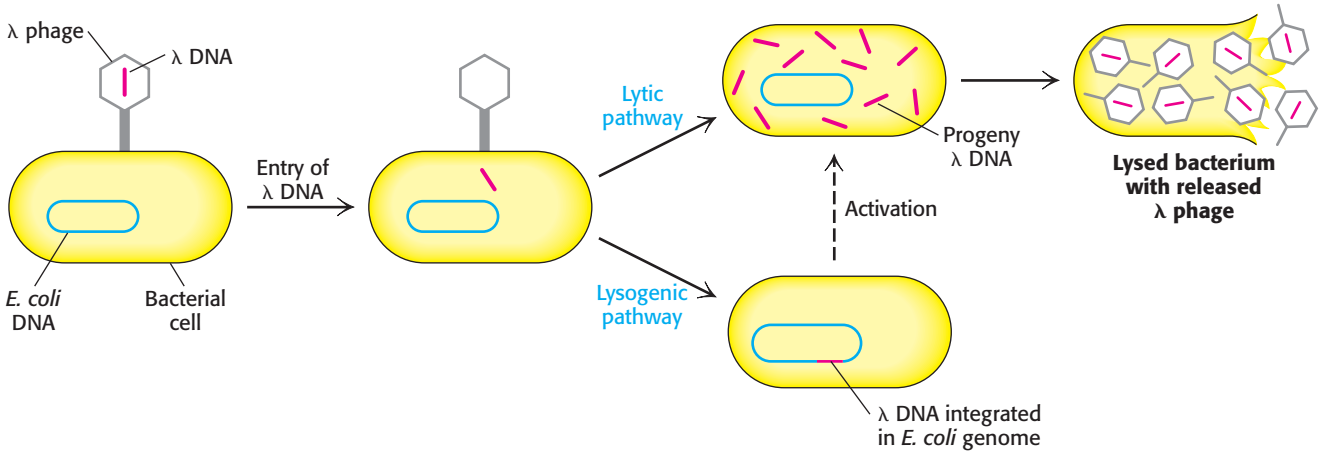


Figure 5.14 Alternative infection modes for λ phage. Lambda phage can multiply within a host and lyse it (lytic pathway) or its DNA can become integrated into the host genome (lysogenic pathway), where it is dormant until activated.

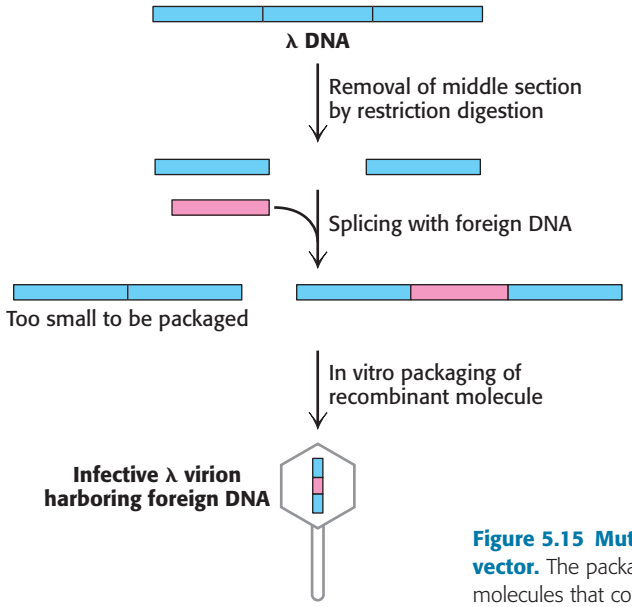


Figure 5.15 Mutant λ phage as a cloning vector. The packaging process selects DNA molecules that contain an insert.

Mutant λ phages designed for cloning have been constructed. An especially useful one called λgt-λβ contains only two *Eco*RI cleavage sites instead of the five normally present (Figure 5.15). After cleavage, the middle segment of this λ DNA molecule can be removed. The two remaining pieces of DNA (called arms) have a combined length equal to 72% of a normal genome length. This amount of DNA is too little to be packaged into a λ particle, which can take up only DNA measuring from 78% to 105% of a normal genome. However, a suitably long DNA insert (such as 10 kb) between the two ends of λ DNA enables such a recombinant DNA molecule (93% of normal length) to be packaged. Nearly all infectious λ particles formed in this way will contain an inserted piece of foreign DNA. Another advantage of using these modified viruses as vectors is that they enter bacteria much more easily than do plasmids. Among the variety of λ mutants that have been constructed for use as cloning vectors, one of them, called a *cosmid,* is essentially a hybrid of λ phage and a plasmid that can serve as a vector for large DNA inserts (as large as 45 kb).

Bacterial and yeast artificial chromosomes

Much larger pieces of DNA can be propagated in *bacterial artificial chromosomes* (BACs) or *yeast artificial chromosomes* (YACs). BACs are highly engineered versions of the *E. coli* fertility (F factor) that can include inserts as large as 300 kb. YACs contain a centromere, an *autonomously replicating sequence* (ARS, where replication begins), a pair of telomeres (normal ends of eukaryotic chromosomes), selectable marker genes, and a cloning site (Figure 5.16). Inserts as large as 1000 kb can be cloned into YAC vectors.

Specific genes can be cloned from digests of genomic DNA

Ingenious cloning and selection methods have made it possible to isolate small stretches of DNA in a genome containing more than 3×10^6 kb. The approach is to prepare a large collection (*library*) of DNA fragments and then to identify those members of the collection that have the gene of interest. Hence, to clone a gene that is present just once in an entire genome, two critical components must be available: a specific oligonucleotide probe for the gene of interest and a DNA library that can be screened rapidly.

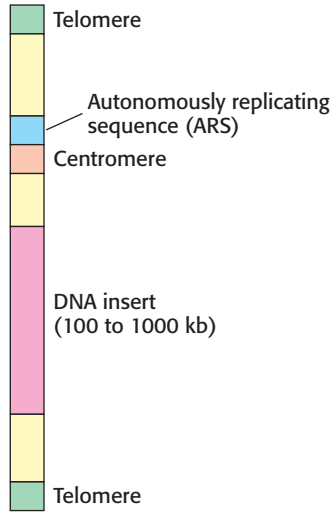


Figure 5.16 Diagram of a yeast artificial chromosome (YAC). These vectors include features necessary for replication and stability in yeast cells.

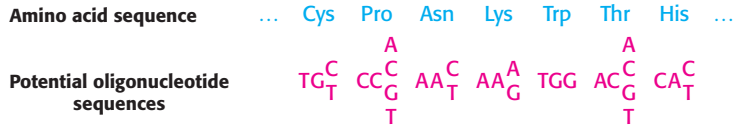


Figure 5.17 Probes generated from a protein sequence. A probe can be generated by synthesizing all possible oligonucleotides encoding a particular sequence of amino acids. Because of the degeneracy of the genetic code, 256 distinct oligonucleotides must be synthesized to ensure that the probe matching the sequence of seven amino acids in this example is present.

How is a specific probe obtained? In one approach, a probe for a gene can be prepared if a part of the amino acid sequence of the protein encoded by the gene is known. Peptide sequencing of a purified protein (Chapter 3) or knowledge of the sequence of a homologous protein from a related species (Chapter 6) are two potential sources of such information. However, a problem arises because a single peptide sequence can be encoded by a number of different oligonucleotides (Figure 5.17). Thus, for this purpose, peptide sequences containing tryptophan and methionine are preferred, because these amino acids are specified by a single codon, whereas other amino acid residues have between two and six codons (see Table 4.5). All the DNA sequences (or their complements) that encode the selected peptide sequence are synthesized by the solid-phase method and made radioactive by phosphorylating their 5′ ends with ^{32}P.

Alternatively, probes can be obtained from the corresponding mRNA from cells in which it is abundant. For example, precursors of red blood cells contain large amounts of mRNA for hemoglobin, and plasma cells are rich in mRNAs for antibody molecules. The mRNAs from these cells can be fractionated by size to enrich for the mRNA of interest. As will be described shortly, a DNA complementary to this mRNA can be synthesized in vitro and cloned to produce a highly specific probe.

To prepare the DNA library, a sample containing many copies of total genomic DNA is first mechanically sheared or partly digested by restriction enzymes into large fragments (Figure 5.18). This process yields a nearly random population of overlapping DNA fragments. These fragments are then separated by gel electrophoresis to isolate the set of all fragments that are about 15 kb long. Synthetic linkers are attached to the ends of these frag-

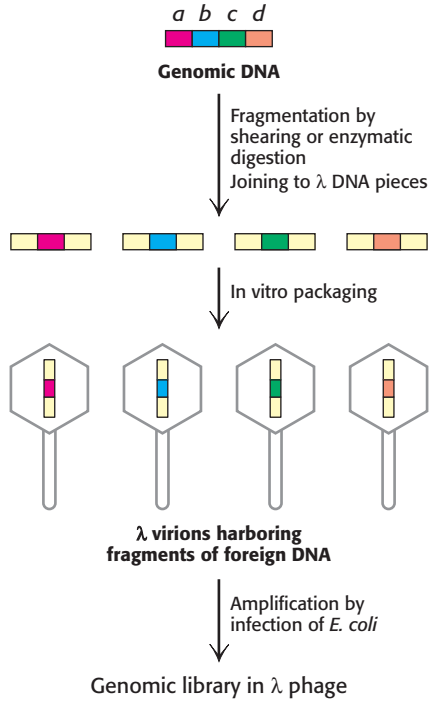


Figure 5.18 Creation of a genomic library. A genomic library can be created from a digest of a whole complex genome. On fragmentation of the genomic DNA into overlapping segments, the DNA is inserted into the λ phage vector (shown in yellow). Packaging into virions and amplification by infection in *E. coli* yields a genomic library.

ments, cohesive ends are formed, and the fragments are then inserted into a vector, such as λ phage DNA, prepared with the same cohesive ends. *E. coli* bacteria are then infected with these recombinant phages. These phages replicate themselves and then lyse their bacterial hosts. The resulting lysate contains fragments of human DNA housed in a sufficiently large number of virus particles to ensure that nearly the entire genome is represented. These phages constitute a *genomic library*. Phages can be propagated indefinitely, and so the library can be used repeatedly over long periods.

This genomic library is then screened to find the very small number of phages harboring the gene of interest. For the human genome, a calculation shows that a 99% probability of success requires screening about 500,000 clones; hence, a very rapid and efficient screening process is essential. Rapid screening can be accomplished by DNA hybridization.

A dilute suspension of the recombinant phages is first plated on a lawn of bacteria (Figure 5.19). Where each phage particle has landed and infected a bacterium, a *plaque* containing identical phages develops on the plate. A replica of this master plate is then made by applying a sheet of nitrocellulose. Infected bacteria and phage DNA released from lysed cells adhere to the sheet in a pattern of spots corresponding to the plaques. Intact bacteria on this sheet are lysed with NaOH, which also serves to denature the DNA so that it becomes accessible for hybridization with a ^{32}P-labeled probe. The presence of a specific DNA sequence in a single spot on the replica can be detected by using a radioactive complementary DNA or RNA molecule as a probe. Autoradiography then reveals the positions of spots harboring recombinant DNA. The corresponding plaques are picked out of the intact master plate and grown. A single investigator can readily screen a million clones in a day. This method makes it possible to isolate virtually any gene, provided that a probe is available.

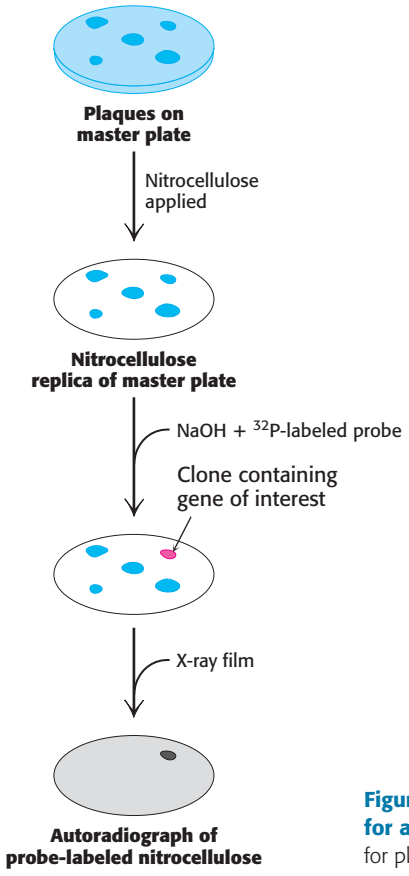


Figure 5.19 Screening a genomic library for a specific gene. Here, a plate is tested for plaques containing gene *a* of Figure 5.18.

Complementary DNA prepared from mRNA can be expressed in host cells

The preparation of eukaryotic DNA libraries presents unique challenges, especially if the researcher is interested primarily in the protein-coding region of a particular gene. Recall that most mammalian genes are mosaics of introns and exons. These interrupted genes cannot be expressed by bacteria, which lack the machinery to splice introns out of the primary transcript. However, this difficulty can be circumvented by causing bacteria to take up recombinant DNA that is complementary to mRNA, where the intronic sequences have been removed.

The key to forming *complementary DNA* is the enzyme *reverse transcriptase*. As discussed in Section 4.3, a retrovirus uses this enzyme to form a DNA–RNA hybrid in replicating its genomic RNA. Reverse transcriptase synthesizes a DNA strand complementary to an RNA template if the transcriptase is provided with a DNA primer that is base-paired to the RNA and contains a free 3′-OH group. We can use a simple sequence of linked thymidine [oligo(T)] residues as the primer. This oligo(T) sequence pairs with the poly(A) sequence at the 3′ end of most eukaryotic mRNA molecules (Section 4.4), as shown in Figure 5.20. The reverse transcriptase then synthesizes the rest of the cDNA strand in the presence of the four deoxyribonucleoside triphosphates. The RNA strand of this RNA–DNA hybrid is subsequently hydrolyzed by raising the pH. Unlike RNA, DNA is resistant to alkaline hydrolysis. The single-stranded DNA is converted into double-stranded DNA by creating another primer site. The enzyme *terminal transferase* adds nucleotides—for instance, several residues of dG—to the 3′ end of DNA. Oligo(dC) can bind to dG residues and prime the synthesis of the second DNA strand. Synthetic linkers can be added to this double-helical DNA for ligation to a suitable vector. Complementary DNA for all mRNA that a cell contains can be made, inserted into vectors, and then inserted into bacteria. Such a collection is called a *cDNA library*.

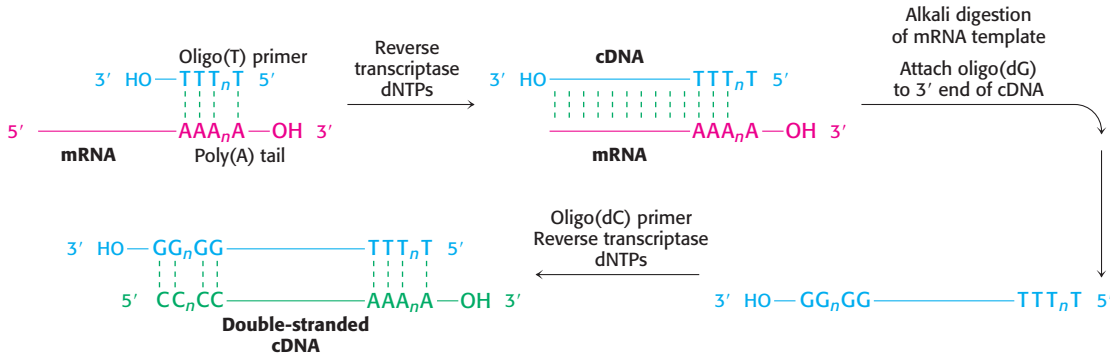


Figure 5.20 Formation of a cDNA duplex. A complementary DNA (cDNA) duplex is created from mRNA by using reverse transcriptase to synthesize a cDNA strand, first along the mRNA template and then, after digestion of the mRNA, along that same newly synthesized cDNA strand.

Complementary DNA molecules can be inserted into expression vectors to enable the production of the corresponding protein of interest. Clones of cDNA can be screened on the basis of their capacity to direct the synthesis of a foreign protein in bacteria, a technique referred to as *expression cloning*. A radioactive antibody specific for the protein of interest can be used to identify colonies of bacteria that express the corresponding protein product (Figure 5.21). As described earlier, spots of bacteria on a replica plate are lysed to release proteins, which bind to an applied nitrocellulose filter. With the addition of ^{125}I-labeled antibody specific for the protein of interest, autoradiography reveals the location of the desired colonies on the master

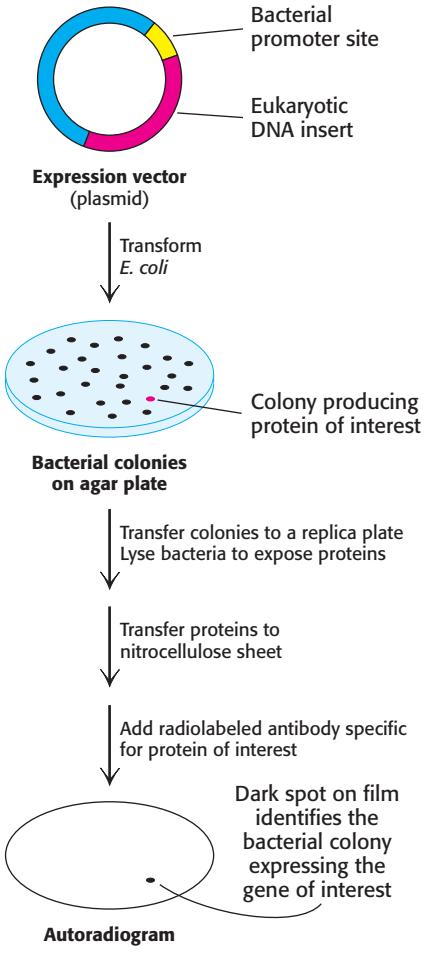


Figure 5.21 Screening of cDNA clones. A method of screening for cDNA clones is to identify expressed products by staining with specific antibody.

plate. This immunochemical screening approach can be used whenever a protein is expressed and corresponding antibody is available.

Complementary DNA has many applications beyond the generation of genetic libraries. The overproduction and purification of most eukaryotic proteins in prokaryotic cells necessitates the insertion of cDNA into plasmid vectors. For example, proinsulin, a precursor of insulin, is synthesized by bacteria harboring plasmids that contain DNA complementary to mRNA for proinsulin (Figure 5.22). Indeed, bacteria produce much of the insulin used today by millions of diabetics.

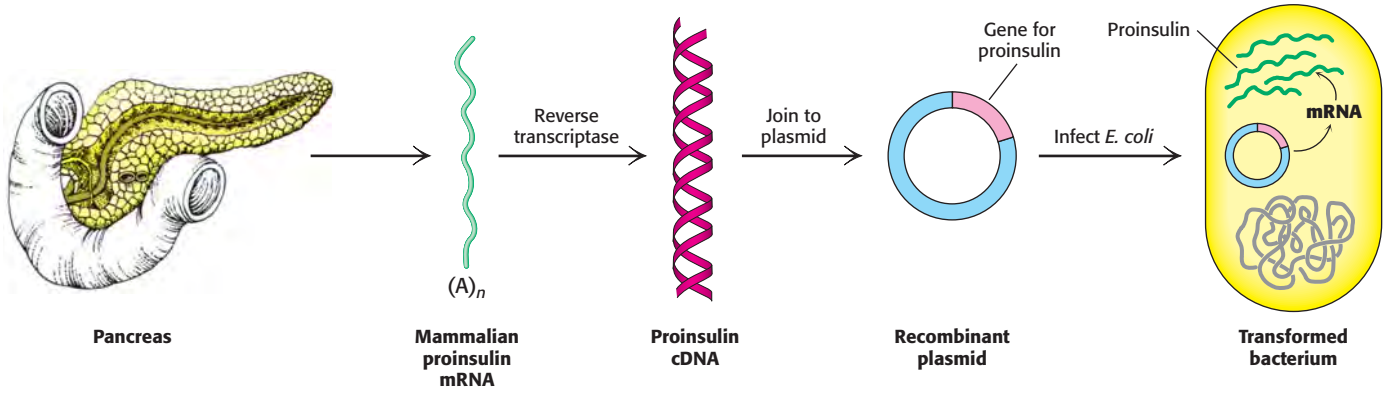


Figure 5.22 Synthesis of proinsulin by bacteria. Proinsulin, a precursor of insulin, can be synthesized by transformed (genetically altered) clones of *E. coli.* The clones contain the mammalian proinsulin gene.

Proteins with new functions can be created through directed changes in DNA

Much has been learned about genes and proteins by analyzing the effects that mutations have on their structure and function. In the classic genetic approach, mutations are generated randomly throughout the genome of a host organism, and those individuals exhibiting a phenotype of interest are selected. Analysis of these mutants then reveals which genes are altered, and DNA sequencing identifies the precise nature of the changes. *Recombinant DNA technology now makes the creation of specific mutations feasible in vitro.* We can construct new genes with designed properties by making three kinds of directed changes: *deletions, insertions,* and *substitutions*.

Deletions. A specific deletion can be produced by cleaving a plasmid at two sites with a restriction enzyme and ligating to form a smaller circle. This simple approach usually removes a large block of DNA. A smaller deletion can be made by cutting a plasmid at a single site. The ends of the linear DNA are then digested by an exonuclease that removes nucleotides from both strands. The shortened piece of DNA is then ligated to form a circle that is missing a short length of DNA about the restriction site.

Substitutions: oligonucleotide-directed mutagenesis. Mutant proteins with single amino acid substitutions can be readily produced by *oligonucleotide-directed mutagenesis* (Figure 5.23). Suppose that we want to replace a particular serine residue with cysteine. This mutation can be made if (1) we have a plasmid containing the gene or cDNA for the protein and (2) we know the base sequence around the site to be altered. If the serine of interest is encoded by TCT, mutation of the central base from C to G yields the TGT codon, which encodes cysteine. This type of mutation is called a *point mutation* because only one base is altered. To introduce this mutation into our plasmid, we prepare an oligonucleotide primer that is complementary to this region of the gene except that it contains TGT instead of TCT. The two strands of the plasmid are separated, and the primer is then annealed to the complementary strand. The mismatch of 1 of 15 base pairs is tolerable if the annealing is carried out at an appropriate temperature. After annealing to the complementary strand, the primer is elongated by DNA polymerase, and the double-stranded circle is closed by adding DNA ligase. Subsequent replication of this duplex yields two kinds of progeny plasmid, half with the original TCT sequence and half with the mutant TGT sequence. Expression of the plasmid containing the new TGT sequence will produce a protein with the desired substitution of cysteine for serine at a unique site. We will encounter many examples of the use of oligonucleotide-directed mutagenesis to precisely alter regulatory regions of genes and to produce proteins with tailor-made features.

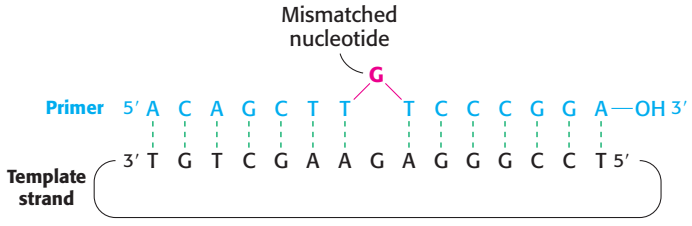


Figure 5.23 Oligonucleotide-directed mutagenesis. A primer containing a mismatched nucleotide is used to produce a desired change in the DNA sequence.

Insertions: cassette mutagenesis. In *cassette mutagenesis,* a variety of mutations, including insertions, deletions, and multiple point mutations, can be introduced into the gene of interest. A plasmid harboring the original gene is cut with a pair of restriction enzymes to remove a short segment (Figure 5.24). A synthetic double-stranded oligonucleotide (the *cassette*) carrying the genetic alterations of interest is prepared with cohesive ends that are complementary to the ends of the cut plasmid. Ligation of the cassette into the plasmid yields the desired mutated gene product.

Designer genes. Novel proteins can also be created by splicing together gene segments that encode domains that are not associated in nature. For example, a gene for an antibody can be joined to a gene for a toxin to produce a chimeric protein that kills cells that are recognized by the antibody. These *immunotoxins* are being evaluated as anticancer agents. Furthermore, noninfectious coat proteins of viruses can be produced in large amounts by recombinant DNA methods. They can serve as *synthetic vaccines* that are safer than conventional vaccines prepared by inactivating pathogenic viruses. A subunit of the hepatitis B virus produced in yeast is proving to be an effective vaccine against this debilitating viral disease. Finally, entirely new genes can be synthesized de novo by the solid-phase method. These genes can encode proteins with no known counterparts in nature.

Recombinant methods enable the exploration of the functional effects of disease-causing mutations

The application of recombinant DNA technology to the production of mutated proteins has had a significant effect in the study of ALS. Recall that genetic studies had identified a number of ALS-inducing mutations within the gene encoding Cu/Zn superoxide dismutase. As we shall learn in Section 18.3, SOD1 catalyzes the conversion of the superoxide radical anion into hydrogen peroxide, which, in turn, is converted into molecular oxygen and water by catalase. To study the potential effect of ALS-causing mutations on SOD1 structure and function, the *SOD1* gene was isolated from a human cDNA library by PCR amplification. The amplified fragments containing the gene were then digested with by an appropriate restriction enzyme and inserted into a similarly digested plasmid vector. Mutations corresponding to those observed in ALS patients were introduced into these plasmids by oligonucleotide-directed mutagenesis and the protein products were expressed and assayed for their catalytic activity. Surprisingly, these mutations did not significantly alter the enzymatic activity of the corresponding recombinant proteins. These observations have led to the prevailing notion that these mutations impart toxic properties to SOD1. Although the nature of this toxicity is not yet completely understood, one hypothesis is that mutant SOD1 is prone to form toxic aggregates in the cytoplasm of neuronal cells.

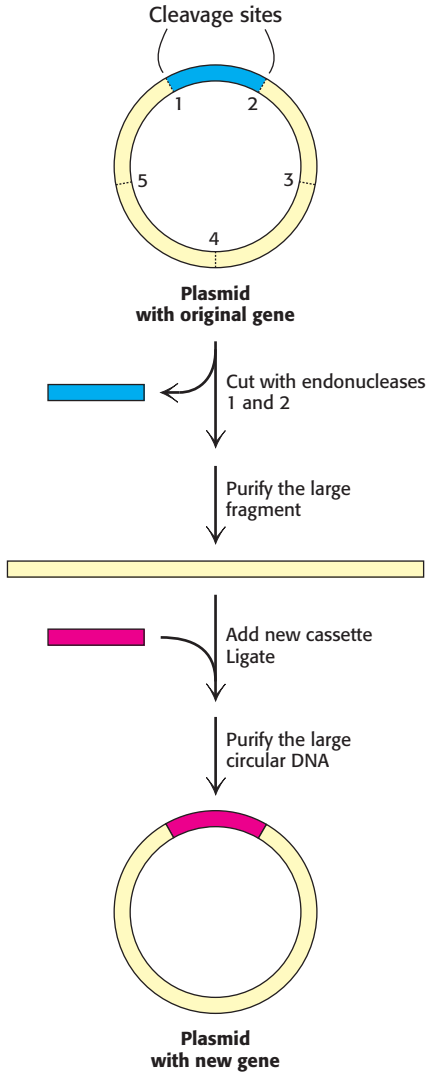


Figure 5.24 Cassette mutagenesis. DNA is cleaved at a pair of unique restriction sites by two different restriction endonucleases. A synthetic oligonucleotide with ends that are complementary to these sites (the cassette) is then ligated to the cleaved DNA. The method is highly versatile because the inserted DNA can have any desired sequence.

5.3 Complete Genomes Have Been Sequenced and Analyzed

The methods just described are extremely effective for the isolation and characterization of fragments of DNA. However, the genomes of organisms ranging from viruses to human beings contain longer sequences of DNA, arranged in very specific ways crucial for their integrated functions. Is it possible to sequence complete genomes and analyze them? For small genomes, this sequencing was accomplished soon after DNA-sequencing

methods were developed. Sanger and his coworkers determined the complete sequence of the 5,386 bases in the DNA of the ϕX174 DNA virus in 1977, just a quarter century after Sanger's pioneering elucidation of the amino acid sequence of a protein. This tour de force was followed several years later by the determination of the sequence of human mitochondrial DNA, a double-stranded circular DNA molecule containing 16,569 base pairs. It encodes 2 ribosomal RNAs, 22 transfer RNAs, and 13 proteins. Many other viral genomes were sequenced in subsequent years. However, the genomes of free-living organisms presented a great challenge because even the simplest comprises more than 1 million base pairs. Thus, sequencing projects require both rapid sequencing techniques and efficient methods for assembling many short stretches of 300 to 500 base pairs into a complete sequence.

The genomes of organisms ranging from bacteria to multicellular eukaryotes have been sequenced

With the development of automatic DNA sequencers based on fluorescent dideoxynucleotide chain terminators, high-volume, rapid DNA sequencing became a reality. The genome sequence of the bacterium *Haemophilus influenzae* was determined in 1995 by using a "shotgun" approach. The genomic DNA was sheared randomly into fragments that were then sequenced. Computer programs assembled the complete sequence by matching up overlapping regions between fragments. The *H. influenzae* genome comprises 1,830,137 base pairs and encodes approximately 1,740 proteins (Figure 5.25). Using similar approaches, investigators have determined the sequences of more than 100 bacterial and archaeal species including key model organisms such as *E. coli, Salmonella typhimurium,* and *Archaeoglobus fulgidus,* as well as pathogenic organisms such as *Yersina pestis* (causing bubonic plague) and *Bacillus anthracis* (anthrax).

The first eukaryotic genome to be completely sequenced was that of baker's yeast, *Saccharomyces cerevisiae,* in 1996. The yeast genome comprises approximately 12 million base pairs, distributed on 16 chromosomes, and encodes more than 6,000 proteins. This achievement was followed in 1998 by the first complete sequencing of the genome of a multicellular

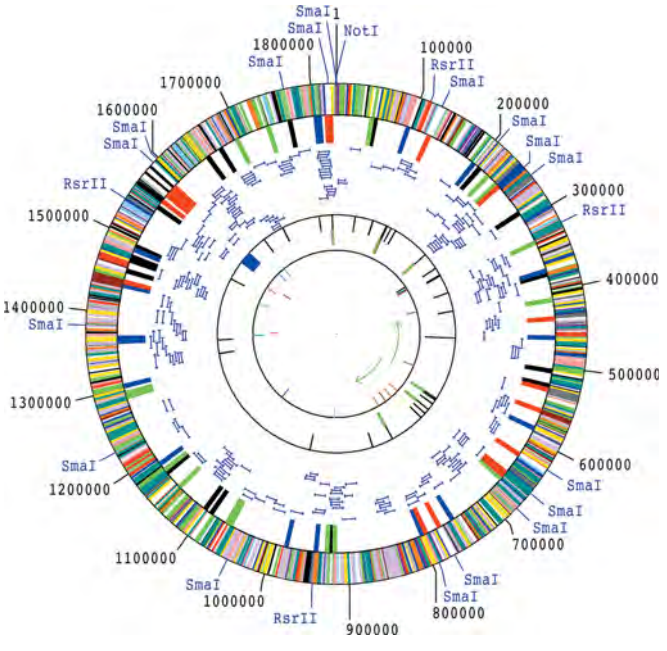


Figure 5.25 A complete genome. The diagram depicts the genome of *Haemophilus influenzae,* the first complete genome of a free-living organism to be sequenced. The genome encodes more than 1700 proteins and 70 RNA molecules. The likely function of approximately one-half of the proteins was determined by comparisons with sequences of proteins already characterized in other species. [From R. D. Fleischmann et al., *Science* 269:496–512, 1995; scan courtesy of The Institute for Genomic Research.]

organism, the nematode *Caenorhabditis elegans,* which contains 97 million base pairs. This genome includes more than 19,000 genes. The genomes of many additional organisms widely used in biological and biomedical research have now been sequenced, including those of the fruit fly *Drosophila melanogaster,* the model plant *Arabidopsis thaliana,* the mouse, the rat, and the dog. Note that the sequencing of a complex genome proceeds in various stages from "draft" through "completed" to "finished." Even after a sequence has been declared "finished," some sections, such as the repetitive sequences that make up heterochromatin, may be missing because these DNA sequences are very difficult to manipulate with the use of standard techniques.

The sequencing of the human genome has been finished

The ultimate goal of much of genomics research has been the sequencing and analysis of the human genome. Given that the human genome comprises approximately 3 billion base pairs of DNA distributed among 24 chromosomes, the challenge of producing a complete sequence was daunting. However, through an organized international effort of academic laboratories and private companies, the human genome has now progressed from a draft sequence first reported in 2001 to a finished sequence reported in late 2004 (Figure 5.26).

The human genome is a rich source of information about many aspects of humanity including biochemistry and evolution. Analysis of the genome will continue for many years to come. Developing an inventory of protein-encoding genes is one of the first tasks. At the beginning of the genome-sequencing project, the number of such genes was estimated to be approximately 100,000. With the availability of the completed (but not finished) genome, this estimate was reduced to between 30,000 and 35,000. With the finished sequence, the estimate fell to 20,000 to 25,000. We will use the estimate of 23,000 throughout this book. The reduction in this estimate is due, in part, to the realization that there are a large number of *pseudogenes,* many of which are formerly functional genes that have picked up mutations and are no longer expressed. For example, more than half of the genomic regions that correspond to olfactory receptors—key molecules responsible for our sense of smell—are pseudogenes (Section 33.1). The corresponding regions in the genomes of other primates and rodents encode functional olfactory receptors. Nonetheless, the surprisingly small number of genes belies the complexity of the human proteome. *Many genes encode more than one protein through mechanisms such as alternative splicing of mRNA and posttranslational modifications of proteins.* The different proteins encoded by a single gene often display important variations in functional properties.

The human genome contains a large amount of DNA that does not encode proteins. A great challenge in modern biochemistry and genetics is to elucidate the roles of this noncoding DNA. Much of this DNA is present because of the existence of *mobile genetic elements.* These elements, related to retroviruses (Section 4.3), have inserted themselves throughout the genome over time. Most of these elements have accumulated mutations and

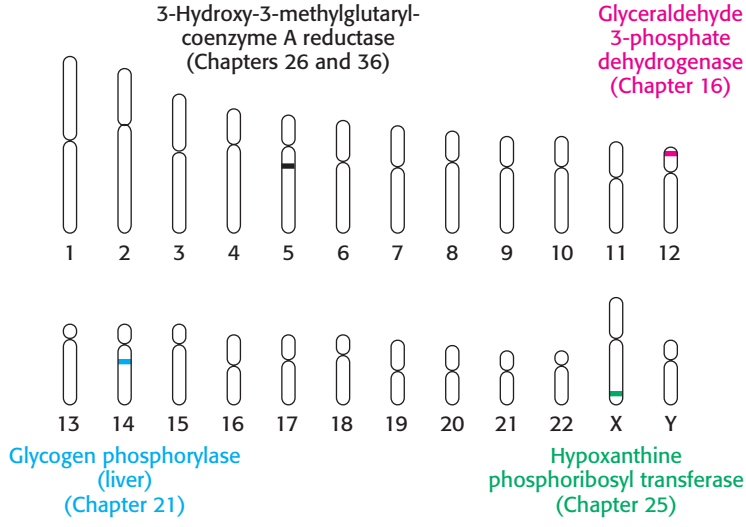


Figure 5.26 The human genome. The human genome is arrayed on 46 chromosomes—22 pairs of autosomes and the X and Y sex chromosomes. The locations of several genes associated with important pathways in biochemistry are highlighted.

are no longer functional. More than 1 million *Alu sequences,* each approximately 300 bases in length, are present in the human genome. *Alu* sequences are examples of *SINES, short interspersed elements.* The human genome also includes nearly 1 million *LINES, long interspersed elements,* DNA sequences that can be as long as 10 kilobase pairs (kbp). The roles of these elements as neutral genetic parasites or instruments of genome evolution are under current investigation.

"Next-generation" sequencing methods enable the rapid determination of a whole genome sequence

Since the introduction of Sanger dideoxy method in the mid-1970s, significant advances have been made in DNA-sequencing technologies, enabling the readout of progressively longer sequences with higher fidelity and shorter run times. The recent development of "next-generation" sequencing methods has extended this capability to formerly unforeseen levels. By combining technological breakthroughs in the handling of very small amounts of liquid, high-resolution optics, and computing power, these methods enable the parallel sequencing of more than 400,000 individual DNA fragments, at several hundred bases per fragment. Hence, a single 10-hour sequencing experiment can generate more than 100,000,000 bases (100 megabases). Although significant hurdles remain, this sequencing capacity suggests that the rapid sequencing of anyone's genome at low cost is a very real possibility. Individual genome sequences will provide information about genetic variation within populations and may usher in an era of personalized medicine, when these data can be used to guide treatment decisions.

Comparative genomics has become a powerful research tool

Comparisons with genomes from other organisms are a source of insight into the human genome. The sequencing of the genome of the chimpanzee, our closest living relative, is nearing completion. The genomes of other mammals that are widely used in biological research, such as the mouse and

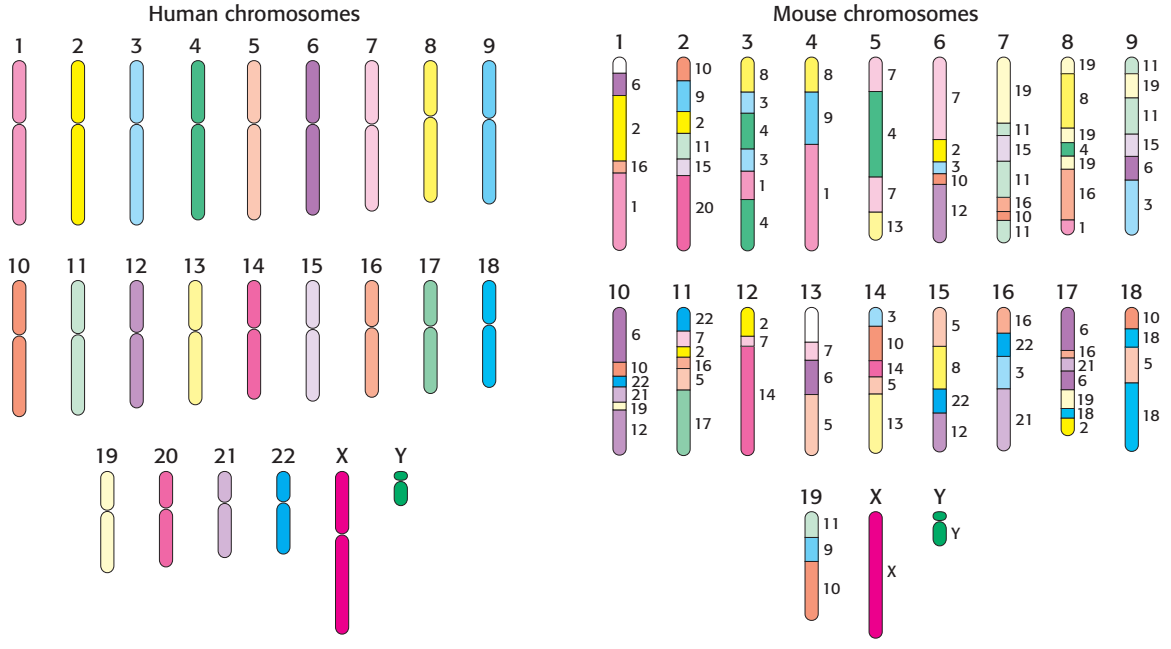


Figure 5.27 Genome comparison. A schematic comparison of the human genome and the mouse genome shows reassortment of large chromosomal fragments.

the rat, have been completed. Comparisons reveal that an astonishing 99% of human genes have counterparts in these rodent genomes. However, these genes have been substantially reassorted among chromosomes in the estimated 75 million years of evolution since humans and rodents had a common ancestor (Figure 5.27).

The genomes of other organisms also have been determined specifically for use in comparative genomics. For example, the genomes of two species of puffer fish, *Takifugu rubripes* and *Tetraodon nigroviridis,* have been determined. These genomes were selected because they are very small and lack much of the intergenic DNA present in such abundance in the human genome. The puffer fish genomes include fewer than 400 megabase pairs (Mbp), one-eighth of the number in the human genome, yet the puffer fish and human genomes contain essentially the same number of genes. Comparison of the genomes of these species with that of humans revealed more than 1000 formerly unrecognized human genes. Furthermore, comparison of the two species of puffer fish, which had a common ancestor approximately 25 million years ago, is a source of insight into more-recent events in evolution. Comparative genomics is a powerful tool, both for interpreting the human genome and for understanding major events in the origin of genera and species.

A puffer fish. [Fred Bavendam/Peter Arnold.]

5.4 Eukaryotic Genes Can Be Quantitated and Manipulated with Considerable Precision

After a gene of interest has been identified, cloned, and sequenced, it is often desirable to understand how that gene and its corresponding protein product function in the context of a whole cell or organism. It is now possible to determine how the expression of a particular gene is regulated, how mutations in the gene affect the function of the corresponding protein product, and how the behavior of an entire cell or model organism is altered by the introduction of mutations within specific genes. Levels of transcription of large families of genes within cells and tissues can be readily quantitated and compared across a range of environmental conditions. Eukaryotic genes can be introduced into bacteria, and the bacteria can be used as factories to produce a desired protein product. DNA can also be introduced into the cells of higher organisms. Genes introduced into animals are valuable tools for examining gene action, and they are the basis of gene therapy. Genes introduced into plants can make the plants resistant to pests, able to grow in harsh conditions, or carry greater quantities of essential nutrients. The manipulation of eukaryotic genes holds much promise as a source of medical and agricultural benefits, but it is also a source of controversy.

Gene-expression levels can be comprehensively examined

Most genes are present in the same quantity in every cell—namely, one copy per haploid cell or two copies per diploid cell. However, the level at which a gene is expressed, as indicated by mRNA quantities, can vary widely, ranging from no expression to hundreds of mRNA copies per cell. Gene-expression patterns vary from cell type to cell type, distinguishing, for example, a muscle cell from a nerve cell. Even within the same cell, gene-expression levels may vary as the cell responds to changes in physiological circumstances. Note that mRNA levels sometimes correlate with the levels of proteins expressed, but this correlation does not always hold. Thus, care must be exercised when interpreting the results of mRNA levels alone.

The quantity of individual mRNA transcripts can be determined by *quantitative PCR* (qPCR), or real-time PCR. RNA is first isolated from the

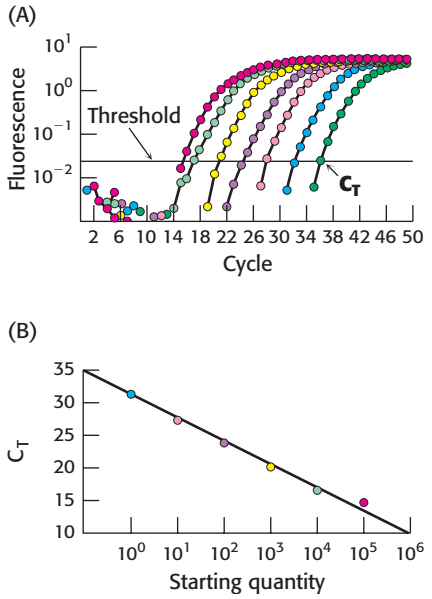


Figure 5.28 Quantitative PCR. (A) In qPCR, fluorescence is monitored in the course of PCR amplification to determine C_T, the cycle at which this signal exceeds a defined threshold. Each color represents a different starting quantity of DNA. (B) C_T values are inversely proportional to the number of copies of the original cDNA template. [After N. J. Walker, *Science* 296:557–559, 2002.]

cell or tissue of interest. With the use of reverse transcriptase, cDNA is prepared from this RNA sample. In one qPCR approach, the transcript of interest is PCR amplified with the appropriate primers in the presence of the dye SYBR Green I, which fluoresces brightly when bound to double-stranded DNA. In the initial PCR cycles, not enough duplex is present to allow a detectable fluorescence signal. However, after repeated PCR cycles, the fluorescence intensity exceeds the detection threshold and continues to rise as the number of duplexes corresponding to the transcript of interest increases (Figure 5.28). Importantly, the cycle number at which the fluorescence becomes detectable over a defined threshold (or C_T) is indirectly proportional to the number of copies of the original template. After the relation between the original copy number and the C_T has been established with the use of a known standard, subsequent qPCR experiments can be used to determine the number of copies of any desired transcript in the original sample, provided the appropriate primers are available.

Although qPCR is a powerful technique for quantitation of a small number of transcripts in any given experiment, we can now use our knowledge of complete genome sequences to investigate an entire *transcriptome*, the pattern and level of expression of all genes in a particular cell or tissue. One of the most powerful methods developed to date for this purpose is based on hybridization. Oligonucleotides or cDNAs are affixed to a solid support such as a microscope slide, creating a *DNA microarray*. Fluorescently labeled cDNA is hybridized to the slide to reveal the expression level for each gene, identifiable by its known position within the microarray (Figure 5.29). The intensity of the fluorescent spot on the chip reveals the extent of the transcription of a particular gene. DNA chips have been prepared that contain oligonucleotides complementary to all known protein-encoding genes, 6200 in number, within the yeast genome (Figure 5.30). An analysis of mRNA pools with the use of these chips revealed, for example, that approximately 50% of all yeast genes are expressed at steady-state levels of 0.1 to 1.0 mRNA copy per cell. This method readily detected variations in expression levels displayed by specific genes under different growth conditions.

Microarray analyses can be quite informative in the study of gene-expression changes in diseased mammals compared with their healthy

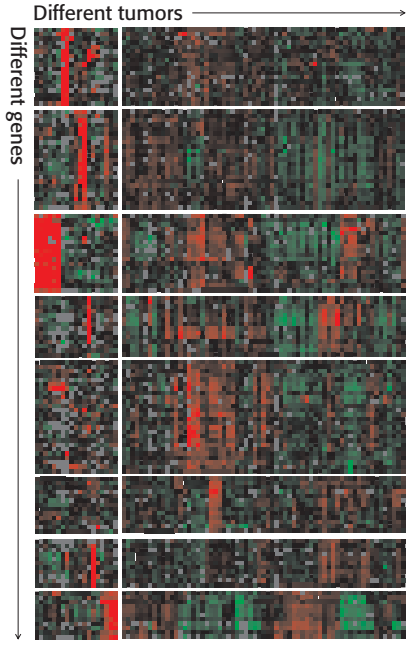


Figure 5.29 Gene-expression analysis with the use of microarrays. The expression levels of thousands of genes can be simultaneously analyzed by using DNA microarrays (gene chips). Here, an analysis of 1733 genes in 84 breast-tumor samples reveals that the tumors can be divided into distinct classes on the basis of their gene-expression patterns. Red corresponds to gene induction and green corresponds to gene repression. [After C. M. Perou et al., *Nature* 406:747–752, 2000.]

counterparts. As noted earlier, although ALS-causing mutations within the *SOD1* gene had been identified, the mechanism by which the mutant SOD1 protein ultimately leads to motor-neuron loss remains a mystery. Many research groups have used microarray analysis of neuronal cells isolated from humans and mice carrying *SOD1* mutations to search for clues into the pathways of disease progression and to suggest potential avenues for treatment. These studies have implicated the participation of a variety of biochemical pathways, including immunological activation, handling of oxidative stress, and protein degradation, in the cellular response to the mutant, toxic forms of SOD1.

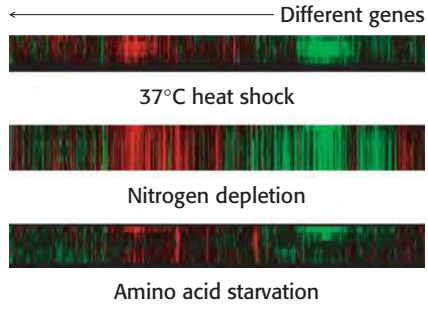


Figure 5.30 Monitoring changes in yeast gene expression. This microarray analysis shows levels of gene expression for yeast genes under different conditions. [After V. R. Iyer et al., *Nature* 409:533–538, 2001.]

New genes inserted into eukaryotic cells can be efficiently expressed

Bacteria are ideal hosts for the amplification of DNA molecules. They can also serve as factories for the production of a wide range of prokaryotic and eukaryotic proteins. However, bacteria lack the necessary enzymes to carry out posttranslational modifications such as the specific cleavage of polypeptides and the attachment of carbohydrate units. Thus, many eukaryotic genes can be correctly expressed only in eukaryotic host cells. The introduction of recombinant DNA molecules into cells of higher organisms can also be a source of insight into how their genes are organized and expressed. How are genes turned on and off in embryological development? How does a fertilized egg give rise to an organism with highly differentiated cells that are organized in space and time? These central questions of biology can now be fruitfully approached by expressing foreign genes in mammalian cells.

Recombinant DNA molecules can be introduced into animal cells in several ways. In one method, foreign DNA molecules precipitated by calcium phosphate are taken up by animal cells. A small fraction of the imported DNA becomes stably integrated into the chromosomal DNA. The efficiency of incorporation is low, but the method is useful because it is easy to apply. In another method, DNA is *microinjected* into cells. A fine-tipped (0.1-mm-diameter) glass micropipette containing a solution of foreign DNA is inserted into a nucleus (Figure 5.31). A skilled investigator can inject hundreds of cells per hour. About 2% of injected mouse cells are viable and contain the new gene. In a third method, *viruses* are used to introduce new genes into animal cells. The most effective vectors are *retroviruses,* which replicate through DNA intermediates, the reverse of the normal flow of information. A striking feature of the life cycle of a retrovirus is that the double-helical DNA form of its genome, produced by the action of reverse transcriptase, becomes randomly incorporated into host chromosomal DNA. This DNA version of the viral genome, called *proviral DNA,* can be efficiently expressed by the host cell and replicated along with normal cellular DNA. Retroviruses do not usually kill their hosts. Foreign genes have been efficiently introduced into mammalian cells by infecting them with vectors derived from the *Moloney murine leukemia virus,* which can accept inserts as long as 6 kb. Some genes introduced by this retroviral vector into the genome of a transformed host cell are efficiently expressed.

Two other viral vectors are extensively used. *Vaccinia virus,* a large DNA-containing virus, replicates in the cytoplasm of mammalian cells, where it shuts down host-cell protein synthesis. *Baculovirus* infects insect cells, which can

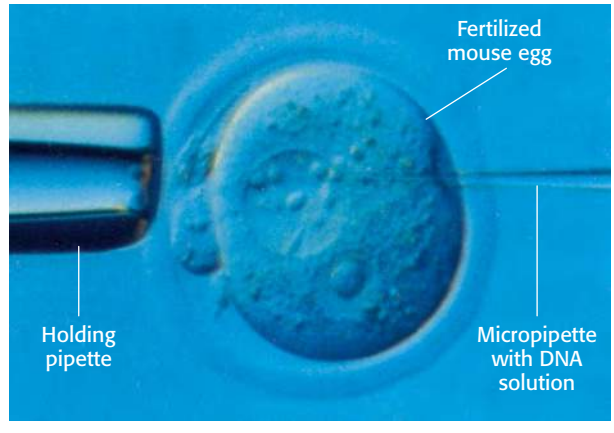


Figure 5.31 Microinjection of DNA. Cloned plasmid DNA is being microinjected into the male pronucleus of a fertilized mouse egg.

be conveniently cultured. Insect larvae infected with this virus can serve as efficient protein factories. Vectors based on these large-genome viruses have been engineered to express DNA inserts efficiently.

Transgenic animals harbor and express genes introduced into their germ lines

As shown in Figure 5.31, plasmids harboring foreign genes can be microinjected into the male pronucleus of fertilized mouse eggs, which are then inserted into the uterus of a foster-mother mouse. A subset of the resulting embryos in this host will then harbor the foreign gene; these embryos may develop into mature animals. Southern blotting of the DNA of the progeny can be used to determine which offspring carry the introduced gene. These *transgenic mice* are a powerful means of exploring the role of a specific gene in the development, growth, and behavior of an entire organism. Transgenic animals often serve as useful models for a particular disease process, enabling researchers to test the efficacy and safety of a newly developed therapy.

Let us return to our example of ALS. Research groups have generated transgenic mouse lines that express forms of human superoxide dismutase that harbor mutations matching those identified in earlier genetic analyses. Many of these strains exhibit a clinical picture similar to that observed in ALS patients: progressive weakness of voluntary muscles and eventual paralysis, motor-neuron loss, and rapid progression to death (Figure 5.32). Since their first characterization in 1994, these strains continue to serve as valuable sources of information for the exploration of the mechanism, and potential treatment, of ALS.

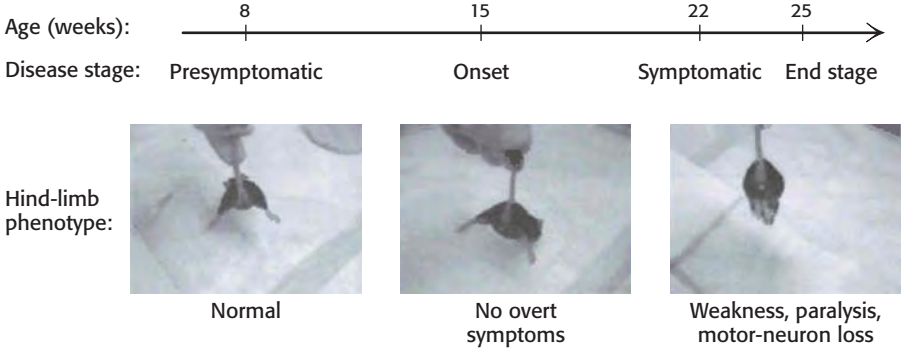


Figure 5.32 Transgenic mice. Mice expressing human SOD1 harboring a known ALS-causing mutation exhibit a phenotype similar to the human disease, including the loss of motor neurons, voluntary muscle weakness, and paralysis. [After C. S. Lobsinger et al., *PNAS* 104:7319–7326, 2007. Copyright 2007 National Academy of Sciences, U. S. A.]

Gene disruption provides clues to gene function

A gene's function can also be probed by inactivating the gene and looking for resulting abnormalities. Powerful methods have been developed for accomplishing *gene disruption* (also called *gene knockout*) in organisms such as yeast and mice. These methods rely on the process of *homologous recombination*. Through this process, regions of strong sequence similarity exchange segments of DNA. Foreign DNA inserted into a cell can thus disrupt any gene that is at least partly homologous by the exchange of segments (Figure 5.33). Specific genes can be targeted if their nucleotide sequences are known.

For example, the gene-knockout approach has been applied to the genes encoding gene-regulatory proteins (also called *transcription factors*) that

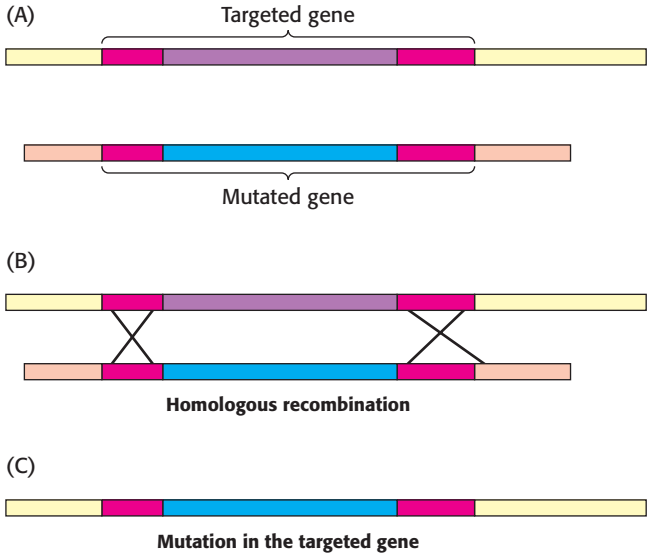


Figure 5.33 Gene disruption by homologous recombination.
(A) A mutated version of the gene to be disrupted is constructed, maintaining some regions of homology with the normal gene (red). When the foreign mutated gene is introduced into an embryonic stem cell, (B) recombination takes place at regions of homology and (C) the normal (targeted) gene is replaced, or "knocked out," by the foreign gene. The cell is inserted into embryos, and mice lacking the gene (knockout mice) are produced.

control the differentiation of muscle cells. When both copies of the gene for the regulatory protein *myogenin* are disrupted, an animal dies at birth because it lacks functional skeletal muscle. Microscopic inspection reveals that the tissues from which muscle normally forms contain precursor cells that have failed to differentiate fully (Figure 5.34). Heterozygous mice containing one normal myogenin gene and one disrupted gene appear normal, suggesting that the level of gene expression is not essential for its function. Analogous studies have probed the function of many other genes to generate animal models for known human genetic diseases.

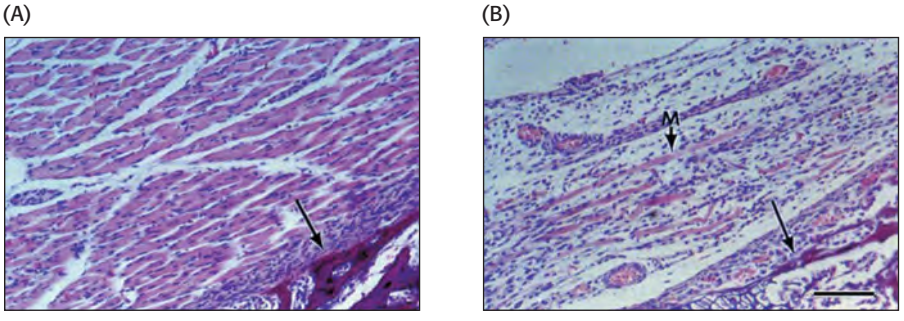


Figure 5.34 Consequences of gene disruption. Sections of muscle from normal (A) and gene-disrupted (B) mice, as viewed under the light microscope. Muscles do not develop properly in mice having both myogenin genes disrupted. [From P. Hasty, A. Bradley, J. H. Morris, D. G. Edmondson, J. M. Venuti, E. N. Olson, and W. H. Klein, *Nature* 364:501–506, 1993.]

RNA interference provides an additional tool for disrupting gene expression

An extremely powerful tool for disrupting gene expression was serendipitously discovered in the course of studies that required the introduction of RNA into a cell. The introduction of a specific double-stranded RNA molecule into a cell was found to suppress the transcription of genes that contained sequences present in the double-stranded RNA molecule. Thus, the introduction of a specific RNA molecule can interfere with the expression of a specific gene.

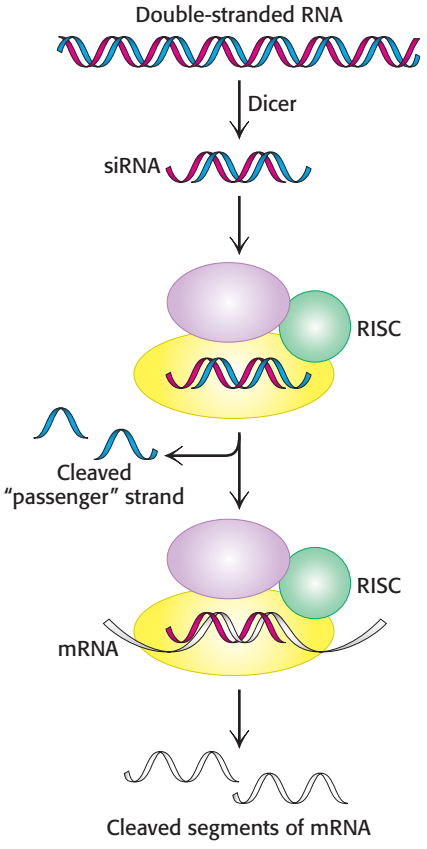


Figure 5.35 RNA interference mechanism. A double-stranded RNA molecule is cleaved into 21-bp fragments by the enzyme Dicer to produce siRNAs. These siRNAs are incorporated into the RNA-induced silencing complex (RISC), where the single-stranded RNAs guide the cleavage of mRNAs that contain complementary sequences.

The mechanism of *RNA interference* has been largely established (Figure 5.35). When a double-stranded RNA molecule is introduced into an appropriate cell, the RNA is cleaved by an enzyme referred to as *Dicer* into fragments approximately 21 nucleotides in length. Each fragment, termed a small interfering RNA (siRNA), consists of 19 bp of double-stranded RNA and 2 bases of unpaired RNA on each 5′ end. The siRNA is loaded into an assembly of several proteins referred to as the *RNA-induced silencing complex* (RISC), which unwinds the RNA duplex and cleaves one of the strands, the so-called *passenger strand*. The uncleaved single-stranded RNA segment, the *guide strand*, remains incorporated into the enzyme. The fully assembled RISC cleaves mRNA molecules that contain exact complements of the guide-strand sequence. Thus, levels of such mRNA molecules are dramatically reduced.

The machinery necessary for RNA interference is found in many cells. In some organisms such as *C. elegans,* RNA interference is quite efficient. Indeed, RNA interference can be induced simply by feeding *C. elegans* strains of *E. coli* that have been engineered to produce appropriate double-stranded RNA molecules. Although not as efficient in mammalian cells, RNA interference has emerged as a powerful research tool for reducing the expression of specific genes. Moreover, initial clinical trials of therapies based on RNA interference are underway.

Tumor-inducing plasmids can be used to introduce new genes into plant cells

The common soil bacterium *Agrobacterium tumefaciens* infects plants and introduces foreign genes into plants cells (Figure 5.36). A lump of tumor tissue called a *crown gall* grows at the site of infection. Crown galls synthesize opines, a group of amino acid derivatives that are metabolized by the infecting bacteria. In essence, the metabolism of the plant cell is diverted to satisfy the highly distinctive appetite of the intruder. *Tumor-inducing plasmids* (Ti plasmids) that are carried by *A. tumefaciens* carry instructions for the switch to the tumor state and the synthesis of opines. A small part of the Ti plasmid becomes integrated into the genome of infected plant cells; this 20-kb segment is called *T-DNA* (transferred DNA; Figure 5.37).

Ti-plasmid derivatives can be used as vectors to deliver foreign genes into plant cells. First, a segment of foreign DNA is inserted into the T-DNA

Figure 5.36 Tumors in plants. Crown gall, a plant tumor, is caused by a bacterium (*Agrobacterium tumefaciens*) that carries a tumor-inducing plasmid (Ti plasmid). [From M. Escobar et al., *PNAS* 98:13437–13442, 2001. Copyright 2001 National Academy of Sciences, U. S. A.]

region of a small plasmid through the use of restriction enzymes and ligases. This synthetic plasmid is added to *A. tumefaciens* colonies harboring naturally occurring Ti plasmids. By recombination, Ti plasmids containing the foreign gene are formed. These Ti vectors hold great promise as tools for exploring the genomes of plant cells and modifying plants to improve their agricultural value and crop yield. However, they are not suitable for transforming all types of plants. Ti-plasmid transfer is effective with dicots (broad-leaved plants such as grapes) and a few kinds of monocots but not as effective with economically important cereal monocots.

Foreign DNA can be introduced into cereal monocots as well as dicots by applying intense electric fields, a technique called *electroporation* (Figure 5.38). First, the cellulose wall surrounding plant cells is removed by adding cellulase; this treatment produces *protoplasts,* plant cells with exposed plasma membranes. Electric pulses are then applied to a suspension of protoplasts and plasmid DNA. Because high electric fields make membranes transiently permeable to large molecules, plasmid DNA molecules enter the cells. The cell wall is then allowed to reform, and the plant cells are again viable. Maize cells and carrot cells have been stably transformed in this way with the use of plasmid DNA that includes genes for resistance to antibiotics. Moreover, the transformed cells efficiently express the plasmid DNA. Electroporation is also an effective means of delivering foreign DNA into animal cells and bacterial cells.

The most effective means of transforming plant cells is through the use of *"gene guns,"* or *bombardment-mediated transformation*. DNA is coated onto 1-mm-diameter tungsten pellets, and these microprojectiles are fired at the target cells with a velocity greater than 400 m s^{-1}. Despite its apparent crudeness, this technique is proving to be the most effective way of transforming plants, especially important crop species such as soybean, corn, wheat, and rice. The gene-gun technique affords an opportunity to develop genetically modified organisms (GMOs) with beneficial characteristics. Such characteristics could include the ability to grow in poor soils, resistance to natural climatic variation, resistance to pests, and nutritional fortification. These crops might be most useful in developing countries. The use of genetically modified organisms is highly controversial at this point because of fears of unexpected side effects.

The first GMO to come to market was a tomato characterized by delayed ripening, rendering it ideal for shipment. Pectin is a polysaccharide that gives tomatoes their firmness and is naturally destroyed by the enzyme *polygalacturonase*. As pectin is destroyed, the tomatoes soften, making shipment difficult. DNA was introduced that disrupts the polygalacturonase gene. Less of the enzyme was produced, and the tomatoes stayed fresh longer. However, the tomato's poor taste hindered its commercial success.

Human gene therapy holds great promise for medicine

The field of *gene therapy* attempts to express specific genes within the human body in such a way that beneficial results are obtained. The gene targeted for expression may be already present or specially introduced. Alternatively, gene therapy may attempt to modify genes containing sequence variations that have harmful consequences. A tremendous amount of research remains to be done before gene therapy becomes practical. Nonetheless, considerable progress has been made. For example, some people lack functional genes for *adenosine deaminase* and succumb to infections if exposed to a normal environment, a condition called *severe combined*

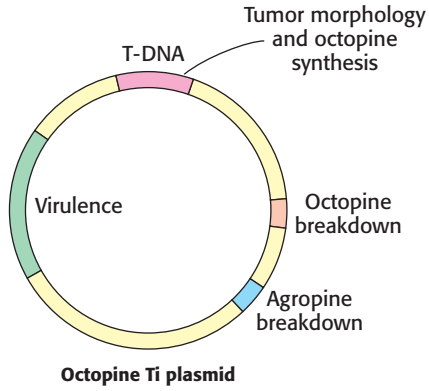


Figure 5.37 Ti plasmids. Agrobacteria containing Ti plasmids can deliver foreign genes into some plant cells. [After M. Chilton. A vector for introducing new genes into plants. Copyright © 1983 by Scientific American, Inc. All rights reserved.]

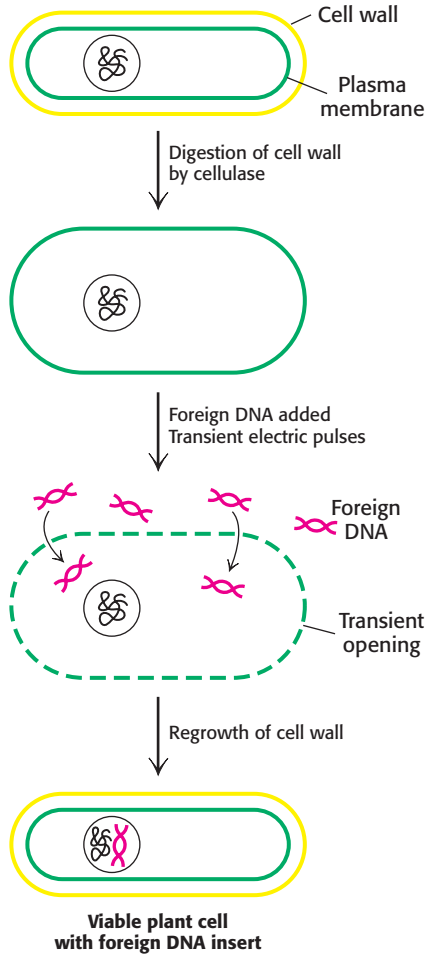


Figure 5.38 Electroporation. Foreign DNA can be introduced into plant cells by electroporation, the application of intense electric fields to make their plasma membranes transiently permeable.

immunodeficiency (SCID). Functional genes for this enzyme have been introduced by using gene-therapy vectors based on retroviruses. Although these vectors have produced functional enzyme and reduced the clinical symptoms, challenges remain. These challenges include increasing the longevity of the effects and eliminating unwanted side effects. Future research promises to transform gene therapy into an important tool for clinical medicine.

Summary

5.1 The Exploration of Genes Relies on Key Tools

The recombinant DNA revolution in biology is rooted in the repertoire of enzymes that act on nucleic acids. Restriction enzymes are a key group among them. These endonucleases recognize specific base sequences in double-helical DNA and cleave both strands of the duplex, forming specific fragments of DNA. These restriction fragments can be separated and displayed by gel electrophoresis. The pattern of these fragments on the gel is a fingerprint of a DNA molecule. A DNA fragment containing a particular sequence can be identified by hybridizing it with a labeled single-stranded DNA probe (Southern blotting).

Rapid sequencing techniques have been developed to further the analysis of DNA molecules. DNA can be sequenced by controlled interruption of replication. The fragments produced are separated by gel electrophoresis and visualized by autoradiography of a ^{32}P label at the 5′ end or by fluorescent tags.

DNA probes for hybridization reactions, as well as new genes, can be synthesized by the automated solid-phase method. The technique is to add deoxyribonucleoside 3′-phosphoramidites to one another to form a growing chain that is linked to an insoluble support. DNA chains a hundred nucleotides long can be readily synthesized. The polymerase chain reaction makes it possible to greatly amplify specific segments of DNA in vitro. The region amplified is determined by the placement of a pair of primers that are added to the target DNA along with a thermostable DNA polymerase and deoxyribonucleoside triphosphates. The exquisite sensitivity of PCR makes it a choice technique in detecting pathogens and cancer markers, in genotyping, and in reading DNA from fossils that are many thousands of years old.

5.2 Recombinant DNA Technology Has Revolutionized All Aspects of Biology

New genes can be constructed in the laboratory, introduced into host cells, and expressed. Novel DNA molecules are made by joining fragments that have complementary cohesive ends produced by the action of a restriction enzyme. DNA ligase seals breaks in DNA chains. Vectors for propagating the DNA include plasmids, λ phage, and bacterial and yeast artificial chromosomes. Specific genes can be cloned from a genomic library with the use of a DNA or RNA probe. Foreign DNA can be expressed after insertion into prokaryotic and eukaryotic cells by the appropriate vector. Specific mutations can be generated in vitro to engineer novel proteins. A mutant protein with a single amino acid substitution can be produced by priming DNA replication with an oligonucleotide encoding the new amino acid. Plasmids can be engineered to

permit the facile insertion of a DNA cassette containing any desired mutation. The techniques of protein and nucleic acid chemistry are highly synergistic. Investigators now move back and forth between gene and protein with great facility.

5.3 Complete Genomes Have Been Sequenced and Analyzed

The sequences of many important genomes are known in their entirety. More than 100 bacterial and archaeal genomes have been sequenced, including those from key model organisms and important pathogens. The sequence of the human genome has now been completed with nearly full coverage and high precision. Only from 20,000 to 25,000 protein-encoding genes appear to be present in the human genome, a substantially smaller number than earlier estimates. Comparative genomics has become a powerful tool for analyzing individual genomes and for exploring evolution. Genomewide gene-expression patterns can be examined through the use of DNA microarrays.

5.4 Eukaryotic Genes Can Be Quantitated and Manipulated with Considerable Precision

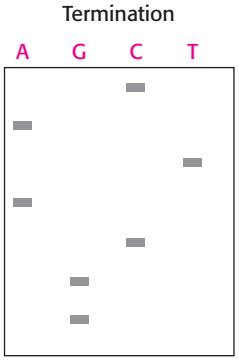
Changes in gene expression can be readily determined by such techniques as quantitative PCR and hybridization to microarrays. The production of transgenic mice carrying mutations known to cause ALS in humans has been a source of considerable insight into the disease mechanism and its possible treatment. The functions of particular genes can be investigated by disruption. One method of disrupting the expression of a particular gene is through RNA interference, which depends on the introduction of specific double-stranded RNA molecules into eukaryotic cells. New DNA can be brought into plant cells by the soil bacterium *Agrobacterium tumefaciens*, which harbors Ti plasmids. DNA can also be introduced into plant cells by applying intense electric fields, which render them transiently permeable to very large molecules, or by bombarding them with DNA-coated microparticles. Gene therapy holds great promise for clinical medicine, but many challenges remain.

Key Terms

restriction enzyme (p. 141)
palindrome (p. 141)
DNA probe (p. 142)
Southern blotting (p. 142)
northern blotting (p. 142)
controlled termination of replication (Sanger dideoxy method) (p. 143)
polymerase chain reaction (PCR) (p. 145)
polymorphism (p. 147)
vector (p. 148)
plasmid (p. 148)
sticky ends (p. 148)
DNA ligase (p. 148)
expression vector (p. 149)
lambda (λ) phage (p. 150)
bacterial artificial chromosome (BAC) (p. 151)
yeast artificial chromosome (YAC) (p. 151)
genomic library (p. 153)
complementary DNA (cDNA) (p. 154)
reverse transcriptase (p. 154)
cDNA library (p. 154)
oligonucleotide-directed mutagenesis (p. 156)
cassette mutagenesis (p. 157)
pseudogene (p. 159)
mobile genetic element (p. 159)
short interspersed elements (SINES) (p. 160)
long interspersed elements (LINES) (p. 160)
quantitative PCR (qPCR) (p. 161)
transcriptome (p. 162)
DNA microarray (gene chip) (p. 162)
transgenic mouse (p. 164)
gene disruption (gene knockout) (p. 164)
RNA interference (p. 166)
RNA-induced silencing complex (RISC) (p. 166)
tumor-inducing plasmid (Ti plasmid) (p. 166)
gene gun (bombardment-mediated transformation) (p. 167)

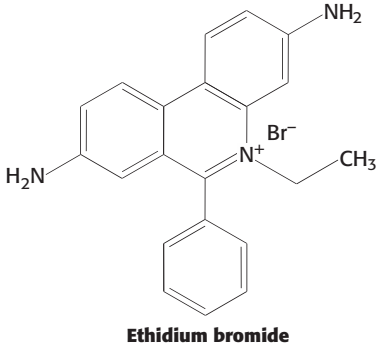
Problems

1. *Reading sequences.* An autoradiogram of a sequencing gel containing four lanes of DNA fragments is shown in the adjoining illustration. (a) What is the sequence of the DNA fragment? (b) Suppose that the Sanger dideoxy method shows that the template strand sequence is 5′-TGCAATGGC-3′. Sketch the gel pattern that would lead to this conclusion.



2. *The right template.* Ovalbumin is the major protein of egg white. The chicken ovalbumin gene contains eight exons separated by seven introns. Should ovalbumin cDNA or ovalbumin genomic DNA be used to form the protein in *E. coli?* Why?

3. *Handle with care.* Ethidium bromide is a commonly used stain for DNA molecules after separation by gel electrophoresis. The chemical structure of ethidium bromide is shown here. Based on this structure, suggest how this stain binds to DNA.



Ethidium bromide

4. *Cleavage frequency.* The restriction enzyme *Alu*I cleaves at the sequence 5′-AGCT-3′, and *Not*I cleaves at 5′-GCGGCCGC-3′. What would be the average distance between cleavage sites for each enzyme on digestion of double-stranded DNA? Assume that the DNA contains equal proportions of A, G, C, and T.

5. *The right cuts.* Suppose that a human genomic library is prepared by exhaustive digestion of human DNA with the *Eco*RI restriction enzyme. Fragments averaging about 4 kb in length would be generated. Is this procedure suitable for cloning large genes? Why or why not?

6. *A revealing cleavage.* Sickle-cell anemia arises from a mutation in the gene for the β chain of human hemoglobin. The change from GAG to GTG in the mutant eliminates a cleavage site for the restriction enzyme *Mst*II, which recognizes the target sequence CCTGAGG. These findings form the basis of a diagnostic test for the sickle-cell gene. Propose a rapid procedure for distinguishing between the normal and the mutant gene. Would a positive result prove that the mutant contains GTG in place of GAG?

7. *Sticky ends?* The restriction enzymes *Kpn*I and *Acc*65I recognize and cleave the same 6-bp sequence. However, the sticky end formed from *Kpn*I cleavage cannot be ligated directly to the sticky end formed from *Acc*65I cleavage. Explain why.

8. *Many melodies from one cassette.* Suppose that you have isolated an enzyme that digests paper pulp and have obtained its cDNA. The goal is to produce a mutant that is effective at high temperature. You have engineered a pair of unique restriction sites in the cDNA that flank a 30-bp coding region. Propose a rapid technique for generating many different mutations in this region.

9. *A blessing and a curse.* The power of PCR can also create problems. Suppose someone claims to have isolated dinosaur DNA by using PCR. What questions might you ask to determine if it is indeed dinosaur DNA?

10. *Rich or poor?* DNA sequences that are highly enriched in G–C base pairs typically have high melting temperatures. Moreover, once separated, single strands containing these regions can form rigid secondary structures. How might the presence of G–C-rich regions in a DNA template affect PCR amplification?

11. *Questions of accuracy.* The stringency of PCR amplification can be controlled by altering the temperature at which the primers and the target DNA undergo hybridization. How would altering the temperature of hybridization affect the amplification? Suppose that you have a particular yeast gene *A* and that you wish to see if it has a counterpart

in humans. How would controlling the stringency of the hybridization help you?

12. *Terra incognita.* PCR is typically used to amplify DNA that lies between two known sequences. Suppose that you want to explore DNA on both sides of a single known sequence. Devise a variation of the usual PCR protocol that would enable you to amplify entirely new genomic terrain.

13. *A puzzling ladder.* A gel pattern displaying PCR products shows four strong bands. The four pieces of DNA have lengths that are approximately in the ratio of 1:2:3:4. The largest band is cut out of the gel, and PCR is repeated with the same primers. Again, a ladder of four bands is evident in the gel. What does this result reveal about the structure of the encoded protein?

14. *Chromosome walking.* Propose a method for isolating a DNA fragment that is adjacent in the genome to a previously isolated DNA fragment. Assume that you have access to a complete library of DNA fragments in a BAC vector but that the sequence of the genome under study has not yet been determined.

15. *Probe design.* Which of the following amino acid sequences would yield the most optimal oligonucleotide probe?

```
Ala-Met-Ser-Leu-Pro-Trp
Gly-Trp-Asp-Met-His-Lys
Cys-Val-Trp-Asn-Lys-Ile
Arg-Ser-Met-Leu-Gln-Asn
```

16. *Man's best friend.* Why might the genomic analysis of dogs be particularly useful for investigating the genes responsible for body size and other physical characteristics?

17. *Of mice and men.* You have identified a gene that is located on human chromosome 20 and wish to identify its location within the mouse genome. On which chromosome would you be most likely to find the mouse counterpart of this gene?

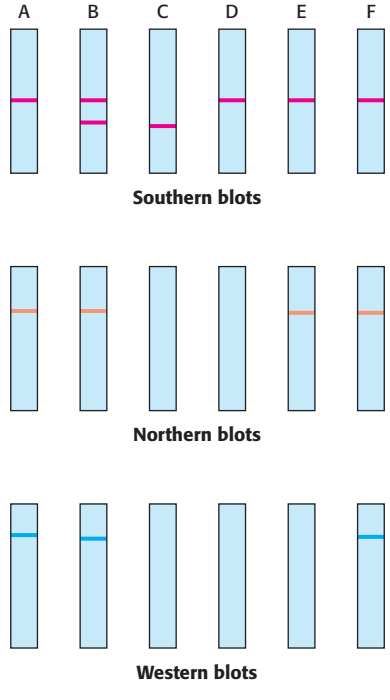
Chapter Integration Problems

18. *Designing primers I.* A successful PCR experiment often depends on designing the correct primers. In particular, the T_m for each primer should be approximately the same. What is the basis of this requirement?

19. *Designing primers II.* You wish to amplify a segment of DNA from a plasmid template by PCR with the use of the following primers: 5′-GGATCGATGCTCGCGA-3′ and 5′-AGGATCGGGTCGCGAG-3′. Despite repeated attempts, you fail to observe a PCR product of the expected length after electrophoresis on an agarose gel. Instead, you observe a bright smear on the gel with an approximate length of 25 to 30 base pairs. Explain these results.

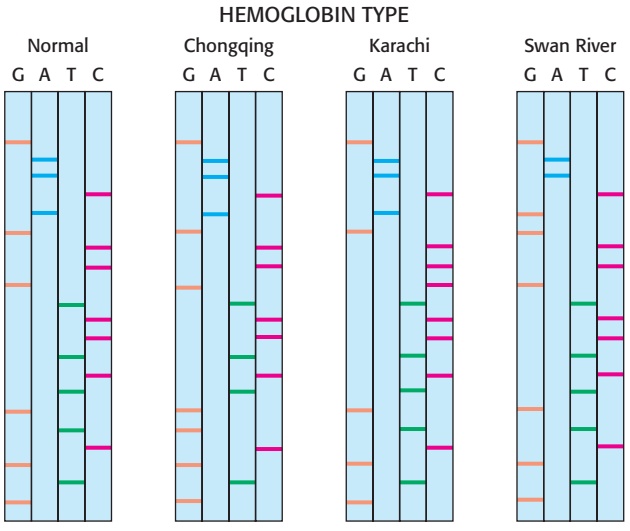
Chapter Integration and Data Interpretation Problem

20. *Any direction but east.* A series of people are found to have difficulty eliminating certain types of drugs from their bloodstreams. The problem has been linked to a gene *X*, which encodes an enzyme Y. Six people were tested with the use of various techniques of molecular biology. Person A is a normal control, person B is asymptomatic but some of his children have the metabolic problem, and persons C through F display the trait. Tissue samples from each person were obtained. Southern analysis was performed on the DNA after digestion with the restriction enzyme *Hin*dIII. Northern analysis of mRNA also was done. In both types of analysis, the gels were probed with labeled *X* cDNA. Finally, a western blot with an enzyme-linked monoclonal antibody was used to test for the presence of protein Y. The results are shown here. Why is person B without symptoms? Suggest possible defects in the other people.

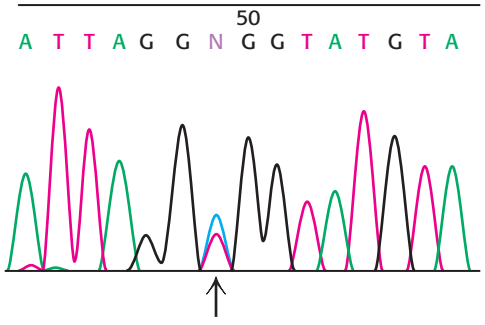


Data Interpretation Problems

21. *DNA diagnostics.* Representations of sequencing gels for variants of the α chain of human hemoglobin are shown here. What is the nature of the amino acid change in each of the variants? The first triplet encodes valine.



22. *Two peaks.* In the course of studying a gene and its possible mutation in humans, you obtain genomic DNA samples from a collection of persons and PCR amplify a region of interest within this gene. For one of the samples, you obtain the sequencing chromatogram shown here. Provide an explanation for the appearance of these data at position 49 (indicated by the arrow):



Animated Techniques

Visit www.whfreeman.com/Berg7e to see animations of Dideoxy Sequencing of DNA, Polymerase Chain Reaction, Synthesizing an Oligonucleotide Array, Screening an Oligonucleotide Array for Patterns of Gene Expression, Plasmid Cloning, In Vitro Mutagenesis of Cloned Genes, Creating a Transgenic Mouse. [Courtesy of H. Lodish et al., *Molecular Cell Biology*, 5th ed. (W. H. Freeman and Company, 2004).]

CHAPTER 6

Exploring Evolution and Bioinformatics

Evolutionary relationships are manifest in protein sequences. The close kinship between human beings and chimpanzees, hinted at by the mutual interest shown by Jane Goodall and a chimpanzee in the photograph, is revealed in the amino acid sequences of myoglobin. The human sequence (red) differs from the chimpanzee sequence (blue) in only one amino acid in a protein chain of 153 residues. [(Left) Kennan Ward/Corbis.]

```
GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGIL-
GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGIL-

KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLHSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG
KKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG
```

Like members of a human family, members of molecular families often have features in common. Such family resemblance is most easily detected by comparing three-dimensional structure, the aspect of a molecule most closely linked to function. Consider as an example ribonuclease from cows, which was introduced in our consideration of protein folding (Section 2.6). Comparing structures reveals that the three-dimensional structure of this protein and that of a human ribonuclease are quite similar (Figure 6.1). Although the degree of overlap between these two structures is not unexpected, given their nearly identical biological functions, similarities revealed by other such comparisons are sometimes surprising. For example, angiogenin, a protein that stimulates the growth of new blood vessels, also turns out to be structurally similar to ribonuclease—so similar that both angiogenin and ribonuclease are clearly members of the same protein family (Figure 6.2). Angiogenin and ribonuclease must have had a common ancestor at some earlier stage of evolution.

Three-dimensional structures have been determined for only a small proportion of the total number of proteins. In contrast, gene sequences and the corresponding amino acid sequences are available for a great number of

OUTLINE

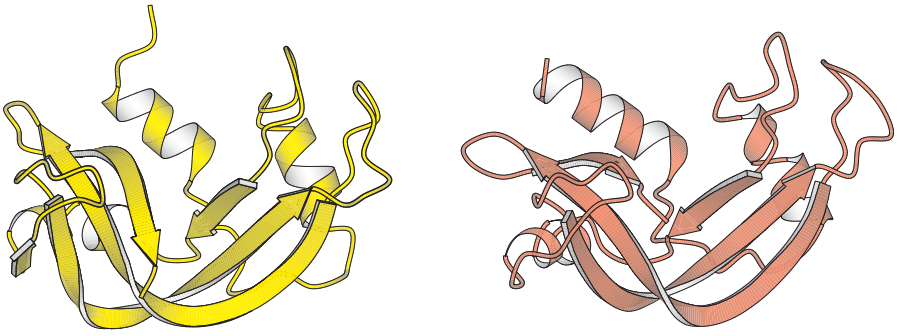


Figure 6.1 Structures of ribonucleases from cows and human beings. Structural similarity often follows functional similarity. [Drawn from 8RAT.pdb. and 2RNF.pdb.]

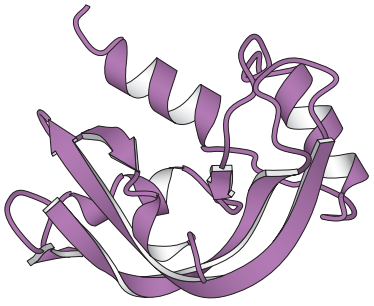


Figure 6.2 Structure of angiogenin. The protein angiogenin, identified on the basis of its ability to stimulate blood-vessel growth, is highly similar in three-dimensional structure to ribonuclease. [Drawn from 2ANG.pdb.]

proteins, largely owing to the tremendous power of DNA cloning and sequencing techniques including applications to complete-genome sequencing. Evolutionary relationships also are manifest in amino acid sequences. For example, 35% of the amino acids in corresponding positions are identical in the sequences of bovine ribonuclease and angiogenin. Is this level sufficiently high to ensure an evolutionary relationship? If not, what level is required? In this chapter, we shall examine the methods that are used to compare amino acid sequences and to deduce such evolutionary relationships.

Sequence-comparison methods have become powerful tools in modern biochemistry. Sequence databases can be probed for matches to a newly elucidated sequence to identify related molecules. This information can often be a source of considerable insight into the function and mechanism of the newly sequenced molecule. When three-dimensional structures are available, they can be compared to confirm relationships suggested by sequence comparisons and to reveal others that are not readily detected at the level of sequence alone.

By examining the footprints present in modern protein sequences, the biochemist can become a molecular archeologist able to learn about events in the evolutionary past. Sequence comparisons can often reveal pathways of evolutionary descent and estimated dates of specific evolutionary landmarks. This information can be used to construct evolutionary trees that trace the evolution of a particular protein or nucleic acid in many cases from Archaea and Bacteria through Eukarya, including human beings. Molecular evolution can also be studied experimentally. In some cases, DNA from fossils can be amplified by PCR methods and sequenced, giving a direct view into the past. In addition, investigators can observe molecular evolution taking place in the laboratory, through experiments based on nucleic acid replication. The results of such studies are revealing more about how evolution proceeds.

6.1 Homologs Are Descended from a Common Ancestor

The exploration of biochemical evolution consists largely of an attempt to determine how proteins, other molecules, and biochemical pathways have been transformed through time. The most fundamental relationship between two entities is *homology;* two molecules are said to be *homologous* if they have been derived from a common ancestor. Homologous molecules, or *homologs,* can be divided into two classes (Figure 6.3). *Paralogs* are homologs that are present within one species. Paralogs often differ in their detailed biochemical functions. *Orthologs* are homologs that are present within

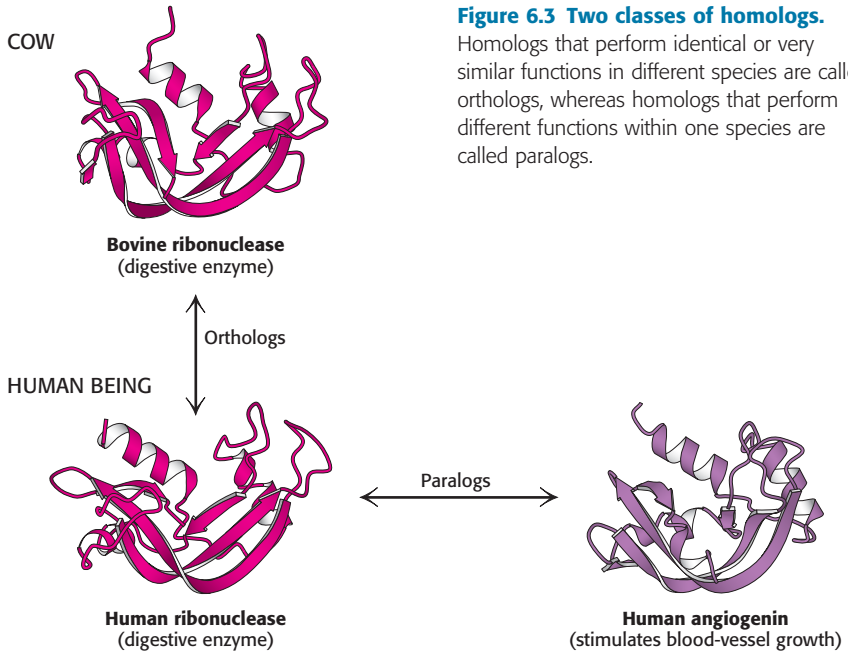


Figure 6.3 Two classes of homologs. Homologs that perform identical or very similar functions in different species are called orthologs, whereas homologs that perform different functions within one species are called paralogs.

different species and have very similar or identical functions. Understanding the homology between molecules can reveal the evolutionary history of the molecules as well as information about their function; if a newly sequenced protein is homologous to an already characterized protein, we have a strong indication of the new protein's biochemical function.

How can we tell whether two human proteins are paralogs or whether a yeast protein is the ortholog of a human protein? As will be discussed in Section 6.2, *homology is often detectable by significant similarity in nucleotide or amino acid sequence and almost always manifested in three-dimensional structure.*

6.2 Statistical Analysis of Sequence Alignments Can Detect Homology

A significant sequence similarity between two molecules implies that they are likely to have the same evolutionary origin and, therefore, similar three-dimensional structures, functions, and mechanisms. Both nucleic acid and protein sequences can be compared to detect homology. However, the possibility exists that the observed agreement between any two sequences is solely a product of chance. Because nucleic acids are composed of fewer building blocks than proteins (4 bases versus 20 amino acids), the likelihood of random agreement between two DNA or RNA sequences is significantly greater than that for protein sequences. For this reason, detection of homology between protein sequences is typically far more effective.

To illustrate sequence-comparison methods, let us consider a class of proteins called the *globins.* Myoglobin is a protein that binds oxygen in muscle, whereas hemoglobin is the oxygen-carrying protein in blood (Chapter 7). Both proteins cradle a heme group, an iron-containing organic molecule that binds the oxygen. Each human hemoglobin molecule is composed of four heme-containing polypeptide chains, two identical α chains and two identical β chains. Here, we consider only the α chain. To examine the similarity between the amino acid sequence of the human α chain and

Human hemoglobin (α chain)

VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHG
SAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLS
HCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR

Human myoglobin

GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKS
EDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVK
YLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG

Figure 6.4 Amino acid sequences of human hemoglobin (α chain) and human myoglobin. α-Hemoglobin is composed of 141 amino acids; myoglobin consists of 153 amino acids. (One-letter abbreviations designating amino acids are used; see Table 2.2.)

that of human myoglobin (Figure 6.4), we apply a method, referred to as a *sequence alignment*, in which the two sequences are systematically aligned with respect to each other to identify regions of significant overlap.

How can we tell where to align the two sequences? In the course of evolution, the sequences of two proteins that have an ancestor in common will have diverged in a variety of ways. Insertions and deletions may have occurred at the ends of the proteins or within the functional domains themselves. Individual amino acids may have been mutated to other residues of varying degrees of similarity. To understand how the methods of sequence alignment take these potential sequence variations into account, let us first consider the simplest approach, where we slide one sequence past the other, one amino acid at a time, and count the number of matched residues, or *sequence identities* (Figure 6.5). For α-hemoglobin and myoglobin, the best

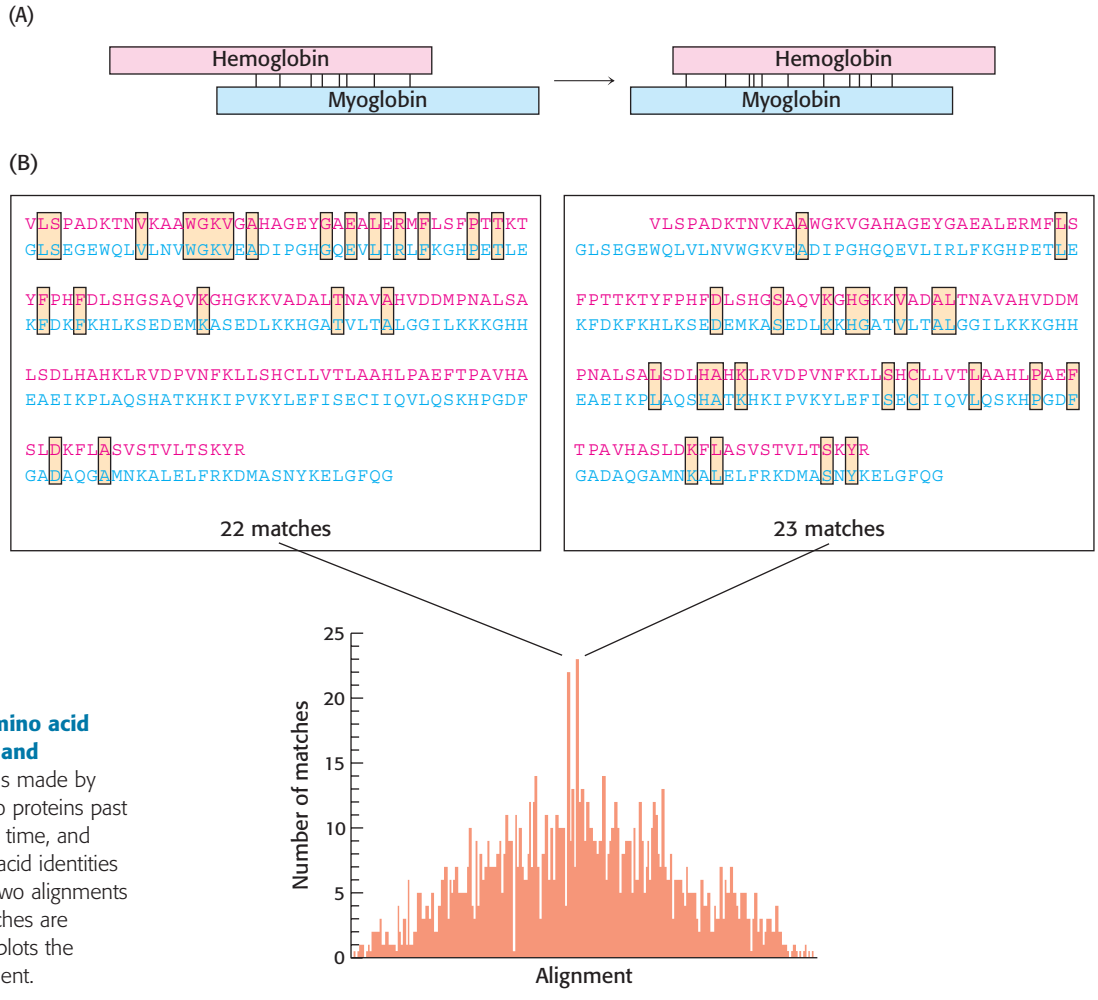


Figure 6.5 Comparing the amino acid sequences of α-hemoglobin and myoglobin. (A) A comparison is made by sliding the sequences of the two proteins past each other, one amino acid at a time, and counting the number of amino acid identities between the proteins. (B) The two alignments with the largest number of matches are shown above the graph, which plots the matches as a function of alignment.

alignment reveals 23 sequence identities, spread throughout the central parts of the sequences.

However, careful examination of all the possible alignments and their scores suggests that important information regarding the relationship between myoglobin and hemoglobin α has been lost with this method. In particular, we see that another alignment, featuring 22 identities, is nearly as good. This alignment is shifted by six residues relative to the preceding alignment and yields identities that are concentrated toward the amino-terminal end of the sequences. By introducing a *gap* into one of the sequences, the identities found in *both* alignments will be represented (Figure 6.6). Insertion of gaps allows the alignment method to compensate for the insertions or deletions of nucleotides that may have taken place in the gene for one molecule but not the other in the course of evolution.

The use of gaps substantially increases the complexity of sequence alignment because a vast number of possible gaps, varying in both position and length, must be considered throughout each sequence. Moreover, the introduction of an excessive number of gaps can yield an artificially high number of identities. Nevertheless, methods have been developed for the insertion of gaps in the automatic alignment of sequences. These methods use scoring systems to compare different alignments, including penalties for gaps to prevent the insertion of an unreasonable number of them. Here is an example of such a scoring system: each identity between aligned sequences is counted as +10 points, whereas each gap introduced, regardless of size, counts for −25 points. For the alignment shown in Figure 6.6, there are 38 identities ($38 \times 10 = 380$) and 1 gap ($1 \times -25 = -25$), producing a score of ($380 + -25 = 355$). Overall, there are 38 matched amino acids in an average length of 147 residues; so the sequences are 25.9% identical. Next, we must determine the significance of this score and level of identity.

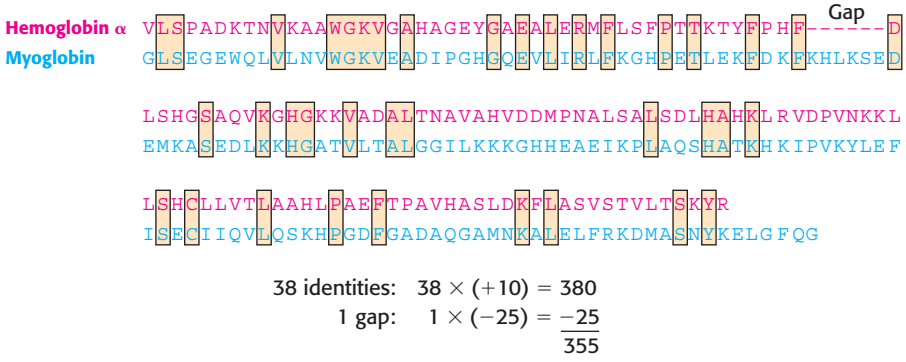


Figure 6.6 Alignment with gap insertion. The alignment of α-hemoglobin and myoglobin after a gap has been inserted into the hemoglobin α sequence.

The statistical significance of alignments can be estimated by shuffling

The similarities in sequence in Figure 6.5 appear striking, yet there remains the possibility that a grouping of sequence identities has occurred by chance alone. Because proteins are composed of the same set of 20 amino acid monomers, the alignment of any two unrelated proteins will yield some identities, particularly if we allow the introduction of gaps. Even if two proteins have identical amino acid composition, they may not be linked by evolution. It is the order of the residues within their sequences that implies a relationship between them. Hence, we can assess the significance of our alignment by "shuffling," or randomly rearranging, one of the sequences (Figure 6.7), repeat the sequence alignment, and determine a new alignment score. This process is repeated many times to yield a histogram showing, for each

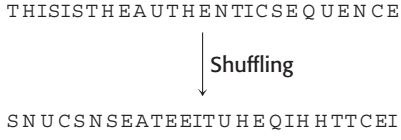


Figure 6.7 The generation of a shuffled sequence.

possible score, the number of shuffled sequences that received that score (Figure 6.8). If the original score is not appreciably different from the scores from the shuffled alignments, then we cannot exclude the possibility that the original alignment is merely a consequence of chance.

When this procedure is applied to the sequences of myoglobin and α-hemoglobin, the authentic alignment clearly stands out (see Figure 6.8). Its score is far above the mean for the alignment scores based on shuffled sequences. The probability that such a deviation occurred by chance alone is approximately 1 in 10^{20}. Thus, we can comfortably conclude that the two sequences are genuinely similar; the simplest explanation for this similarity is that these sequences are homologous—that is, that the two molecules have descended by divergence from a common ancestor.

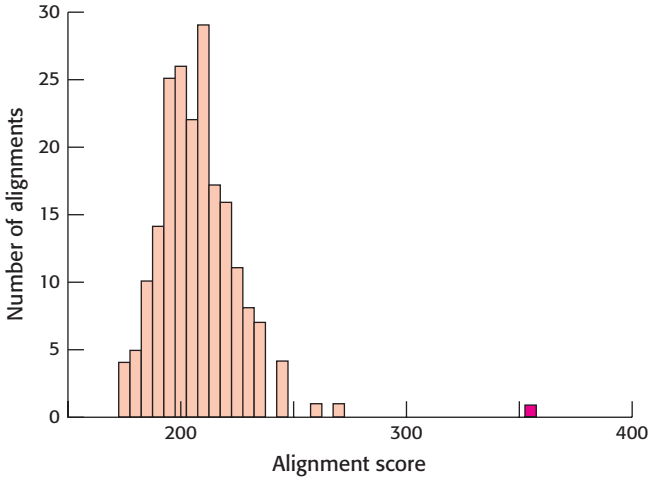


Figure 6.8 Statistical comparison of alignment scores. Alignment scores are calculated for many shuffled sequences, and the number of sequences generating a particular score is plotted against the score. The resulting plot is a distribution of alignment scores occurring by chance. The alignment score for unshuffled α-hemoglobin and myoglobin (shown in red) is substantially greater than any of these scores, strongly suggesting that the sequence similarity is significant.

Distant evolutionary relationships can be detected through the use of substitution matrices

The scoring scheme heretofore described assigns points only to positions occupied by identical amino acids in the two sequences being compared. No credit is given for any pairing that is not an identity. However, as already discussed, two proteins related by evolution undergo amino acid substitutions as they diverge. A scoring system based solely on amino acid identity cannot account for these changes. To add greater sensitivity to the detection of evolutionary relationships, methods have been developed to compare two amino acids and assess their degree of similarity.

Not all substitutions are equivalent. For example, amino acid changes can be classified as structurally conservative or nonconservative. A *conservative substitution* replaces one amino acid with another that is similar in size and chemical properties. Conservative substitutions may have only minor effects on protein structure and often can be tolerated without compromising protein function. In contrast, in a *nonconservative substitution*, an amino acid is replaced by one that is structurally dissimilar. Amino acid changes can also be classified by the fewest number of nucleotide changes necessary to achieve the corresponding amino acid change. Some substitutions arise from the replacement of only a single nucleotide in the gene sequence; whereas others require two or three replacements. Conservative and single-nucleotide substitutions are likely to be more common than are substitutions with more radical effects.

How can we account for the type of substitution when comparing sequences? We can approach this problem by first examining the substitutions that have actually taken place in evolutionarily related proteins. From an examination of appropriately aligned sequences, substitution matrices have been deduced. A *substitution matrix* describes a scoring system for the replacement of any amino acid with each of the other 19 amino acids. In these matrices, a large positive score corresponds to a substitution that occurs relatively frequently, whereas a large negative score corresponds to a substitution that occurs only rarely. A commonly used substitution matrix, the Blosum-62 (for *Blo*cks of amino acid *su*bstitution *m*atrix), is illustrated in Figure 6.9. In this depiction, each column in this matrix represents one of the 20 amino acids, whereas the position of the single-letter codes within

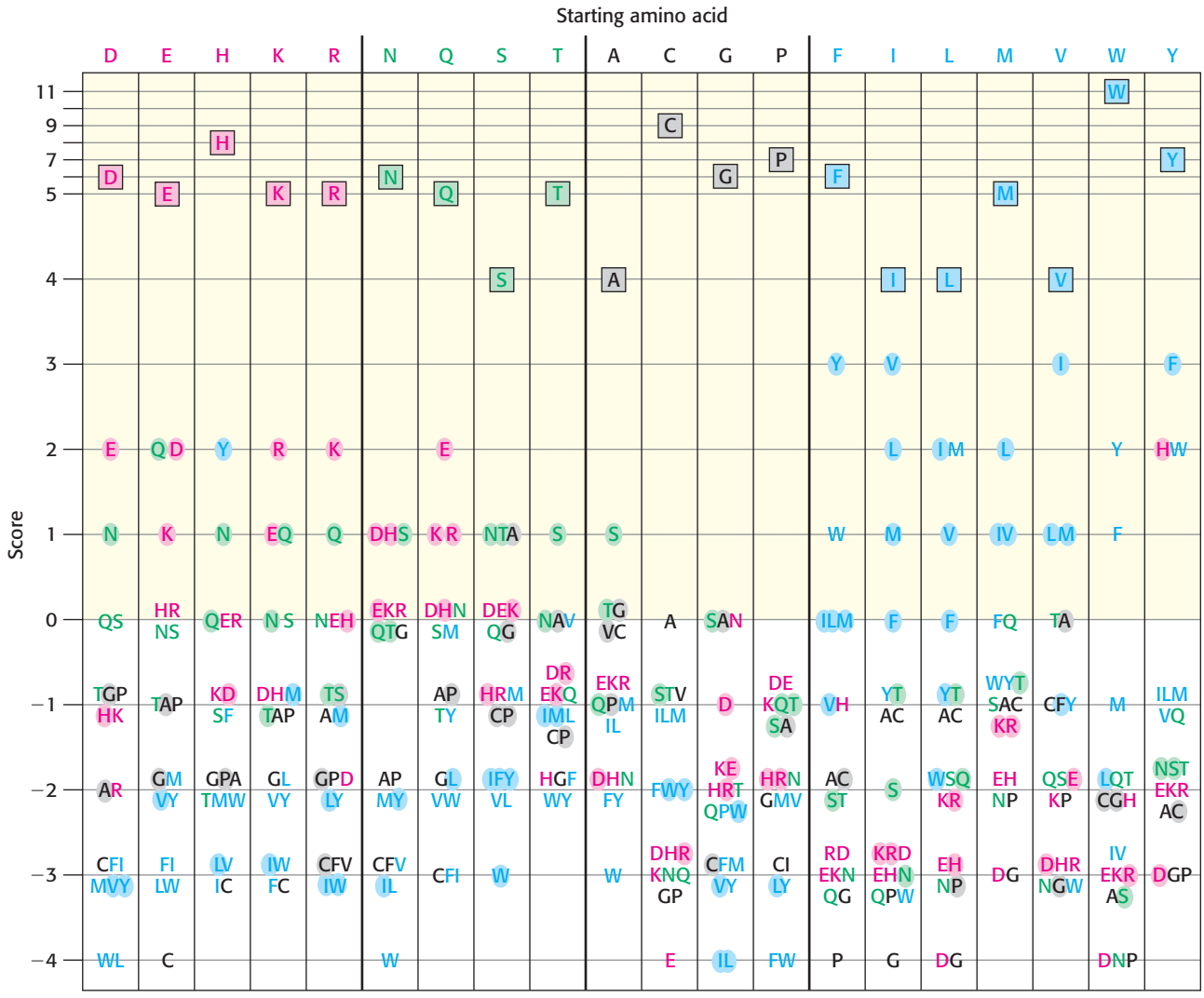


Figure 6.9 A graphic view of the Blosum-62. This substitution matrix was derived by examining substitutions within aligned sequence blocks in related proteins. Amino acids are classified into four groups (charged, red; polar, green; large and hydrophobic, blue; other, black). Substitutions that require the change of only a single nucleotide are shaded. Identities are boxed. To find the score for a substitution of, for instance, a Y for an H, you find the Y in the column having H at the top and check the number at the left. In this case, the resulting score is 2.

each column specifies the score for the corresponding substitution. Notice that scores corresponding to identity (the boxed codes at the top of each column) are not the same for each residue, owing to the fact that less frequently occurring amino acids such as cysteine (C) and tryptophan (W) will align by chance less often than the more common residues align. Furthermore, structurally conservative substitutions such as lysine (K) for arginine (R) and isoleucine (I) for valine (V) have relatively high scores, whereas nonconservative substitutions such as lysine for tryptophan result in negative scores (Figure 6.10). When two sequences are compared, each pair of aligned residues is assigned a score based on the matrix. In addition, gap penalties are often assessed. For example, the introduction of a single-residue gap lowers the alignment score by 12 points and the extension of an existing gap costs 2 points per residue. With the use of this scoring system, the alignment shown in Figure 6.6 receives a score of 115. In many regions, most substitutions are conservative (defined as those substitutions with

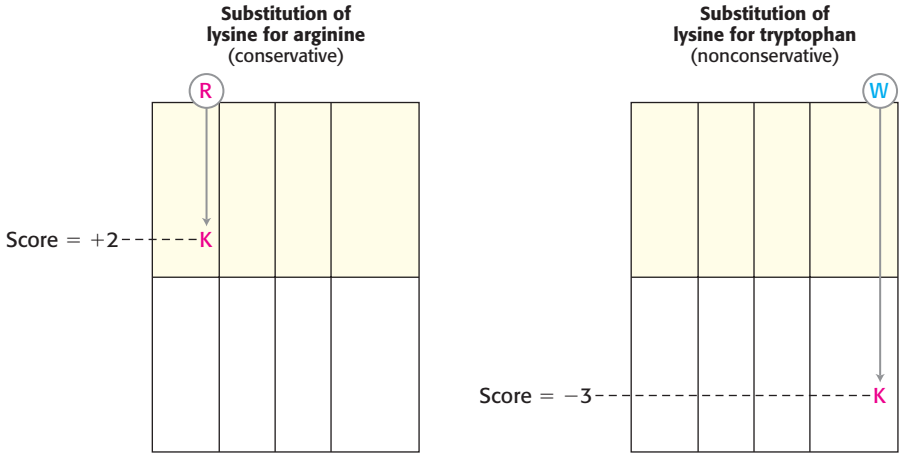


Figure 6.10 Scoring of conservative and nonconservative substitutions. The Blosum-62 indicates that a conservative substitution (lysine for arginine) receives a positive score, whereas a nonconservative substitution (lysine for tryptophan) is scored negatively. The matrix is depicted as an abbreviated form of Figure 6.9.

scores greater than 0) and relatively few are strongly disfavored types (Figure 6.11).

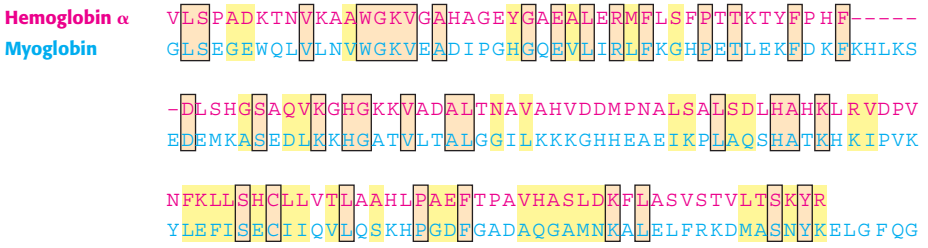


Figure 6.11 Alignment with conservative substitutions noted. The alignment of α-hemoglobin and myoglobin with conservative substitutions indicated by yellow shading and identities by orange.

This scoring system detects homology between less obviously related sequences with greater sensitivity than would a comparison of identities only. Consider, for example, the protein leghemoglobin, an oxygen-binding protein found in the roots of some plants. The amino acid sequence of leghemoglobin from the herb lupine can be aligned with that of human myoglobin and scored by using either the simple scoring scheme based on identities only or the Blosum-62 (see Figure 6.9). Repeated shuffling and scoring provides a distribution of alignment scores (Figure 6.12). Scoring based on identities only indicates that the probability of the alignment between myoglobin and leghemoglobin occurring by chance alone is 1 in 20. Thus, although the level of similarity suggests a relationship, there is a 5% chance that the similarity is accidental on the basis of this analysis. In contrast, users of the substitution matrix are able to incorporate the effects of conservative substitutions. From such an analysis, the odds of the alignment occurring by chance are calculated to be approximately 1 in 300. Thus, an analysis performed by using the substitution matrix reaches a much firmer conclusion about the evolutionary relationship between these proteins (Figure 6.13).

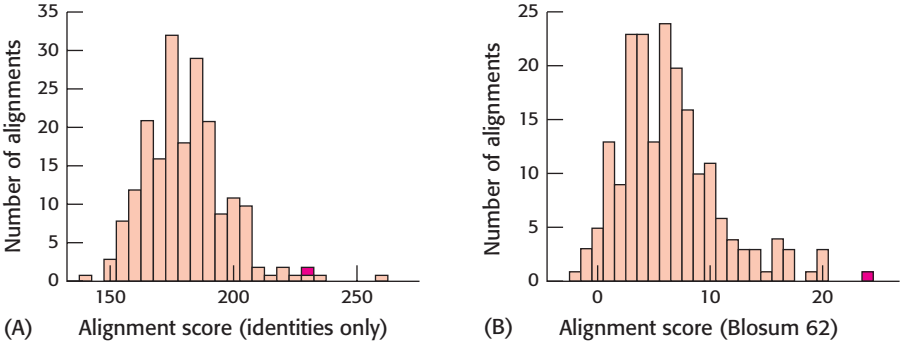


Figure 6.12 Alignment of identities only versus the Blosum-62. Repeated shuffling and scoring reveal the significance of sequence alignment for human myoglobin versus lupine leghemoglobin with the use of either (A) the simple, identity-based scoring system or (B) the Blosum-62. The scores for the alignment of the authentic sequences are shown in red. Accounting for amino acid similarity in addition to identity reveals a greater separation between the authentic alignment and the population of shuffled alignments.

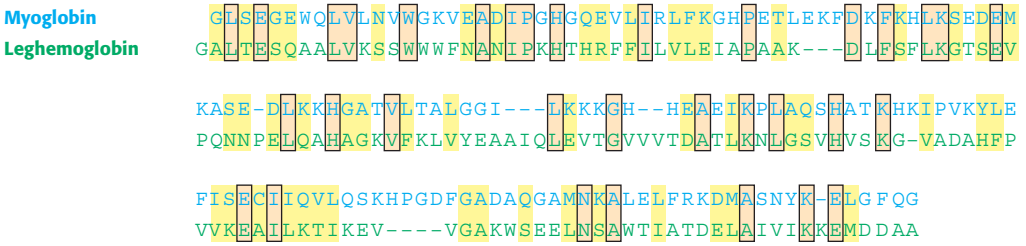


Figure 6.13 Alignment of human myoglobin and lupine leghemoglobin. The use of Blosum-62 yields the alignment shown between human myoglobin and lupine leghemoglobin, illustrating identities (orange boxes) and conservative substitutions (yellow). These sequences are 23% identical.

Experience with sequence analysis has led to the development of simpler rules of thumb. For sequences longer than 100 amino acids, sequence identities greater than 25% are almost certainly not the result of chance alone; such sequences are probably homologous. In contrast, if two sequences are less than 15% identical, their alignment alone is unlikely to indicate statistically significant similarity. For sequences that are between 15 and 25% identical, further analysis is necessary to determine the statistical significance of the alignment. It must be emphasized that *the lack of a statistically significant degree of sequence similarity does not rule out homology*. The sequences of many proteins that have descended from common ancestors have diverged to such an extent that the relationship between the proteins can no longer be detected from their sequences alone. As we will see, such homologous proteins can often be detected by examining three-dimensional structures.

Databases can be searched to identify homologous sequences

When the sequence of a protein is first determined, comparing it with all previously characterized sequences can be a source of tremendous insight into its evolutionary relatives and, hence, its structure and function. Indeed, an extensive sequence comparison is almost always the first analysis performed on a newly elucidated sequence. The sequence-alignment methods just described are used to compare an individual sequence with all members of a database of known sequences.

Database searches for homologous sequences are most often accomplished by using resources available on the Internet at the National Center for Biotechnology Information (www.ncbi.nih.gov). The procedure used is referred to as a *BLAST* (Basic Local Alignment Search Tool) *search*. An amino acid sequence is typed or pasted into the Web browser, and a search is performed, most often against a nonredundant database of all known sequences. At the end of 2009, this database included more than 10 million sequences. A BLAST search yields a list of sequence alignments, each accompanied by an estimate giving the likelihood that the alignment occurred by chance (Figure 6.14).

In 1995, investigators reported the first complete sequence of the genome of a free-living organism, the bacterium *Haemophilus influenzae*. With the sequences available, they performed a BLAST search with each deduced protein sequence. Of 1,743 identified protein-coding regions, also called *open reading frames*, 1,007 (58%) could be linked to some protein of known function that had been previously characterized in another organism. An additional 347 open reading frames could be linked to sequences in the database for which no function had yet been assigned ("hypothetical proteins"). The remaining 389 sequences did not match any sequence present in the database at that time. Thus, investigators were able to identify likely functions for more than half the proteins within this organism solely by sequence comparisons.

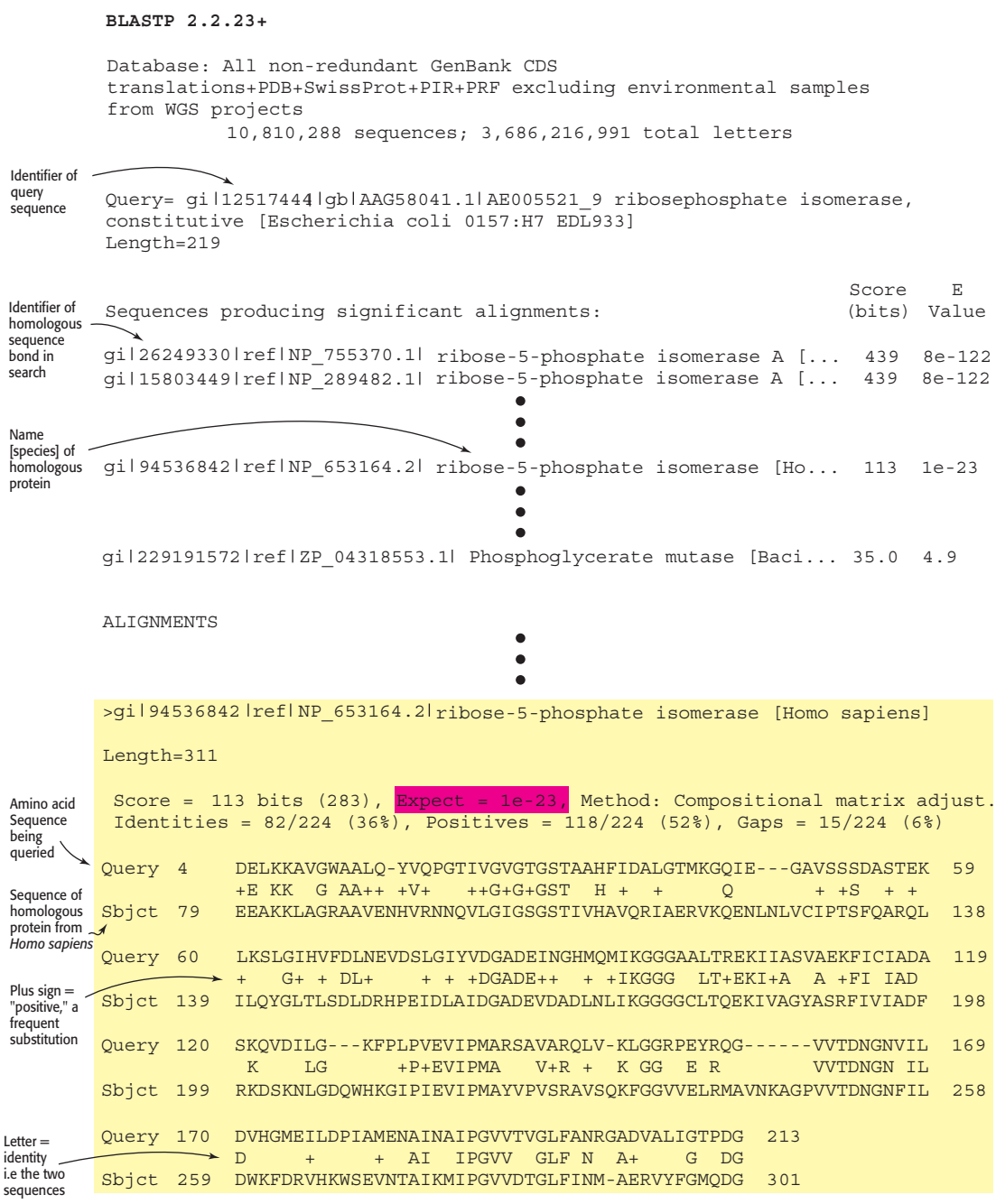


Figure 6.14 BLAST search results. Part of the results from a BLAST search of the nonredundant (nr) protein sequence database using the sequence of ribose 5-phosphate isomerase (also called phosphopentose isomerase, Chapter 20) from *E. coli* as a query. Among the thousands of sequences found is the orthologous sequence from human beings, and the alignment between these sequences is shown (highlighted in yellow). The number of sequences with this level of similarity expected to be in the database by chance is 1×10^{-23} as shown by the E value (highlighted in red). Because this value is much less than 1, the observed sequence alignment is highly significant.

6.3 Examination of Three-Dimensional Structure Enhances Our Understanding of Evolutionary Relationships

Sequence comparison is a powerful tool for extending our knowledge of protein function and kinship. However, biomolecules generally function as intricate three-dimensional structures rather than as linear polymers. Mutations occur at the level of sequence, but the effects of the mutations are at the level of function, and function is directly related to tertiary structure. Consequently, to gain a deeper understanding of evolutionary relationships

between proteins, we must examine three-dimensional structures, especially in conjunction with sequence information. The techniques of structural determination are presented in Chapter 3.

Tertiary structure is more conserved than primary structure

Because three-dimensional structure is much more closely associated with function than is sequence, tertiary structure is more evolutionarily conserved than is primary structure. This conservation is apparent in the tertiary structures of the globins (Figure 6.15), which are extremely similar even though the similarity between human myoglobin and lupine leghemoglobin is just barely detectable at the sequence level and that between human α-hemoglobin and lupine leghemoglobin is not statistically significant (15.6% identity). This structural similarity firmly establishes that the framework that binds the heme group and facilitates the reversible binding of oxygen has been conserved over a long evolutionary period.

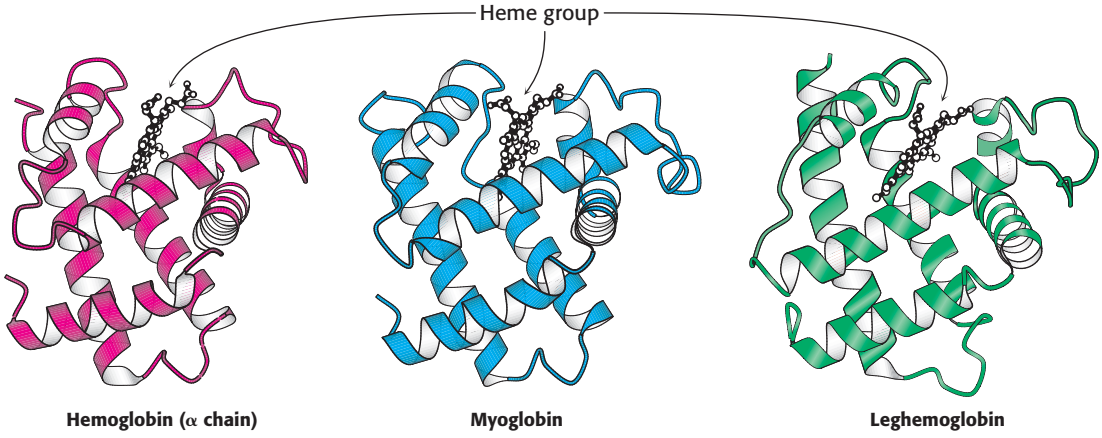


Figure 6.15 Conservation of three-dimensional structure. The tertiary structures of human hemoglobin (α chain), human myoglobin, and lupine leghemoglobin are conserved. Each heme group contains an iron atom to which oxygen binds. [Drawn from 1HBB.pdb, 1MBD.pdb, and 1GDJ.pdb.]

Anyone aware of the similar biochemical functions of hemoglobin, myoglobin, and leghemoglobin could expect the structural similarities. In a growing number of other cases, however, a comparison of three-dimensional structures has revealed striking similarities between proteins that were *not* expected to be related, on the basis of their diverse functions. A case in point is the protein actin, a major component of the cytoskeleton (Section 35.2), and heat shock protein 70 (Hsp-70), which assists protein folding inside cells. These two proteins were found to be noticeably similar in structure despite only 15.6% sequence identity (Figure 6.16). On the basis of their

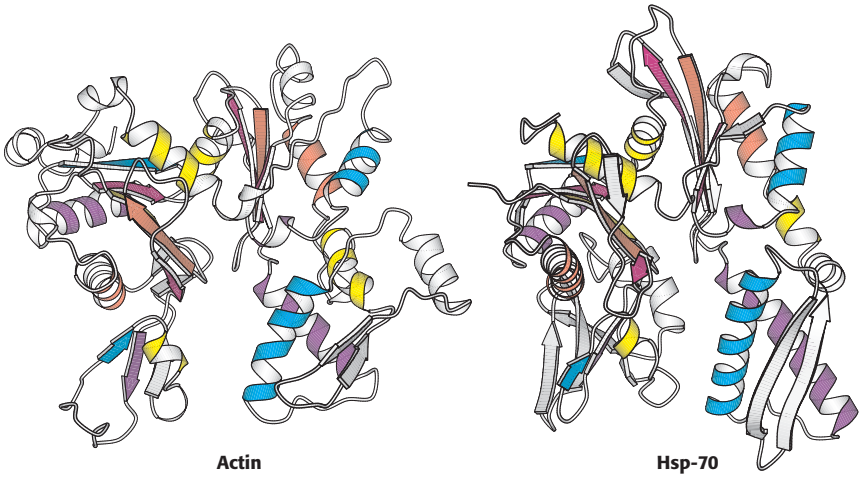


Figure 6.16 Structures of actin and a large fragment of heat shock protein 70 (Hsp-70). A comparison of the identically colored elements of secondary structure reveals the overall similarity in structure despite the difference in biochemical activities. [Drawn from 1ATN.pdb and 1ATR.pdb.]

three-dimensional structures, actin and Hsp-70 are paralogs. The level of structural similarity strongly suggests that, despite their different biological roles in modern organisms, these proteins descended from a common ancestor. As the three-dimensional structures of more proteins are determined, such unexpected kinships are being discovered with increasing frequency. The search for such kinships relies ever more frequently on computer-based searches that are able to compare the three-dimensional structure of any protein with all other known structures.

Knowledge of three-dimensional structures can aid in the evaluation of sequence alignments

The sequence-comparison methods described thus far treat all positions within a sequence equally. However, we know from examining families of homologous proteins for which at least one three-dimensional structure is known that regions and residues critical to protein function are more strongly conserved than are other residues. For example, each type of globin contains a bound heme group with an iron atom at its center. A histidine residue that interacts directly with this iron atom (residue 64 in human myoglobin) is conserved in all globins. After we have identified key residues or highly conserved sequences within a family of proteins, we can sometimes identify other family members even when the overall level of sequence similarity is below statistical significance. Thus it may be useful to generate a *sequence template*—a map of conserved residues that are structurally and functionally important and are characteristic of particular families of proteins, which makes it possible to recognize new family members that might be undetectable by other means. A variety of other methods for sequence classification that take advantage of known three-dimensional structures also are being developed. Still other methods are able to identify conserved residues within a family of homologous proteins, even without a known three-dimensional structure. These methods often use substitution matrices that differ at each position within a family of aligned sequences. Such methods can often detect quite distant evolutionary relationships.

Repeated motifs can be detected by aligning sequences with themselves

More than 10% of all proteins contain sets of two or more domains that are similar to one another. Sequence search methods can often detect internally repeated sequences that have been characterized in other proteins. Often, however, repeated units do not correspond to previously identified domains. In these cases, their presence can be detected by attempting to align a given sequence with itself. The statistical significance of such repeats can be tested by aligning the regions in question as if these regions were sequences from separate proteins. For the TATA-box-binding protein, a key protein in controlling gene transcription (Section 29.2), such an alignment is highly significant: 30% of the amino acids are identical over 90 residues (Figure 6.17A). The estimated probability of such an alignment occurring by chance is 1 in 10^{13}. The determination of the three-dimensional structure of the TATA-box-binding protein confirmed the presence of repeated structures; the protein is formed of two nearly identical domains (Figure 6.17B). The evidence is convincing that the gene encoding this protein evolved by duplication of a gene encoding a single domain.

(A)

```
  1  MTDQGLEGSNPVDLSKHPS

 20  GIVPTLQNIVSTVNLDCKLDLKAIALQ-ARNAEYNPKRFAAVIMRIR
110  FKDFKIQNIVGSCDVKFPIRLEGLAYSHAAFSSYEPELFPGLIYRMK

 66  EPKTTALIFASGKMVCTGAKSEDFSKMAARKYARIVQKLGFPAK
157  VPKIVLLIFVSGKIVITGAKMRDETYKAFENIYPVLSEFRKIQQ
```

(B)

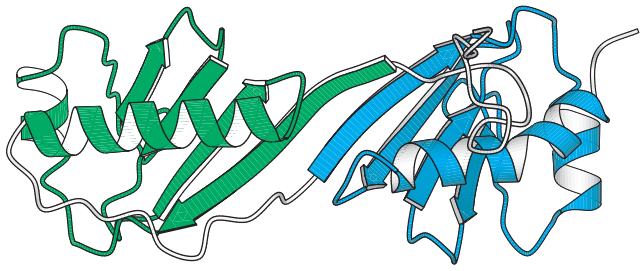


Figure 6.17 Sequence alignment of internal repeats. (A) An alignment of the sequences of the two repeats of the TATA-box-binding protein. The amino-terminal repeat is shown in green and the carboxyl-terminal repeat in blue. (B) Structure of the TATA-box-binding protein. The amino-terminal domain is shown in green and the carboxyl-terminal domain in blue. [Drawn from 1VOK.pdb.]

Convergent evolution illustrates common solutions to biochemical challenges

Thus far, we have been exploring proteins derived from common ancestors—that is, through *divergent evolution*. Other cases have been found of proteins that are structurally similar in important ways but are not descended from a common ancestor. How might two unrelated proteins come to resemble each other structurally? Two proteins evolving independently may have converged on a similar structure to perform a similar biochemical activity. Perhaps that structure was an especially effective solution to a biochemical problem that organisms face. The process by which very different evolutionary pathways lead to the same solution is called *convergent evolution*.

An example of convergent evolution is found among the serine proteases. These enzymes, to be considered in more detail in Chapter 9, cleave peptide bonds by hydrolysis. Figure 6.18 shows the structure of the active sites—

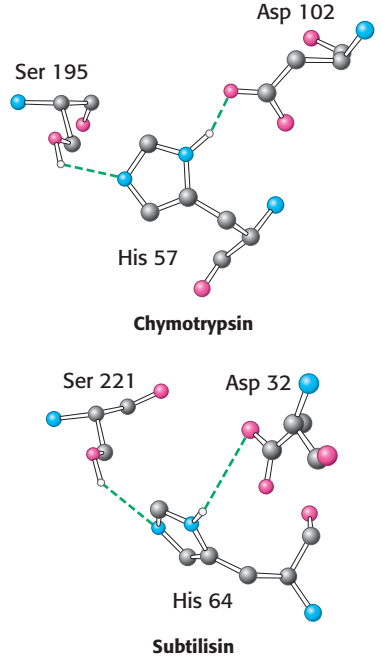


Figure 6.18 Convergent evolution of protease active sites. The relative positions of the three key residues shown are nearly identical in the active sites of the serine proteases chymotrypsin and subtilisin.

that is, the sites on the proteins at which the hydrolysis reaction takes place—for two such enzymes, chymotrypsin and subtilisin. These active-site structures are remarkably similar. In each case, a serine residue, a histidine residue, and an aspartic acid residue are positioned in space in nearly identical arrangements. As we will see, this conserved spatial arrangement is critical for the activity of these enzymes and affords the same mechanistic solution to the problem of peptide hydrolysis. At first glance, this similarity might suggest that these proteins are homologous. However, striking differences in the overall structures of these proteins make an evolutionary relationship extremely unlikely (Figure 6.19). Whereas chymotrypsin consists almost entirely of β sheets, subtilisin contains extensive α-helical structure. Moreover, the key serine, histidine, and aspartic acid residues do not occupy similar positions or even appear in the same order within the two sequences. It is extremely unlikely that two proteins evolving from a common ancestor could have retained similar active-site structures while other aspects of the structure changed so dramatically.

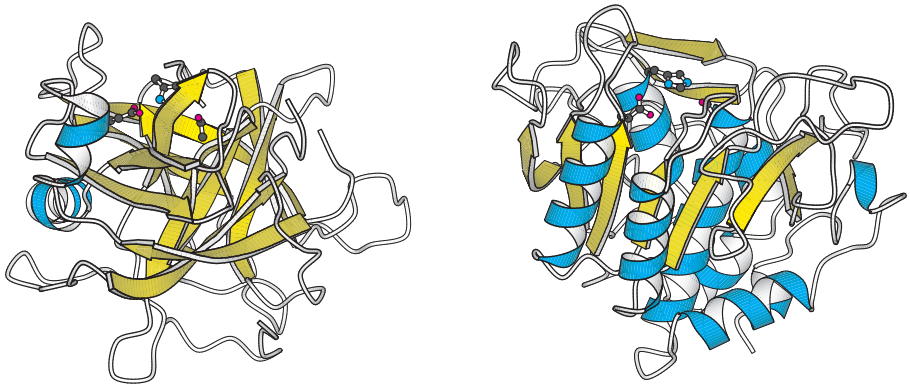
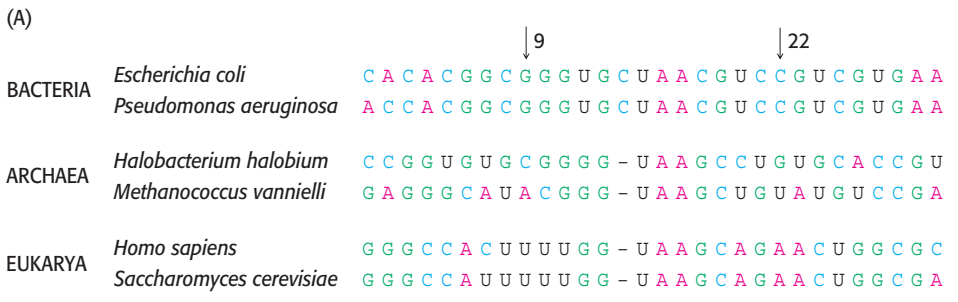


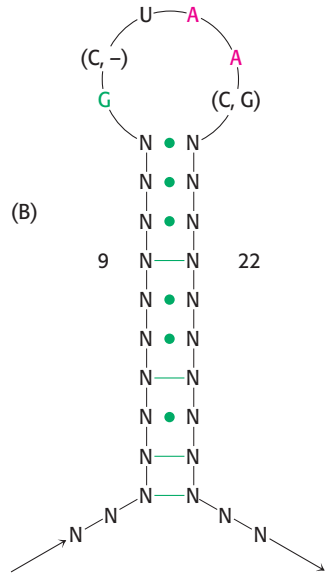
Figure 6.19 Structures of mammalian chymotrypsin and bacterial subtilisin. The overall structures are quite dissimilar, in stark contrast with the active sites, shown at the top of each structure. The β strands are shown in yellow and the α helices in blue. [Drawn from 1GCT.pdb. and 1SUP.pdb.]

Comparison of RNA sequences can be a source of insight into RNA secondary structures

Homologous RNA sequences can be compared in a manner similar to that already described for protein sequences. Such comparisons can be a source of important insights into evolutionary relationships; in addition, they provide clues to the three-dimensional structure of the RNA itself. As noted in Chapter 4, single-stranded nucleic acid molecules fold back on themselves to form elaborate structures held together by Watson–Crick base-pairing and other interactions. In a family of sequences that form similar base-paired structures, base sequences may vary, but base-pairing ability is conserved. Consider, for example, a region from a large RNA molecule present in the ribosomes of all organisms (Figure 6.20). In the region shown, the *E. coli* sequence has a guanine (G) residue in position 9 and a cytosine (C) residue in position 22, whereas the human sequence has uracil (U) in position 9 and adenine (A) in position 22. Examination of the six sequences shown in Figure 6.20 reveals that the bases in positions 9 and 22, as well as several of the neighboring positions, retain the ability to form Watson–Crick base pairs even though the identities of the bases in these positions vary. We can deduce that two segments with paired mutations that maintain base-pairing ability are likely to form a double helix. Where sequences are known for several homologous RNA molecules, this type of sequence analysis can often suggest complete secondary structures as well as some additional

Figure 6.20 Comparison of RNA sequences. (A) A comparison of sequences in a part of ribosomal RNA taken from a variety of species. (B) The implied secondary structure. Green bars indicate positions at which Watson–Crick base-pairing is completely conserved in the sequences shown, whereas dots indicate positions at which Watson–Crick base-pairing is conserved in most cases.





interactions. For this particular ribosomal RNA, the subsequent determination of its three-dimensional structure (Section 30.3) confirmed the predicted secondary structure.

6.4 Evolutionary Trees Can Be Constructed on the Basis of Sequence Information

The observation that homology is often manifested as sequence similarity suggests that the evolutionary pathway relating the members of a family of proteins may be deduced by examination of sequence similarity. This approach is based on the notion that sequences that are more similar to one another have had less evolutionary time to diverge than have sequences that are less similar. This method can be illustrated by using the three globin sequences in Figures 6.11 and 6.13, as well as the sequence for the human hemoglobin β chain. These sequences can be aligned with the additional constraint that gaps, if present, should be at the same positions in all of the proteins. These aligned sequences can be used to construct an *evolutionary tree* in which the length of the branch connecting each pair of proteins is proportional to the number of amino acid differences between the sequences (Figure 6.21).

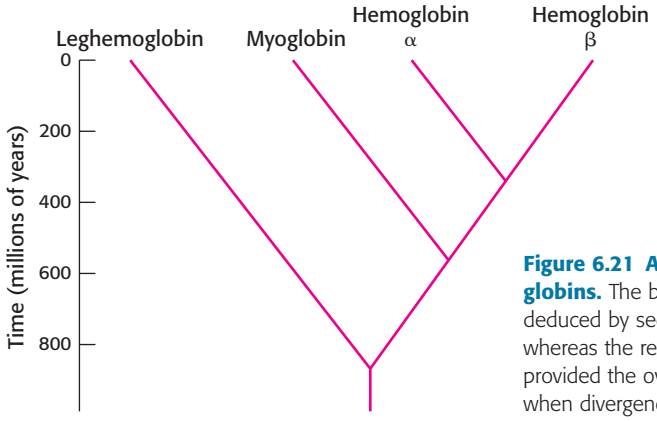


Figure 6.21 An evolutionary tree for globins. The branching structure was deduced by sequence comparison, whereas the results of fossil studies provided the overall time scale showing when divergence occurred.

Such comparisons reveal only the relative divergence times—for example, that myoglobin diverged from hemoglobin twice as long ago as the α chain diverged from the β chain. How can we estimate the approximate dates of gene duplications and other evolutionary events? Evolutionary trees can be calibrated by comparing the deduced branch points with divergence times determined from the fossil record. For example, the duplication leading to the two chains of hemoglobin appears to have occurred 350 million years ago. This estimate is supported by the observation that jawless fish such as the lamprey, which diverged from bony fish approximately 400 million years ago, contain hemoglobin built from a single type of subunit (Figure 6.22).

These methods can be applied to both relatively modern and very ancient molecules, such as the ribosomal RNAs that are found in all organisms. Indeed, such an RNA sequence analysis led to the realization that Archaea are a distinct group of organisms that diverged from Bacteria very early in evolutionary history.

Figure 6.22 The lamprey. A jawless fish whose ancestors diverged from bony fish approximately 400 million years ago, the lamprey contains hemoglobin molecules that contain only a single type of polypeptide chain. [Brent P. Kent.]

6.5 Modern Techniques Make the Experimental Exploration of Evolution Possible

Two techniques of biochemistry have made it possible to examine the course of evolution more directly and not simply by inference. The polymerase chain reaction (Chapter 5) allows the direct examination of ancient DNA sequences, releasing us, at least in some cases, from the constraints of being able to examine existing genomes from living organisms only. Molecular evolution may be investigated through the use of *combinatorial chemistry,* the process of producing large populations of molecules en masse and selecting for a biochemical property. This exciting process provides a glimpse into the types of molecules that may have existed very early in evolution.

Ancient DNA can sometimes be amplified and sequenced

The tremendous chemical stability of DNA makes the molecule well suited to its role as the storage site of genetic information. So stable is the molecule that samples of DNA have survived for many thousands of years under appropriate conditions. With the development of PCR and advanced DNA-sequencing methods, such ancient DNA can be amplified and sequenced. This approach has been applied to mitochondrial DNA from a Neanderthal fossil estimated at 38,000 years of age excavated from Vindija Cave, Croatia, in 1980. Remarkably, investigators have completely sequenced the mitochondrial genome from this specimen. Comparison of

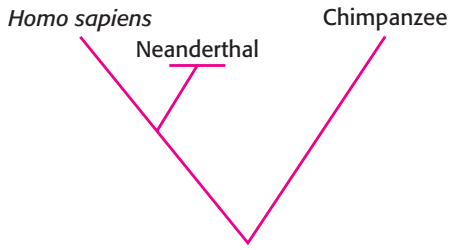


Figure 6.23 Placing Neanderthal on an evolutionary tree. Comparison of DNA sequences revealed that Neanderthal is not on the line of direct descent leading to *Homo sapiens* but, instead, branched off earlier and then became extinct.

the Neanderthal mitochondrial sequence with those from *Homo sapiens* individuals revealed between 201 and 234 substitutions, considerably fewer than the approximately 1,500 differences between human beings and chimpanzees over the same region. Further analysis suggested that the common ancestor of modern human beings and Neanderthals lived approximately 660,000 years ago. An evolutionary tree constructed from these data has revealed that the Neanderthal was not an intermediate between chimpanzees and human beings but, instead, was an evolutionary "dead end" that became extinct (Figure 6.23).

A few earlier studies claimed to determine the sequences of far more ancient DNA such as that found in insects trapped in amber, but these studies appear to have been flawed. The source of these sequences turned out to be contaminating modern DNA. Successful sequencing of ancient DNA requires sufficient DNA for reliable amplification and the rigorous exclusion of all sources of contamination.

Molecular evolution can be examined experimentally

Evolution requires three processes: (1) the generation of a diverse population, (2) the selection of members based on some criterion of fitness, and (3) reproduction to enrich the population in these more-fit members. Nucleic acid molecules are capable of undergoing all three processes in vitro under appropriate conditions. The results of such studies enable us to glimpse how evolutionary processes might have generated catalytic activities and specific binding abilities—important biochemical functions in all living systems.

A diverse population of nucleic acid molecules can be synthesized in the laboratory by the process of combinatorial chemistry, which rapidly produces large populations of a particular type of molecule such as a nucleic acid. A population of molecules of a given size can be generated randomly so that many or all possible sequences are present in the mixture. When an initial population has been generated, it is subjected to a selection process that isolates specific molecules with desired binding or reactivity properties. Finally, molecules that have survived the selection process are replicated through the use of PCR; primers are directed toward specific sequences included at the ends of each member of the population. Errors that occur naturally in the course of the replication process introduce additional variation into the population in each "generation."

Let us consider an application of this approach. Early in evolution, before the emergence of proteins, RNA molecules may have played all major roles in biological catalysis. To understand the properties of potential RNA catalysts, researchers have used the methods heretofore described to create an RNA molecule capable of binding adenosine triphosphate and related nucleotides. An initial population of RNA molecules 169 nucleotides long was created; 120 of the positions differed randomly, with equimolar mixtures of adenine, cytosine, guanine, and uracil. The initial

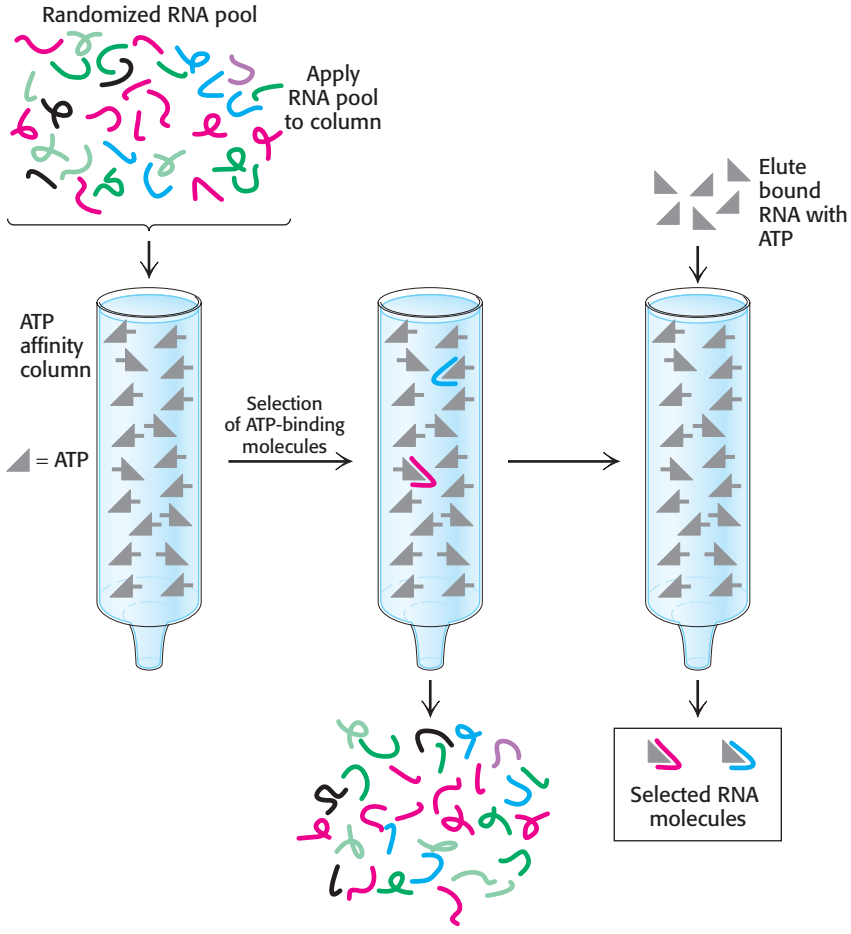


Figure 6.24 Evolution in the laboratory. A collection of RNA molecules of random sequences is synthesized by combinatorial chemistry. This collection is selected for the ability to bind ATP by passing the RNA through an ATP affinity column (Section 3.1). The ATP-binding RNA molecules are released from the column by washing with excess ATP and then replicated. The process of selection and replication is then repeated several times. The final RNA products with significant ATP-binding ability are isolated and characterized.

synthetic pool that was used contained approximately 10^{14} RNA molecules. Note that this number is a very small fraction of the total possible pool of random 120-base sequences. From this pool, those molecules that bound to ATP, which had been immobilized on a column, were selected (Figure 6.24).

The collection of molecules that were bound well by the ATP affinity column were allowed to replicate by reverse transcription into DNA, amplification by PCR, and transcription back into RNA. The somewhat error-prone replication processes introduced additional mutations into the population in each cycle. The new population was subjected to additional rounds of selection for ATP-binding activity. After eight generations, members of the selected population were characterized by sequencing. Seventeen different sequences were obtained, 16 of which could form the structure shown in Figure 6.25. Each of these molecules bound ATP with dissociation constants less than 50 μM.

The folded structure of the ATP-binding region from one of these RNAs was determined by nuclear magnetic resonance (Section 3.6) methods (Figure 6.26). As expected, this 40-nucleotide molecule is composed of two Watson–Crick base-paired helical regions separated by an 11-nucleotide

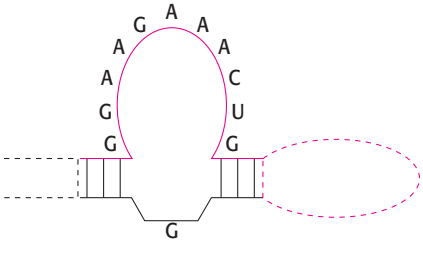


Figure 6.25 A conserved secondary structure. The secondary structure shown is common to RNA molecules selected for ATP binding.

loop. This loop folds back on itself in an intricate way to form a deep pocket into which the adenine ring can fit. Thus, a structure had evolved that was capable of a specific interaction.

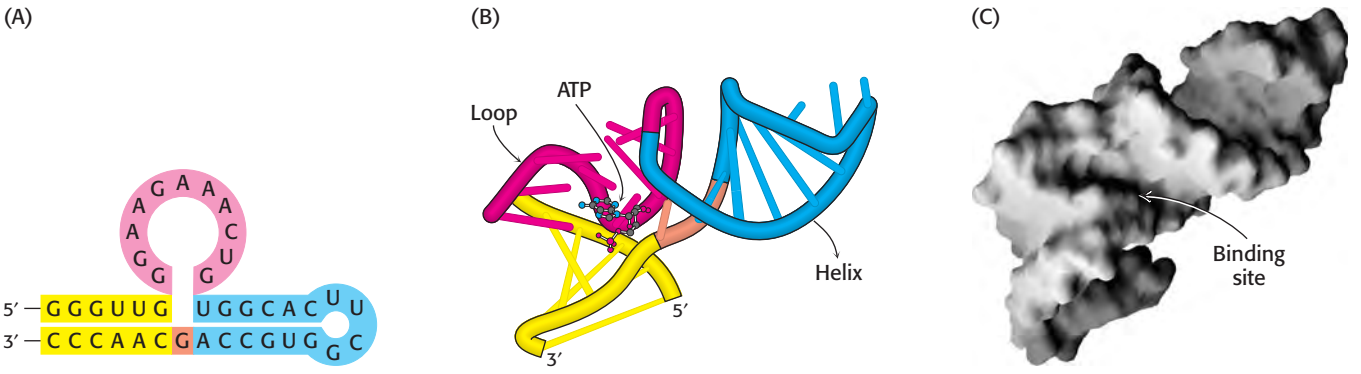


Figure 6.26 An evolved ATP-binding RNA molecule. (A) The Watson–Crick base-pairing pattern, (B) the folding pattern, and (C) a surface representation of an RNA molecule selected to bind adenosine nucleotides. The bound ATP is shown in part B, and the binding site is revealed as a deep pocket in part C.

Summary

6.1 Homologs Are Descended from a Common Ancestor

Exploring evolution biochemically often means searching for homology between molecules, because homologous molecules, or homologs, evolved from a common ancestor. Paralogs are homologous molecules that are found in one species and have acquired different functions through evolutionary time. Orthologs are homologous molecules that are found in different species and have similar or identical functions.

6.2 Statistical Analysis of Sequence Alignments Can Detect Homology

Protein and nucleic acid sequences are two of the primary languages of biochemistry. Sequence-alignment methods are the most powerful tools of the evolutionary detective. Sequences can be aligned to maximize their similarity, and the significance of these alignments can be judged by statistical tests. The detection of a statistically significant alignment between two sequences strongly suggests that two sequences are related by divergent evolution from a common ancestor. The use of substitution matrices makes the detection of more-distant evolutionary relationships possible. Any sequence can be used to probe sequence databases to identify related sequences present in the same organism or in other organisms.

6.3 Examination of Three-Dimensional Structure Enhances Our Understanding of Evolutionary Relationships

The evolutionary kinship between proteins may be even more strikingly evident in the conserved three-dimensional structures. The analysis of three-dimensional structure in combination with the analysis of especially conserved sequences has made it possible to determine evolutionary relationships that cannot be detected by other means. Sequence-comparison methods can also be used to detect imperfectly repeated sequences within a protein, indicative of linked similar domains.

6.4 Evolutionary Trees Can Be Constructed on the Basis of Sequence Information

Evolutionary trees can be constructed with the assumption that the number of sequence differences corresponds to the time since the two sequences diverged. Construction of an evolutionary tree based

on sequence comparisons revealed approximate times for the gene-duplication events separating myoglobin and hemoglobin as well as the α and β subunits of hemoglobin. Evolutionary trees based on sequences can be compared with those based on fossil records.

6.5 Modern Techniques Make the Experimental Exploration of Evolution Possible

The exploration of evolution can also be a laboratory science. In favorable cases, PCR amplification of well-preserved samples allows the determination of nucleotide sequences from extinct organisms. Sequences so determined can help authenticate parts of an evolutionary tree constructed by other means. Molecular evolutionary experiments performed in the test tube can examine how molecules such as ligand-binding RNA molecules might have been generated.

Key Terms

homolog (p. 174)
paralog (p. 174)
ortholog (p. 174)
sequence alignment (p. 176)
conservative substitution (p. 178)
substitution matrix (p. 178)
BLAST search (p. 181)
sequence template (p. 184)
divergent evolution (p. 185)
convergent evolution (p. 185)
evolutionary tree (p. 187)
combinatorial chemistry (p. 188)

Problems

1. *What's the score?* Using the identity-based scoring system (Section 6.2), calculate the score for the following alignment. Do you think the score is statistically significant?

(1) WYLGKITRMDAEVLLKKPTVRDGHFLVTQCESSPGEF-
(2) WYFGKITRRESERLLLNPENPRGTFLVRESETTKGAY-

SISVRFGDSVQ-----HFKVLRDQNGKYYLWAVK-FN-
CLSVSDFDNAKGLNVKHYKIRKLDSGGFYITSRTQFS-

SLNELVAYHRTASVSRTHTILLSDMNV
SSLQQLVAYYSKHADGLCHRLTNV

2. *Sequence and structure.* A comparison of the aligned amino acid sequences of two proteins each consisting of 150 amino acids reveals them to be only 8% identical. However, their three-dimensional structures are very similar. Are these two proteins related evolutionarily? Explain.

3. *It depends on how you count.* Consider the following two sequence alignments:

(a) A-SNLFDIRLIG
GSNDFYEVKIMD

(b) ASNLFDIRLI-G
GSNDFYEVKIMD

Which alignment has a higher score if the identity-based scoring system (Section 6.2) is used? Which alignment has a higher score if the Blosum-62 substitution matrix (Figure 6.9) is used?

4. *Discovering a new base pair.* Examine the ribosomal RNA sequences in Figure 6.20. In sequences that do not contain Watson–Crick base pairs, what base tends to be paired with G? Propose a structure for your new base pair.

5. *Overwhelmed by numbers.* Suppose that you wish to synthesize a pool of RNA molecules that contain all four bases at each of 40 positions. How much RNA must you have in grams if the pool is to have at least a single molecule of each sequence? The average molecular weight of a nucleotide is 330 g mol^{-1}.

6. *Form follows function.* The three-dimensional structure of biomolecules is more conserved evolutionarily than is sequence. Why?

7. *Shuffling.* Using the identity-based scoring system (Section 6.2), calculate the alignment score for the alignment of the following two short sequences:

(1) ASNFLDKAGK
(2) ATDYLEKAGK

Generate a shuffled version of sequence 2 by randomly reordering these 10 amino acids. Align your shuffled sequence with sequence 1 without allowing gaps, and calculate the alignment score between sequence 1 and your shuffled sequence.

8. *Interpreting the score.* Suppose that the sequences of two proteins each consisting of 200 amino acids are aligned and that the percentage of identical residues has been calculated. How would you interpret each of the following results in

regard to the possible divergence of the two proteins from a common ancestor? (a) 80%, (b) 50%, (c) 20%, (d) 10%.

9. *Particularly unique.* Consider the Blosum-62 matrix in Figure 6.9. Replacement of which three amino acids never yields a positive score? What features of these residues might contribute to this observation?

10. *A set of three.* The sequences of three proteins (A, B, and C) are compared with one another, yielding the following levels of identity:

	A	B	C
A	100%	65%	15%
B	65%	100%	55%
C	15%	55%	100%

Assume that the sequence matches are distributed uniformly along each aligned sequence pair. Would you expect protein A and protein C to have similar three-dimensional structures? Explain.

11. *RNA alignment.* Sequences of an RNA fragment from five species have been determined and aligned. Propose a likely secondary structure for these fragments.

(1) UUGGAGAUUCGGUAGAAUCUCCC
(2) GCCGGGAAUCGACAGAUUCCCCG
(3) CCCAAGUCCCGGCAGGGACUUAC
(4) CUCACCUGCCGAUAGGCAGGUCA
(5) AAUACCACCCGGUAGGGUGGUUC

12. *The more the merrier.* When RNA alignments are used to determine secondary structure, it is advantageous to have many sequences representing a wide variety of species. Why?

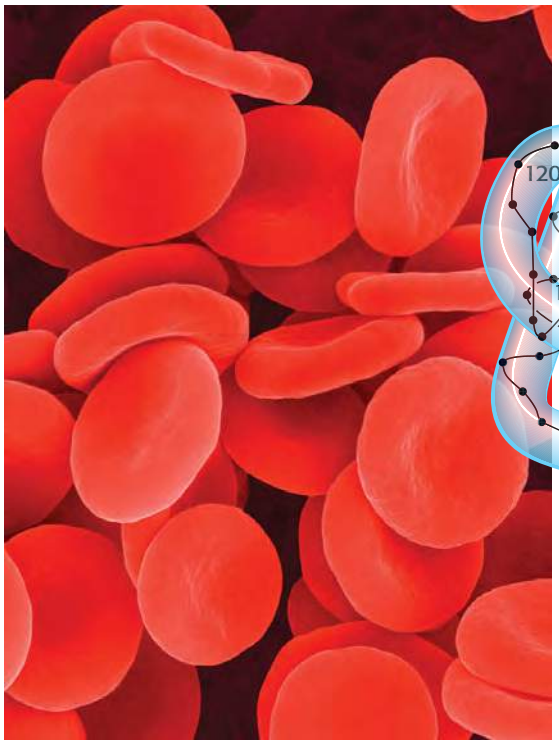
13. *To err is human.* You have discovered a mutant form of a thermostable DNA polymerase with significantly reduced fidelity in adding the appropriate nucleotide to the growing DNA strand, compared with wild-type DNA polymerase. How might this mutant be useful in the molecular-evolution experiments described in Section 6.5?

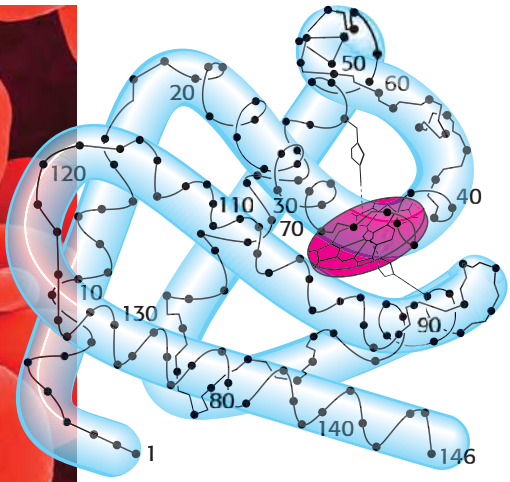
14. *Generation to generation.* When performing a molecular-evolution experiment, such as that described in Section 6.5, why is it important to repeat the selection and replication steps for several generations?

15. *BLAST away.* Using the National Center for Biotechnology Information Web site (www.ncbi.nlm.nih.gov), find the sequence of the enzyme triose phosphate isomerase from *E. coli*. Use this sequence as the query for a protein–protein BLAST search. In the output, find the alignment with the sequence of triose phosphate isomerase from human beings (*Homo sapiens*). How many identities are observed in the alignment?

This page intentionally left blank

Hemoglobin: Portrait of a Protein in Action





Beta chain of hemoglobin

In the bloodstream, red cells carry oxygen from the lungs to the tissues, where demand is high. Hemoglobin, the protein that gives blood its red color, is responsible for the transport of oxygen via its four heme-bound subunits. Hemoglobin was one of the first proteins to have its structure determined; the folding of a single subunit is shown in this hand-drawn view. [Left, Dr. Dennis Kunkel/Visuals Unlimited.]

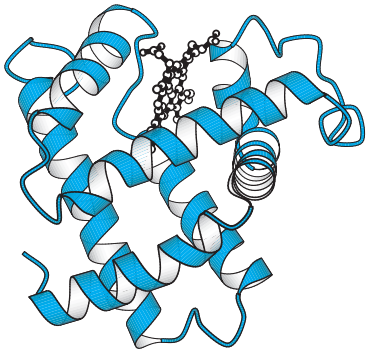
The transition from anaerobic to aerobic life was a major step in evolution because it uncovered a rich reservoir of energy. Fifteen times as much energy is extracted from glucose in the presence of oxygen than in its absence. For single-celled and other small organisms, oxygen can be absorbed into actively metabolizing cells directly from the air or surrounding water. Vertebrates evolved two principal mechanisms for supplying their cells with an adequate supply of oxygen. The first is a circulatory system that actively delivers oxygen to cells throughout the body. The second is the use of the oxygen-transport and oxygen-storage proteins, hemoglobin and myoglobin. Hemoglobin, which is contained in red blood cells, is a fascinating protein, efficiently carrying oxygen from the lungs to the tissues while also contributing to the transport of carbon dioxide and hydrogen ions back to the lungs. Myoglobin, located in muscle, provides a reserve supply of oxygen available in time of need.

A comparison of myoglobin and hemoglobin illuminates some key aspects of protein structure and function. These two evolutionarily related proteins employ nearly identical structures for oxygen binding (Chapter 6). However, hemoglobin is a remarkably efficient oxygen carrier, able to use as much as 90% of its potential oxygen-carrying capacity effectively. Under similar conditions, myoglobin would be able to use only 7% of its potential

OUTLINE

capacity. What accounts for this dramatic difference? Myoglobin exists as a single polypeptide, whereas hemoglobin comprises four polypeptide chains. The four chains in hemoglobin bind oxygen *cooperatively,* meaning that the binding of oxygen to a site in one chain increases the likelihood that the remaining chains will bind oxygen. Furthermore, the oxygen-binding properties of hemoglobin are modulated by the binding of hydrogen ions and carbon dioxide in a manner that enhances oxygen-carrying capacity. Both cooperativity and the response to modulators are made possible by variations in the quaternary structure of hemoglobin when different combinations of molecules are bound.

Hemoglobin and myoglobin have played important roles in the history of biochemistry. They were the first proteins for which three-dimensional structures were determined by x-ray crystallography. Furthermore, the possibility that variations in protein sequence could lead to disease was first proposed and demonstrated for sickle-cell anemia, a blood disease caused by mutation of a single amino acid in one hemoglobin chain. Hemoglobin has been and continues to be a valuable source of knowledge and insight, both in itself and as a prototype for many other proteins that we will encounter throughout our study of biochemistry.



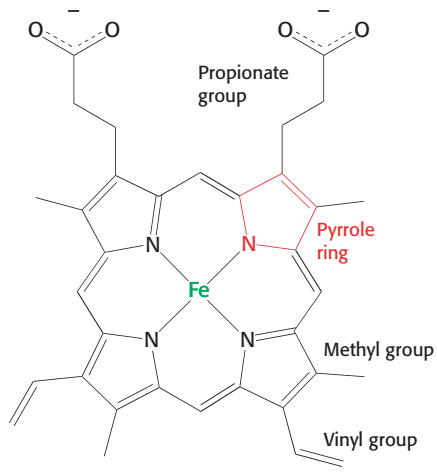
Myoglobin

Figure 7.1 Structure of myoglobin. *Notice* that myoglobin consists of a single polypeptide chain, formed of α helices connected by turns, with one oxygen-binding site. [Drawn from 1MBD.pdb.]

7.1 Myoglobin and Hemoglobin Bind Oxygen at Iron Atoms in Heme

Sperm whale myoglobin was the first protein for which the three-dimensional structure was determined. X-ray crystallographic studies pioneered by John Kendrew revealed the structure of this protein in the 1950s (Figure 7.1). Myoglobin consists largely of α helices that are linked to one another by turns to form a globular structure.

Myoglobin can exist in an oxygen-free form called *deoxymyoglobin* or in a form with an oxygen molecule bound called *oxymyoglobin*. The ability of myoglobin and hemoglobin to bind oxygen depends on the presence of a bound prosthetic group called *heme*.



Heme (Fe-protoporphyrin IX)

The heme group gives muscle and blood their distinctive red color. It consists of an organic component and a central iron atom. The organic component, called *protoporphyrin,* is made up of four pyrrole rings linked by methine bridges to form a tetrapyrrole ring. Four methyl groups, two vinyl groups, and two propionate side chains are attached.

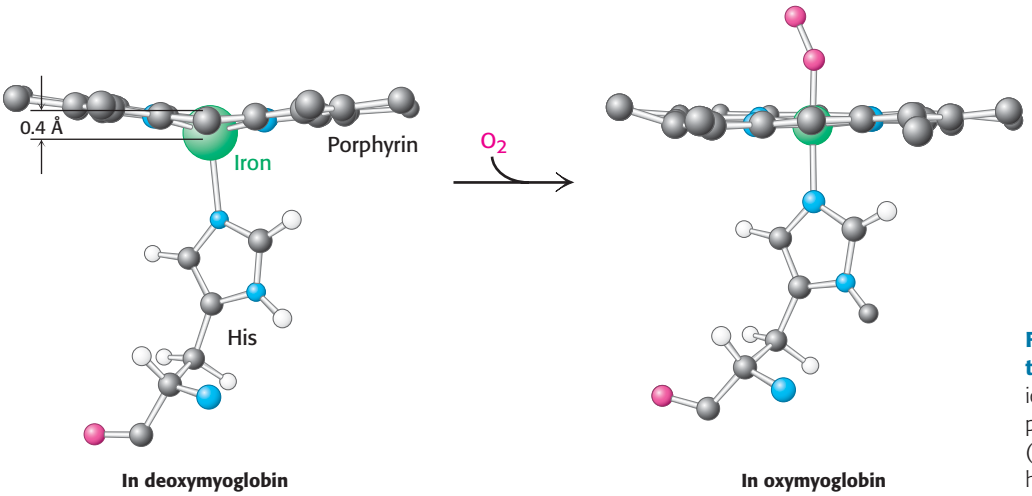


Figure 7.2 Oxygen binding changes the position of the iron ion. The iron ion lies slightly outside the plane of the porphyrin in deoxymyoglobin heme (left), but moves into the plane of the heme on oxygenation (right).

The iron atom lies in the center of the protoporphyrin, bonded to the four pyrrole nitrogen atoms. Although the heme-bound iron can be in either the ferrous (Fe^{2+}) or ferric (Fe^{3+}) oxidation state, only the Fe^{2+} state is capable of binding oxygen. The iron ion can form two additional bonds, one on each side of the heme plane. These binding sites are called the fifth and sixth coordination sites. In myoglobin, the fifth coordination site is occupied by the imidazole ring of a histidine residue from the protein. This histidine is referred to as the *proximal histidine*.

Oxygen binding occurs at the sixth coordination site. In deoxymyoglobin, this site remains unoccupied. The iron ion is slightly too large to fit into the well-defined hole within the porphyrin ring; it lies approximately 0.4 Å outside the porphyrin plane (Figure 7.2, left). Binding of the oxygen molecule at the sixth coordination site substantially rearranges the electrons within the iron so that the ion becomes effectively smaller, allowing it to move within the plane of the porphyrin (Figure 7.2, right). Remarkably, the structural changes that take place on oxygen binding were predicted by Linus Pauling, on the basis of magnetic measurements in 1936, nearly 25 years before the three-dimensional structures of myoglobin and hemoglobin were elucidated.

Changes in heme electronic structure upon oxygen binding are the basis for functional imaging studies

The change in electronic structure that occurs when the iron ion moves into the plane of the porphyrin is paralleled by alterations in the magnetic properties of hemoglobin; these changes are the basis for *functional magnetic resonance imaging* (fMRI), one of the most powerful methods for examining brain function. Nuclear magnetic resonance techniques detect signals that originate primarily from the protons in water molecules and are altered by the magnetic properties of hemoglobin. With the use of appropriate techniques, images can be generated that reveal differences in the relative amounts of deoxy- and oxyhemoglobin and thus the relative activity of various parts of the brain. When a specific part of the brain is active, blood vessels relax to allow more blood flow to that region. Thus, a more-active region of the brain will be richer in oxyhemoglobin.

These noninvasive methods identify areas of the brain that process sensory information. For example, subjects have been imaged while breathing air that either does or does not contain odorants. When odorants are present, fMRI detects an increase in the level of hemoglobin oxygenation (and,

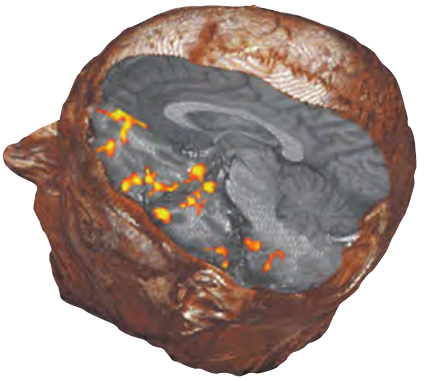


Figure 7.3 Functional magnetic resonance imaging of the brain. A functional magnetic resonance image reveals brain response to odorants. The light spots indicate regions of the brain activated by odorants. [From N. Sobel et al., *J. Neurophysiol.* 83(2000):537–551; courtesy of Dr. Noam Sobel.]

hence, of activity) in several regions of the brain (Figure 7.3). These regions are in the primary olfactory cortex, as well as in areas in which secondary processing of olfactory signals presumably takes place. Further analysis reveals the time course of activation of particular regions. Functional MRI shows tremendous potential for mapping regions and pathways engaged in processing sensory information obtained from all the senses. Thus, a seemingly incidental aspect of the biochemistry of hemoglobin has enabled observation of the brain in action.

The structure of myoglobin prevents the release of reactive oxygen species

Oxygen binding to iron in heme is accompanied by the partial transfer of an electron from the ferrous ion to oxygen. In many ways, the structure is best described as a complex between ferric ion (Fe^{3+}) and *superoxide anion* (O_2^-), as illustrated in Figure 7.4. It is crucial that oxygen, when it is released, leaves as dioxygen rather than superoxide, for two important reasons. First, superoxide and other species generated from it are reactive oxygen species that can be damaging to many biological materials. Second, release of superoxide would leave the iron ion in the ferric state. This species, termed *metmyoglobin*, does not bind oxygen. Thus, potential oxygen-storage capacity is lost. Features of myoglobin stabilize the oxygen complex such that superoxide is less likely to be released. In particular, the binding pocket of myoglobin includes an additional histidine residue (termed the *distal histidine*) that donates a hydrogen bond to the bound oxygen molecule (Figure 7.5). The superoxide character of the bound oxygen species

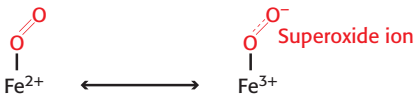


Figure 7.4 Iron–oxygen bonding. The interaction between iron and oxygen in myoglobin can be described as a combination of resonance structures, one with Fe^{2+} and dioxygen and another with Fe^{3+} and superoxide ion.

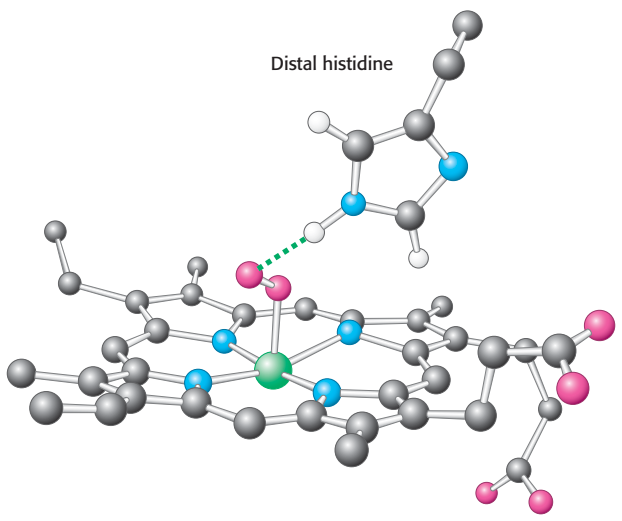


Figure 7.5 Stabilizing bound oxygen. A hydrogen bond (dotted green line) donated by the distal histidine residue to the bound oxygen molecule helps stabilize oxymyoglobin.

strengthens this interaction. Thus, *the protein component of myoglobin controls the intrinsic reactivity of heme,* making it more suitable for reversible oxygen binding.

Human hemoglobin is an assembly of four myoglobin-like subunits

The three-dimensional structure of hemoglobin from horse heart was solved by Max Perutz shortly after the determination of the myoglobin structure. Since then, the structures of hemoglobins from other species including humans have been determined. Hemoglobin consists of four polypeptide chains, two identical *α chains* and two identical *β chains* (Figure 7.6). Each of the subunits consists of a set of α helices in the same arrangement as the α helices in myoglobin (see Figure 6.15 for a comparison of the structures). The recurring structure is called a *globin fold*. Consistent with this structural similarity, alignment of the amino acid sequences of the α and β chains of human hemoglobin with those of sperm whale myoglobin yields 25% and 24% identity, respectively, and good conservation of key residues such as the proximal and distal histidines. Thus, the α and β chains are related to each other and to myoglobin by divergent evolution (Section 6.2).

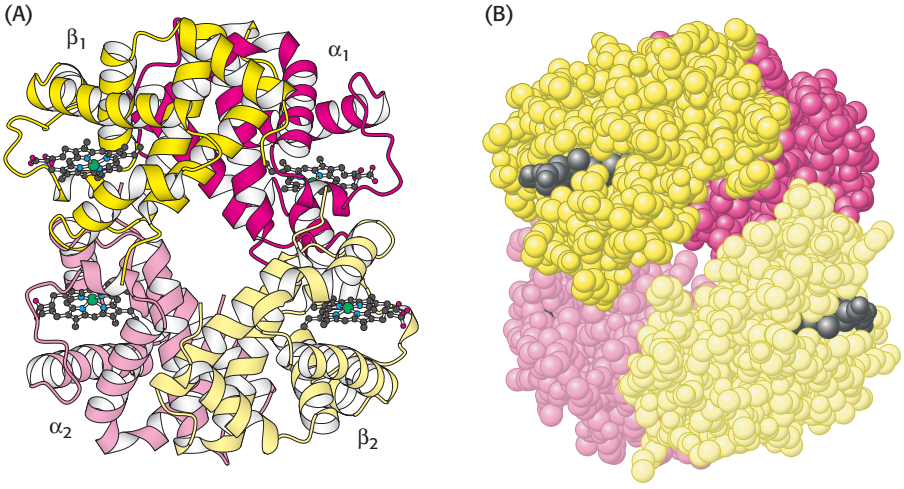


Figure 7.6 Quaternary structure of deoxyhemoglobin. Hemoglobin, which is composed of two α chains and two β chains, functions as a pair of αβ dimers. (A) A ribbon diagram. (B) A space-filling model. [Drawn from 1A3N.pdb.]

The hemoglobin tetramer, referred to as *hemoglobin A* (HbA), is best described as a pair of identical *αβ dimers* ($\alpha_1\beta_1$ and $\alpha_2\beta_2$) that associate to form the tetramer. In deoxyhemoglobin, these αβ dimers are linked by an extensive interface, which includes the carboxyl terminus of each chain. The heme groups are well separated in the tetramer by iron–iron distances ranging from 24 to 40 Å.

7.2 Hemoglobin Binds Oxygen Cooperatively

We can determine the oxygen-binding properties of each of these proteins by observing its *oxygen-binding curve,* a plot of the *fractional saturation* versus the concentration of oxygen. The fractional saturation, *Y,* is defined as the fraction of possible binding sites that contain bound oxygen. The value of *Y* can range from 0 (all sites empty) to 1 (all sites filled). The concentration

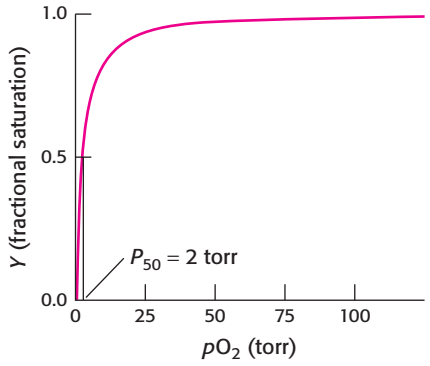


Figure 7.7 Oxygen binding by myoglobin. Half the myoglobin molecules have bound oxygen when the oxygen partial pressure is 2 torr.

of oxygen is most conveniently measured by its *partial pressure*, pO_2. For myoglobin, a binding curve indicating a simple chemical equilibrium is observed (Figure 7.7). Notice that the curve rises sharply as pO_2 increases and then levels off. Half-saturation of the binding sites, referred to as P_{50} (for 50% saturated), is at the relatively low value of 2 torr (mm Hg), indicating that oxygen binds with high affinity to myoglobin.

In contrast, the oxygen-binding curve for hemoglobin in red blood cells shows some remarkable features (Figure 7.8). It does not look like a simple binding curve such as that for myoglobin; instead, it resembles an "S." Such curves are referred to as *sigmoid* because of their S-like shape. In addition, oxygen binding for hemoglobin (P_{50} = 26 torr) is significantly weaker than that for myoglobin. Note that this binding curve is derived from hemoglobin in red blood cells. Inside red cells, hemoglobin interacts with 2,3-bisphosphoglycerate, a molecule that significantly lowers hemoglobin's oxygen affinity, as will be considered in detail shortly.

Torr
A unit of pressure equal to that exerted by a column of mercury 1 mm high at 0°C and standard gravity (1 mm Hg). Named after Evangelista Torricelli (1608–1647), inventor of the mercury barometer.

A sigmoid binding curve indicates that a protein shows a special binding behavior. For hemoglobin, this shape suggests that the binding of oxygen at one site within the hemoglobin tetramer increases the likelihood that oxygen binds at the remaining unoccupied sites. Conversely, the unloading of oxygen at one heme facilitates the unloading of oxygen at the others. This sort of binding behavior is referred to as *cooperative*, because the binding reactions at individual sites in each hemoglobin molecule are not independent of one another. We will return to the mechanism of this cooperativity shortly.

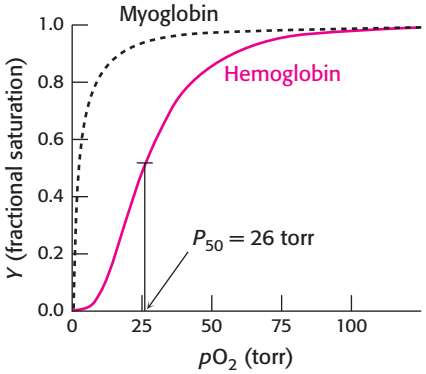


Figure 7.8 Oxygen binding by hemoglobin. This curve, obtained for hemoglobin in red blood cells, is shaped somewhat like an "S," indicating that distinct, but interacting, oxygen-binding sites are present in each hemoglobin molecule. Half-saturation for hemoglobin is 26 torr. For comparison, the binding curve for myoglobin is shown as a dashed black curve.

What is the physiological significance of the cooperative binding of oxygen by hemoglobin? Oxygen must be transported in the blood from the lungs, where the partial pressure of oxygen is relatively high (approximately 100 torr), to the actively metabolizing tissues, where the partial pressure of oxygen is much lower (typically, 20 torr). Let us consider how the cooperative behavior indicated by the sigmoid curve leads to efficient oxygen transport (Figure 7.9). In the lungs, hemoglobin becomes nearly saturated with oxygen such that 98% of the oxygen-binding sites are occupied. When hemoglobin moves to the tissues and releases O_2, the saturation level drops to 32%. Thus, a total of 98 − 32 = 66% of the potential oxygen-binding sites contribute to oxygen transport. The cooperative release of oxygen favors a more-complete unloading of oxygen in the tissues. If myoglobin were employed for oxygen transport, it would be 98% saturated in the lungs, but would remain 91% saturated in the tissues, and so only 98 − 91 = 7% of the sites would contribute to oxygen transport; myoglobin binds oxygen too tightly to be useful in oxygen transport. The situation might have been improved without cooperativity by the evolution of a noncooperative oxygen carrier with an optimized affinity for oxygen. For such a protein, the most oxygen that could be transported from a region in which pO_2 is 100 torr

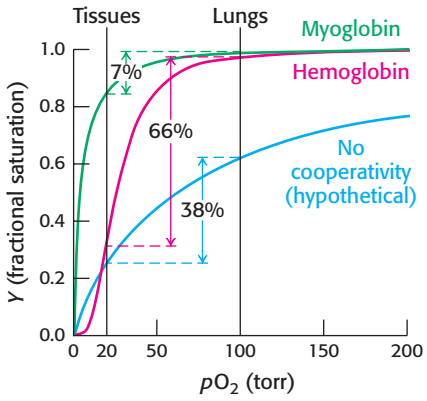


Figure 7.9 Cooperativity enhances oxygen delivery by hemoglobin. Because of cooperativity between O_2 binding sites, hemoglobin delivers more O_2 to tissues than would myoglobin or any noncooperative protein, even one with optimal O_2 affinity.

to one in which it is 20 torr is $63 - 25 = 38\%$. Thus, the cooperative binding and release of oxygen by hemoglobin enables it to deliver nearly 10 times as much oxygen as could be delivered by myoglobin and more than 1.7 times as much as could be delivered by any noncooperative protein.

Closer examination of oxygen concentrations in tissues at rest and during exercise underscores the effectiveness of hemoglobin as an oxygen carrier (Figure 7.10). Under resting conditions, the oxygen concentration in muscle is approximately 40 torr, but during exercise the concentration is reduced to 20 torr. In the decrease from 100 torr in the lungs to 40 torr in resting muscle, the oxygen saturation of hemoglobin is reduced from 98% to 77%, and so $98 - 77 = 21\%$ of the oxygen is released over a drop of 60 torr. In a decrease from 40 torr to 20 torr, the oxygen saturation is reduced from 77% to 32%, corresponding to an oxygen release of 45% over a drop of 20 torr. Thus, because the change in oxygen concentration from rest to exercise corresponds to the steepest part of the oxygen-binding curve, oxygen is effectively delivered to tissues where it is most needed. In Section 7.3, we shall examine other properties of hemoglobin that enhance its physiological responsiveness.

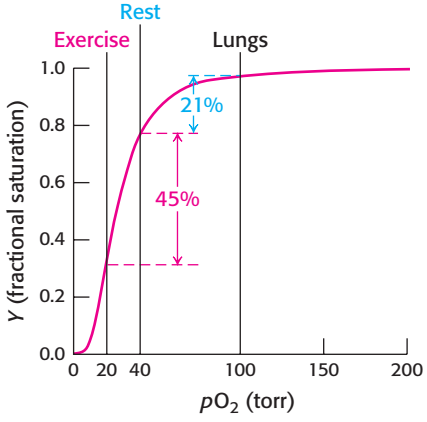


Figure 7.10 Responding to exercise. The drop in oxygen concentration from 40 torr in resting tissues to 20 torr in exercising tissues corresponds to the steepest part of the observed oxygen-binding curve. As shown here, hemoglobin is very effective in providing oxygen to exercising tissues.

Oxygen binding markedly changes the quaternary structure of hemoglobin

The cooperative binding of oxygen by hemoglobin requires that the binding of oxygen at one site in the hemoglobin tetramer influence the oxygen-binding properties at the other sites. Given the large separation between the iron sites, direct interactions are not possible. Thus, indirect mechanisms

for coupling the sites must be at work. These mechanisms are intimately related to the quaternary structure of hemoglobin.

Hemoglobin undergoes substantial changes in quaternary structure on oxygen binding: the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers rotate approximately 15 degrees with respect to one another (Figure 7.11). The dimers themselves are relatively unchanged, although there are localized conformational shifts. Thus, the interface between the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers is most affected by this structural transition. In particular, the $\alpha_1\beta_1$ and $\alpha_2\beta_2$ dimers are freer to move with respect to one another in the oxygenated state than they are in the deoxygenated state.

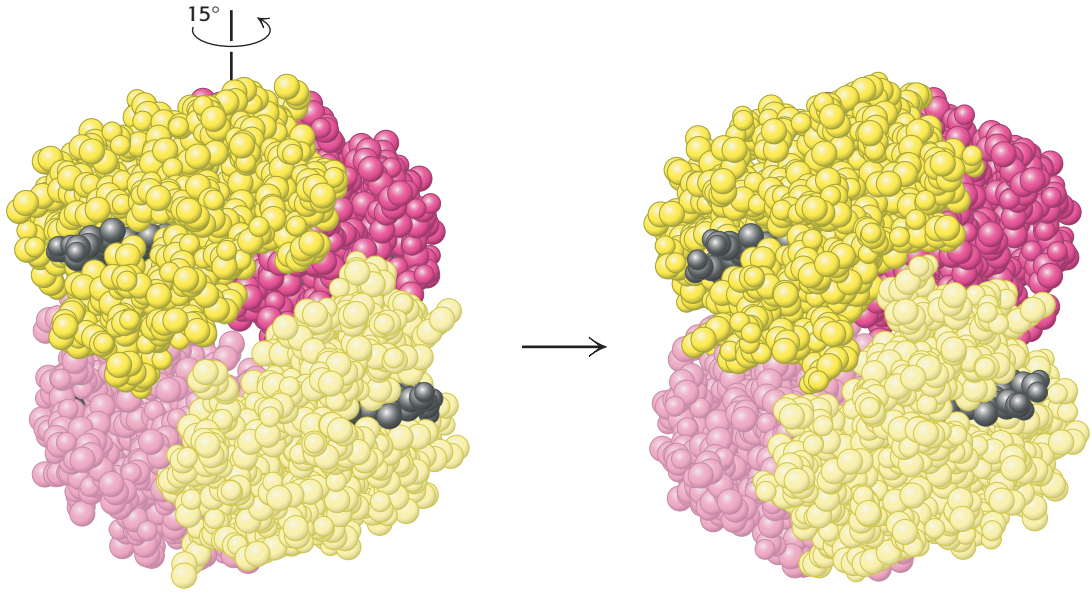


Figure 7.11 Quaternary structural changes on oxygen binding by hemoglobin. *Notice* that, on oxygenation, one αβ dimer shifts with respect to the other by a rotation of 15 degrees. [Drawn from 1A3N.pdb and 1LFQ.pdb.]

The quaternary structure observed in the deoxy form of hemoglobin, *deoxyhemoglobin*, is often referred to as the *T* (for tense) *state* because it is quite constrained by subunit–subunit interactions. The quaternary structure of the fully oxygenated form, *oxyhemoglobin*, is referred to as the *R* (for relaxed) *state*. In light of the observation that the R form of hemoglobin is less constrained, the tense and relaxed designations seem particularly apt. Importantly, in the R state, the oxygen-binding sites are free of strain and are capable of binding oxygen with higher affinity than are the sites in the T state. *By triggering the shift of the hemoglobin tetramer from the T state to the R state, the binding of oxygen to one site increases the binding affinity of other sites.*

Hemoglobin cooperativity can be potentially explained by several models

Two limiting models have been developed to explain the cooperative binding of ligands to a multisubunit assembly such as hemoglobin. In the *concerted model*, also known as the *MWC model* after Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeux, who first proposed it, the overall assembly can exist only in two forms: the T state and the R state. The binding of ligands simply shifts the equilibrium between these two states

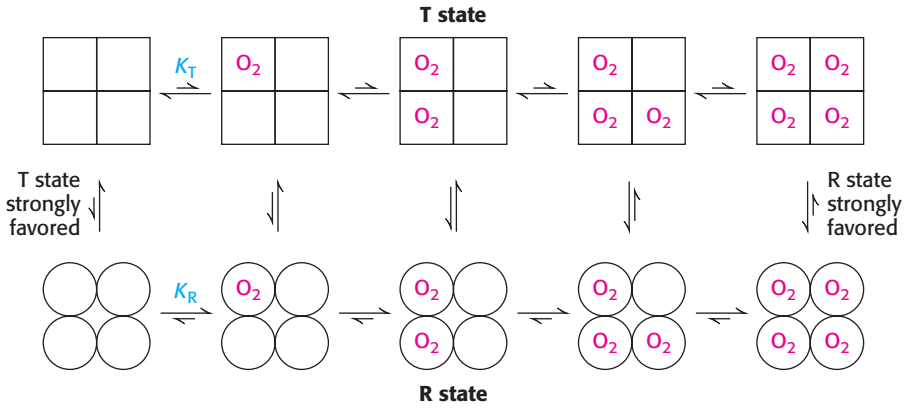


Figure 7.12 Concerted model. All molecules exist either in the T state or in the R state. At each level of oxygen loading, an equilibrium exists between the T and R states. The equilibrium shifts from strongly favoring the T state with no oxygen bound to strongly favoring the R state when the molecule is fully loaded with oxygen. The R state has a greater affinity for oxygen than does the T state.

(Figure 7.12). Thus, as a hemoglobin tetramer binds each oxygen molecule, the probability that the tetramer is in the R state increases. Deoxyhemoglobin tetramers are almost exclusively in the T state. However, the binding of oxygen to one site in the molecule shifts the equilibrium toward the R state. If a molecule assumes the R quaternary structure, the oxygen affinity of its sites increases. Additional oxygen molecules are now more likely to bind to the three unoccupied sites. Thus, the binding curve is shallow at low oxygen concentrations when all of the molecules are in the T state, becomes steeper as the fraction of molecules in the R state increases, and flattens out again when all of the sites within the R-state molecules become filled (Figure 7.13). These events produce the sigmoid binding curve so important for efficient oxygen transport.

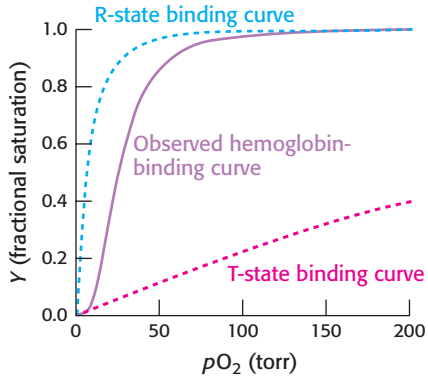


Figure 7.13 T-to-R transition. The observed binding curve for hemoglobin can be seen as a combination of the binding curves that would be observed if all molecules remained in the T state or if all of the molecules were in the R state. The sigmoidal curve is observed because molecules convert from the T state into the R state as oxygen molecules bind.

In the concerted model, each tetramer can exist in only two states, the T state and the R state. In an alternative model, the *sequential model,* the binding of a ligand to one site in an assembly increases the binding affinity of neighboring sites without inducing a full conversion from the T into the R state (Figure 7.14).

Is the cooperative binding of oxygen by hemoglobin better described by the concerted or the sequential model? Neither model in its pure form fully accounts for the behavior of hemoglobin. Instead, a combined model is required. Hemoglobin behavior is concerted in that the tetramer with three sites occupied by oxygen is almost always in the quaternary structure associated with the R state. The remaining open binding site has an affinity for oxygen more than 20-fold greater than that of fully deoxygenated hemoglobin binding its first oxygen. However, the behavior is not fully concerted, because hemoglobin with oxygen bound to only one of four sites remains primarily in the T-state quaternary structure. Yet, this molecule binds oxygen three times as strongly as does fully deoxygenated hemoglobin, an observation consistent only with a sequential model. These results highlight the fact that the concerted and sequential models represent idealized limiting cases, which real systems may approach but rarely attain.

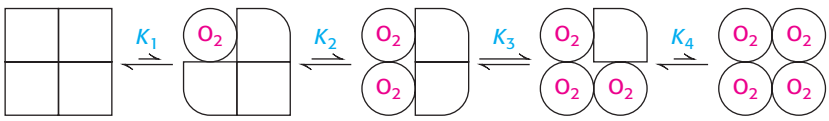


Figure 7.14 Sequential model. The binding of a ligand changes the conformation of the subunit to which it binds. This conformational change induces changes in neighboring subunits that increase their affinity for the ligand.

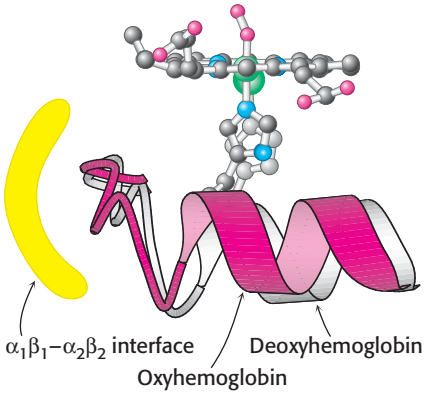


Figure 7.15 Conformational changes in hemoglobin. The movement of the iron ion on oxygenation brings the iron-associated histidine residue toward the porphyrin ring. The associated movement of the histidine-containing α helix alters the interface between the αβ dimers, instigating other structural changes. For comparison, the deoxyhemoglobin structure is shown in gray behind the oxyhemoglobin structure in color.

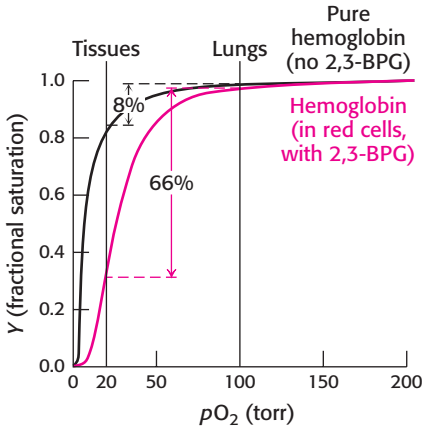


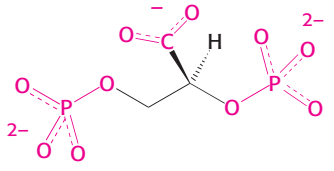
Figure 7.16 Oxygen binding by pure hemoglobin compared with hemoglobin in red blood cells. Pure hemoglobin binds oxygen more tightly than does hemoglobin in red blood cells. This difference is due to the presence of 2,3-bisphosphoglycerate (2,3-BPG) in red blood cells.

Structural changes at the heme groups are transmitted to the $\alpha_1\beta_1$–$\alpha_2\beta_2$ interface

We now examine how oxygen binding at one site is able to shift the equilibrium between the T and R states of the entire hemoglobin tetramer. As in myoglobin, oxygen binding causes each iron atom in hemoglobin to move from outside the plane of the porphyrin into the plane. When the iron atom moves, the histidine residue bound in the fifth coordination site moves with it. This histidine residue is part of an α helix, which also moves (Figure 7.15). The carboxyl terminal end of this α helix lies in the interface between the two αβ dimers. The change in position of the carboxyl terminal end of the helix favors the T-to-R transition. Consequently, *the structural transition at the iron ion in one subunit is directly transmitted to the other subunits.* The rearrangement of the dimer interface provides a pathway for communication between subunits, enabling the cooperative binding of oxygen.

2,3-Bisphosphoglycerate in red cells is crucial in determining the oxygen affinity of hemoglobin

For hemoglobin to function efficiently, the T state must remain stable until the binding of sufficient oxygen has converted it into the R state. In fact, however, the T state of hemoglobin is highly unstable, pushing the equilibrium so far toward the R state that little oxygen would be released in physiological conditions. Thus, an additional mechanism is needed to properly stabilize the T state. This mechanism was discovered by comparing the oxygen-binding properties of hemoglobin in red blood cells with fully purified hemoglobin (Figure 7.16). Pure hemoglobin binds oxygen much more tightly than does hemoglobin in red blood cells. This dramatic difference is due to the presence within these cells of *2,3-bisphosphoglycerate* (2,3-BPG; also known as 2,3-diphosphoglycerate or 2,3-DPG).



2,3-Bisphosphoglycerate (2,3-BPG)

This highly anionic compound is present in red blood cells at approximately the same concentration as that of hemoglobin (~2 mM). Without 2,3-BPG, hemoglobin would be an extremely inefficient oxygen transporter, releasing only 8% of its cargo in the tissues.

How does 2,3-BPG lower the oxygen affinity of hemoglobin so significantly? Examination of the crystal structure of deoxyhemoglobin in the presence of 2,3-BPG reveals that a single molecule of 2,3-BPG binds in the center of the tetramer, in a pocket present only in the T form (Figure 7.17). On T-to-R transition, this pocket collapses and 2,3-BPG is released. Thus, in order for the structural transition from T to R to take place, the bonds between hemoglobin and 2,3-BPG must be broken. In the presence of 2,3-BPG, more oxygen-binding sites within the hemoglobin tetramer must be occupied in order to induce the T-to-R transition, and so hemoglobin remains in the lower-affinity T state until higher oxygen concentrations are reached. This mechanism of regulation is remarkable because 2,3-BPG does not in any way resemble oxygen, the molecule on which hemoglobin carries out its primary function. 2,3-BPG is referred to as an *allosteric*

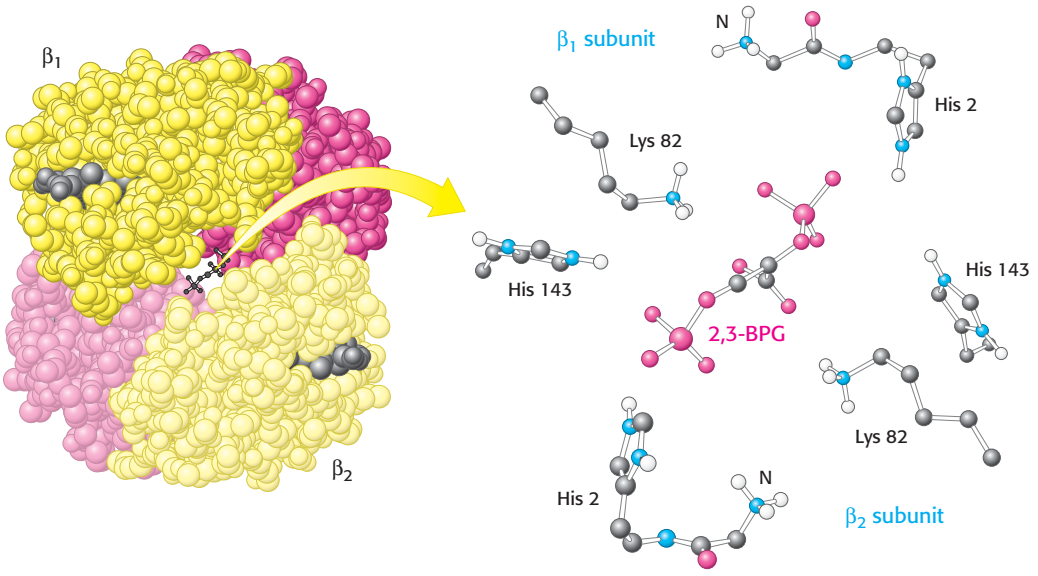


Figure 7.17 Mode of binding of 2,3-BPG to human deoxyhemoglobin. 2,3-Bisphosphoglycerate binds to the central cavity of deoxyhemoglobin (left). There, it interacts with three positively charged groups on each β chain (right). [Drawn from 1B86.pdb.]

effector (from the Greek *allos,* "other," and *stereos,* "structure"). Regulation by a molecule structurally unrelated to oxygen is possible because the allosteric effector binds to a site that is completely distinct from that for oxygen. We will encounter allosteric effects again when we consider enzyme regulation in Chapter 10.

The binding of 2,3-BPG to hemoglobin has other crucial physiological consequences. The globin gene expressed by human fetuses differs from that expressed by adults; *fetal hemoglobin* tetramers include two α chains and two γ chains. The γ chain, a result of a gene duplication, is 72% identical in amino acid sequence with the β chain. One noteworthy change is the substitution of a serine residue for His 143 in the β chain, part of the 2,3-BPG-binding site. This change removes two positive charges from the 2,3-BPG-binding site (one from each chain) and reduces the affinity of 2,3-BPG for fetal hemoglobin. Consequently, the oxygen-binding affinity of fetal hemoglobin is higher than that of maternal (adult) hemoglobin (Figure 7.18). This difference in oxygen affinity allows oxygen to be effectively transferred from maternal to fetal red blood cells. We have here an example in which gene duplication and specialization produced a ready solution to a biological challenge—in this case, the transport of oxygen from mother to fetus.

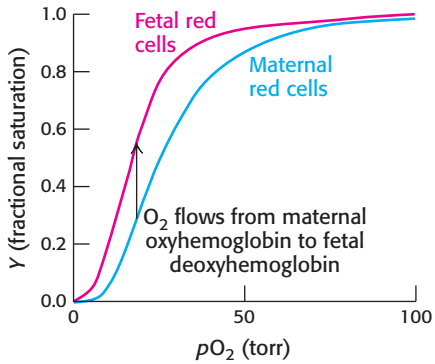


Figure 7.18 Oxygen affinity of fetal red blood cells. Fetal red blood cells have a higher oxygen affinity than do maternal red blood cells because fetal hemoglobin does not bind 2,3-BPG as well as maternal hemoglobin does.

Carbon monoxide can disrupt oxygen transport by hemoglobin

Carbon monoxide (CO) is a colorless, odorless gas that binds to hemoglobin at the same site as oxygen, forming a complex termed *carboxyhemoglobin*. Formation of carboxyhemoglobin exerts devastating consequences on normal oxygen transport in two ways. First, carbon monoxide binds to hemoglobin about 200-fold more tightly than does oxygen. Even at low partial pressures in the blood, carbon monoxide will displace oxygen from hemoglobin, preventing its delivery. Second, carbon monoxide bound to one site in hemoglobin will shift the oxygen saturation curve of the remaining sites to the left, forcing the tetramer into the R state. This results in an increased affinity for oxygen, preventing its dissociation at tissues.

Exposure to carbon monoxide—from gas appliances and running automobiles, for example—can cause carbon monoxide poisoning, in which patients exhibit nausea, vomiting, lethargy, weakness, and disorientation.

One treatment for carbon monoxide poisoning is administration of 100% oxygen, often at pressures greater than atmospheric pressure (this treatment is referred to as *hyperbaric oxygen therapy*). With this therapy, the partial pressure of oxygen in the blood becomes sufficiently high to increase substantially the rate of carbon monoxide displacement from hemoglobin. Exposure to high concentrations of carbon monoxide, however, can be rapidly fatal: in the United States, about 2500 people die each year from carbon monoxide poisoning, about 500 of them from accidental exposures and nearly 2000 by suicide.

7.3 Hydrogen Ions and Carbon Dioxide Promote the Release of Oxygen: The Bohr Effect

We have seen how cooperative release of oxygen from hemoglobin helps deliver oxygen to tissues where it is most needed, as revealed by their low oxygen partial pressures. This ability is enhanced by the facility of hemoglobin to respond to other cues in its physiological environment that signal the need for oxygen. Rapidly metabolizing tissues, such as contracting muscle, generate large amounts of hydrogen ions and carbon dioxide (Chapter 16). To release oxygen where the need is greatest, hemoglobin has evolved to respond to higher levels of these substances. Like 2,3-BPG, hydrogen ions and carbon dioxide are *allosteric effectors* of hemoglobin that bind to sites on the molecule that are distinct from the oxygen-binding sites. The regulation of oxygen binding by hydrogen ions and carbon dioxide is called the *Bohr effect* after Christian Bohr, who described this phenomenon in 1904.

The oxygen affinity of hemoglobin decreases as pH decreases from a value of 7.4 (Figure 7.19). Consequently, as hemoglobin moves into a region of lower pH, its tendency to release oxygen increases. For example, transport from the lungs, with pH 7.4 and an oxygen partial pressure of 100 torr, to active muscle, with a pH of 7.2 and an oxygen partial pressure of 20 torr, results in a release of oxygen amounting to 77% of total carrying capacity. Only 66% of the oxygen would be released in the absence of any change in pH. Structural and chemical studies have revealed much about the chemical basis of the Bohr effect. At least two sets of chemical groups are important for sensing changes in pH: the α-amino groups at the amino termini of the α chain and the side chains of histidines β146 and α122, all of which have pK_a values near pH 7. Consider histidine β146, the residue at the C terminus of the β chain. In deoxyhemoglobin, the terminal carboxylate group of β146 forms a salt bridge with a lysine residue in the α subunit of the other αβ dimer. This interaction locks the side chain of histidine β146 in a

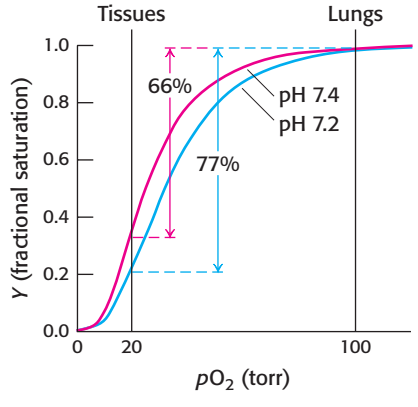


Figure 7.19 Effect of pH on the oxygen affinity of hemoglobin. Lowering the pH from 7.4 (red curve) to 7.2 (blue curve) results in the release of O_2 from oxyhemoglobin.

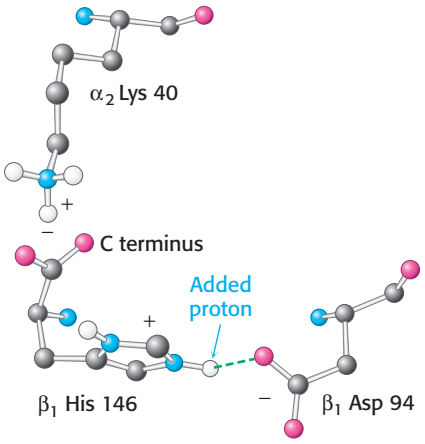


Figure 7.20 Chemical basis of the Bohr effect. In deoxyhemoglobin, three amino acid residues form two salt bridges that stabilize the T quaternary structure. The formation of one of the salt bridges depends on the presence of an added proton on histidine β146. The proximity of the negative charge on aspartate β94 in deoxyhemoglobin favors protonation of this histidine. *Notice* that the salt bridge between histidine β146 and aspartate β94 is stabilized by a hydrogen bond (green dashed line).

position from which it can participate in a salt bridge with negatively charged aspartate β94 in the same chain, provided that the imidazole group of the histidine residue is protonated (Figure 7.20).

The other groups also participate in salt bridges in the T state. *The formation of these salt bridges stabilizes the T state, leading to a greater tendency for oxygen to be released.* For example, at high pH, the side chain of histidine β146 is not protonated and the salt bridge does not form. As the pH drops, however, the side chain of histidine β146 becomes protonated, the salt bridge with aspartate β94 forms, and the T state is stabilized.

Carbon dioxide, a neutral species, passes through the red-blood-cell membrane into the cell. This transport is also facilitated by membrane transporters including proteins associated with Rh blood types. Carbon dioxide stimulates oxygen release by two mechanisms. First, the presence of high concentrations of carbon dioxide leads to a drop in pH within the red blood cell (Figure 7.21). Carbon dioxide reacts with water to form carbonic acid, H_2CO_3. This reaction is accelerated by *carbonic anhydrase,* an enzyme abundant in red blood cells that will be considered extensively in Chapter 9. H_2CO_3 is a moderately strong acid with a pK_a of 3.5. Thus, once formed, carbonic acid dissociates to form bicarbonate ion, HCO_3^-, and H^+, resulting in a drop in pH that stabilizes the T state by the mechanism discussed previously.

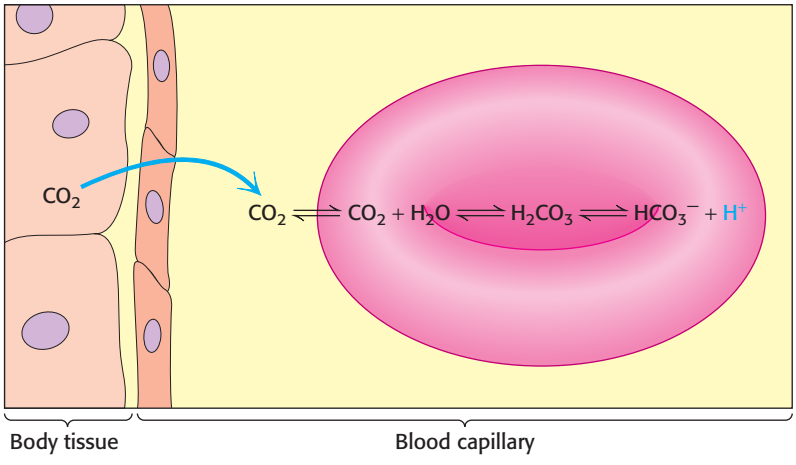


Figure 7.21 Carbon dioxide and pH. Carbon dioxide in the tissues diffuses into red blood cells. Inside a red blood cell, carbon dioxide reacts with water to form carbonic acid, in a reaction catalyzed by the enzyme carbonic anhydrase. Carbonic acid dissociates to form HCO_3^- and H^+, resulting in a drop in pH inside the red cell.

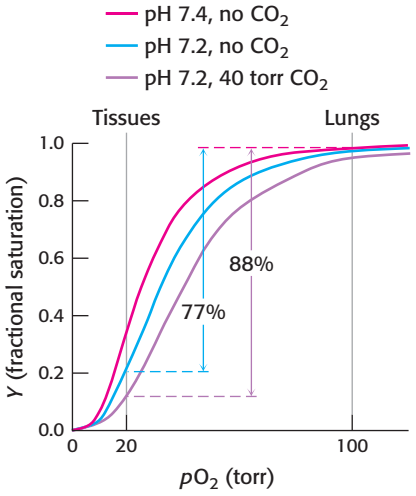
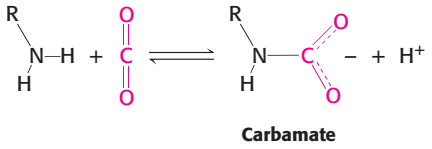


Figure 7.22 Carbon dioxide effects. The presence of carbon dioxide decreases the affinity of hemoglobin for oxygen even beyond the effect due to a decrease in pH, resulting in even more efficient oxygen transport from the tissues to the lungs.

In the second mechanism, a direct chemical interaction between carbon dioxide and hemoglobin stimulates oxygen release. The effect of carbon dioxide on oxygen affinity can be seen by comparing oxygen-binding curves in the absence and in the presence of carbon dioxide at a constant pH (Figure 7.22). In the presence of carbon dioxide at a partial pressure of 40 torr at pH 7.2, the amount of oxygen released approaches 90% of the maximum carrying capacity. Carbon dioxide stabilizes deoxyhemoglobin by reacting with the terminal amino groups to form *carbamate* groups, which are negatively charged, in contrast with the neutral or positive charges on the free amino groups.



Carbamate

The amino termini lie at the interface between the αβ dimers, and these negatively charged carbamate groups participate in salt-bridge interactions that stabilize the T state, favoring the release of oxygen.

Carbamate formation also provides a mechanism for carbon dioxide transport from tissues to the lungs, but it accounts for only about 14% of the total carbon dioxide transport. Most carbon dioxide released from red blood cells is transported to the lungs in the form of HCO_3^- produced from the hydration of carbon dioxide inside the cell (Figure 7.23). Much of the HCO_3^- that is formed leaves the cell through a specific membrane-transport protein that exchanges HCO_3^- from one side of the membrane for Cl^- from the other side. Thus, the serum concentration of HCO_3^- increases. By this means, a large concentration of carbon dioxide is transported from tissues to the lungs in the form of HCO_3^-. In the lungs, this process is reversed: HCO_3^- is converted back into carbon dioxide and exhaled. Thus, carbon dioxide generated by active tissues contributes to a decrease in red-blood-cell pH and, hence, to oxygen release and is converted into a form that can be transported in the serum and released in the lungs.

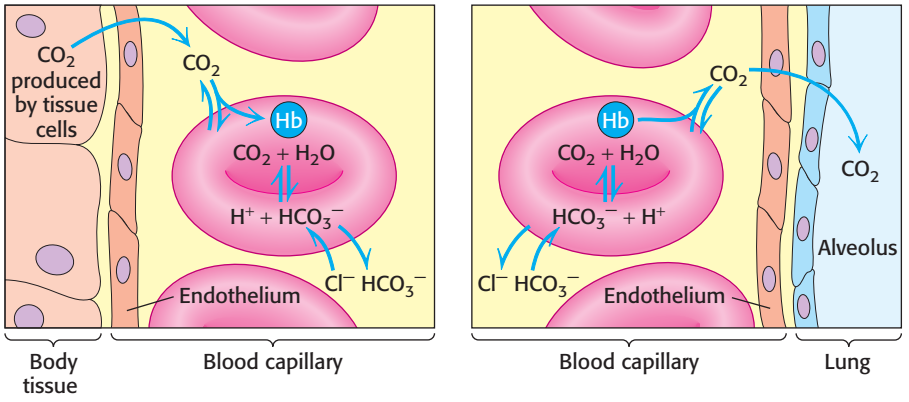


Figure 7.23 Transport of CO_2 from tissues to lungs. Most carbon dioxide is transported to the lungs in the form of HCO_3^- produced in red blood cells and then released into the blood plasma. A lesser amount is transported by hemoglobin in the form of an attached carbamate.

7.4 Mutations in Genes Encoding Hemoglobin Subunits Can Result in Disease

In modern times, particularly after the sequencing of the human genome, to think of genetically encoded variations in protein sequence as a factor in specific diseases is routine. The notion that diseases might be caused by

molecular defects was proposed by Linus Pauling in 1949 (4 years before Watson and Crick's proposal of the DNA double helix) to explain the blood disease *sickle-cell anemia*. The name of the disorder comes from the abnormal sickle shape of red blood cells deprived of oxygen observed in people suffering from this disease (Figure 7.24). Pauling proposed that sickle-cell anemia might be caused by a specific variation in the amino acid sequence of one hemoglobin chain. Today, we know that this bold hypothesis is correct. In fact, approximately 7% of the world's population are carriers of some disorder of hemoglobin caused by a variation in its amino acid sequence. In concluding this chapter, we will focus on the two most important of these disorders, sickle-cell anemia and thalassemia.

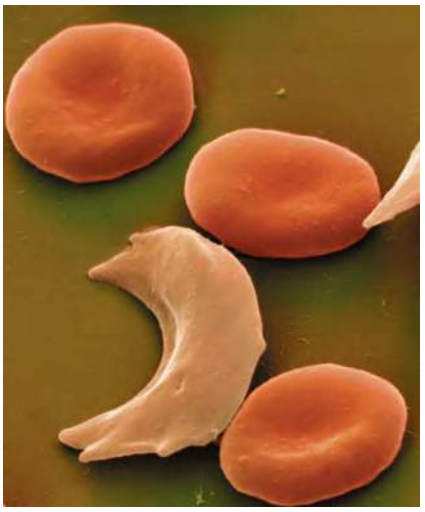


Figure 7.24 Sickled red blood cells. A micrograph showing a sickled red blood cell adjacent to normally shaped red blood cells. [Eye of Science/Photo Researchers.]

Sickle-cell anemia results from the aggregation of mutated deoxyhemoglobin molecules

People with sickled red blood cells experience a number of dangerous symptoms. Examination of the contents of these red cells reveals that the hemoglobin molecules have formed large fibrous aggregates (Figure 7.25). These fibers extend across the red blood cells, distorting them so that they clog small capillaries and impair blood flow. The results may be painful swelling of the extremities and a higher risk of stroke or bacterial infection (due to poor circulation). The sickled red cells also do not remain in circulation as long as normal cells do, leading to anemia.

What is the molecular defect associated with sickle-cell anemia? Using newly developed chromatographic techniques, Vernon Ingram demonstrated in 1956 that a single amino acid substitution in the β chain of hemoglobin is responsible—namely, the replacement of a valine residue with glutamate in position 6. The mutated form is referred to as *hemoglobin S* (HbS). In people with sickle-cell anemia, both alleles of the hemoglobin β-chain gene (HbB) are mutated. The HbS substitution substantially decreases the solubility of deoxyhemoglobin, although it does not markedly alter the properties of oxyhemoglobin.

Examination of the structure of hemoglobin S reveals that the new valine residue lies on the surface of the T-state molecule (Figure 7.26). This new hydrophobic patch interacts with another hydrophobic patch formed by Phe 85 and Val 88 of the β chain of a neighboring molecule to initiate the aggregation process. More-detailed analysis reveals that a single hemoglobin S fiber is formed from 14 chains of multiple interlinked hemoglobin molecules. Why do these aggregates not form when hemoglobin S is oxygenated? Oxyhemoglobin S is in the R state, and residues Phe 85 and Val 88 on the β chain are largely buried inside the hemoglobin assembly.

Figure 7.25 Sickle-cell hemoglobin fibers. An electron micrograph depicting a ruptured sickled red blood cell with fibers of sickle-cell hemoglobin emerging. [Courtesy of Robert Josephs and Thomas E. Wellems, University of Chicago.]

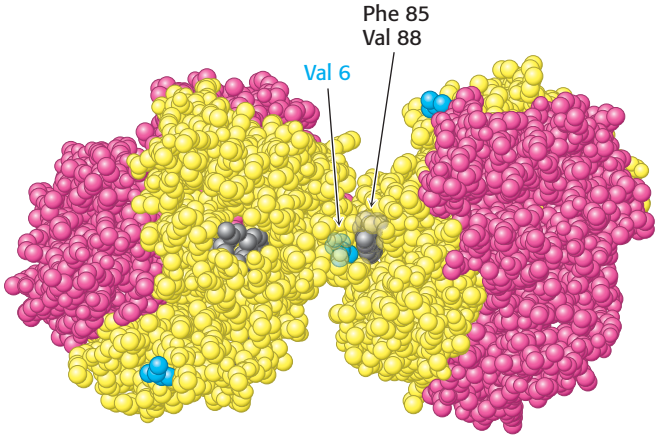


Figure 7.26 Deoxygenated hemoglobin S. The interaction between Val 6 (blue) on a β chain of one hemoglobin molecule and a hydrophobic patch formed by Phe 85 and Val 88 (gray) on a β chain of another deoxygenated hemoglobin molecule leads to hemoglobin aggregation. The exposed Val 6 residues of other β chains participate in other such interactions in hemoglobin S fibers. [Drawn from 2HBS.pdb.]

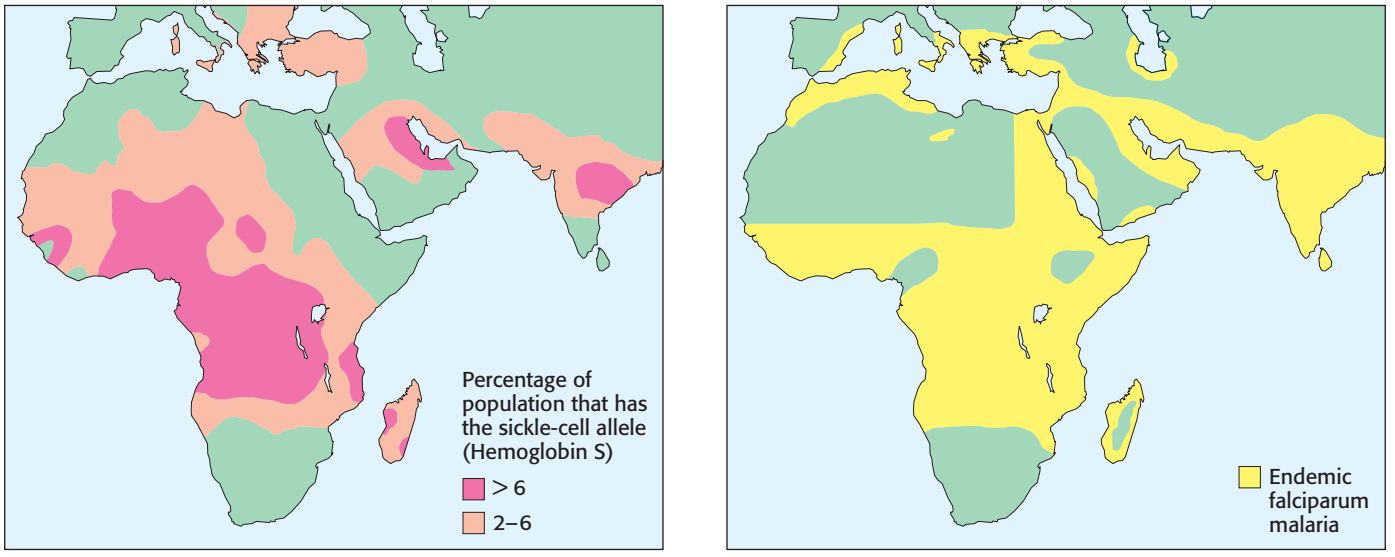


Figure 7.27 Sickle-cell trait and malaria. A significant correlation is observed between regions with a high frequency of the HbS allele and regions with a high prevalence of malaria.

Without a partner with which to interact, the surface Val residue in position 6 is benign.

Approximately 1 in 100 West Africans suffer from sickle-cell anemia. Given the often devastating consequences of the disease, why is the HbS mutation so prevalent in Africa and in some other regions? Recall that both copies of the HbB gene are mutated in people with sickle-cell anemia. People with one copy of the HbB gene and one copy of the HbS are relatively unaffected. They are said to have *sickle-cell trait* because they can pass the HbS gene to their offspring. However, people with sickle-cell trait are resistant to *malaria,* a disease carried by a parasite, *Plasmodium falciparum,* that lives within red blood cells at one stage in its life cycle. The dire effect of malaria on health and reproductive likelihood in regions where malaria has been historically endemic has favored people with sickle-cell trait, increasing the prevalence of the HbS allele (Figure 7.27).

Thalassemia is caused by an imbalanced production of hemoglobin chains

Sickle-cell anemia is caused by the substitution of a single specific amino acid in one hemoglobin chain. *Thalassemia,* the other prevalent inherited disorder of hemoglobin, is caused by the loss or substantial reduction of a single hemoglobin *chain*. The result is low levels of functional hemoglobin and a decreased production of red blood cells, which may lead to anemia, fatigue, pale skin, and spleen and liver malfunction. Thalassemia is a set of related diseases. In α-thalassemia, the α chain of hemoglobin is not produced in sufficient quantity. Consequently, hemoglobin tetramers form that contain only the β chain. These tetramers, referred to as *hemoglobin H* (HbH), bind oxygen with high affinity and no cooperativity. Thus, oxygen release in the tissues is poor. In β-thalassemia, the β chain of hemoglobin is not produced in sufficient quantity. In the absence of β chains, the α chains form insoluble aggregates that precipitate inside immature red blood cells. The loss of red blood cells results in anemia. The most severe form of β-thalassemia is called *thalassemia major* or *Cooley anemia*.

Both α- and β-thalassemia are associated with many different genetic variations and display a wide range of clinical severity. The most severe forms of α-thalassemia are usually fatal shortly before or just after birth. However, these forms are relatively rare. An examination of the repertoire

of hemoglobin genes in the human genome provides one explanation. Normally, humans have not two but four alleles for the α chain, arranged such that the two genes are located adjacent to each other on one end of each chromosome 16. Thus, the complete loss of α-chain expression requires the disruption of four alleles. β-Thalassemia is more common because humans normally have only two alleles for the β chain, one on each copy of chromosome 11.

The accumulation of free alpha-hemoglobin chains is prevented

The presence of four genes expressing the α chain, compared with two for the β chain, suggests that the α chain would be produced in excess (given the overly simple assumption that protein expression from each gene is comparable). If this is correct, why doesn't the excess α chain precipitate? One mechanism for maintaining α chains in solution was revealed by the discovery of an 11-kd protein in red blood cells called *α-hemoglobin stabilizing protein* (AHSP). This protein forms a soluble complex specifically with newly synthesized α-chain monomers. The crystal structure of a complex between AHSP and α-hemoglobin reveals that AHSP binds to the same face of α-hemoglobin as does β-hemoglobin (Figure 7.28). AHSP binds the α chain in both the deoxygenated and oxygenated forms. In the complex with oxygen bound, the distal histidine, rather than the proximal histidine, binds the iron atom.

AHSP serves to bind and ensure the proper folding of α-hemoglobin as it is produced. As β-hemoglobin is expressed, it displaces AHSP because the α-hemoglobin–β-hemoglobin dimer is more stable than the α-hemoglobin–AHSP complex. Thus, AHSP prevents the misfolding, accumulation, and precipitation of free α-hemoglobin. Studies are under way to determine if mutations in the gene encoding AHSP play a role in modulating the severity of β-thalassemia.

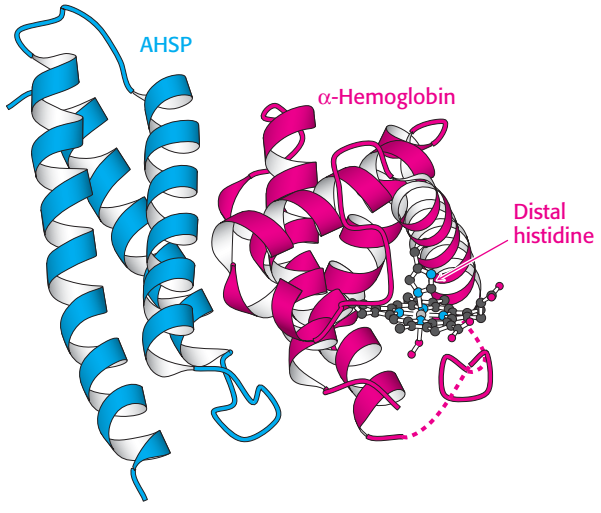


Figure 7.28 Stabilizing free α-hemoglobin. The structure of a complex between AHSP and α-hemoglobin is shown. In this complex, the iron atom is bound to oxygen and to the distal histidine. *Notice* that AHSP binds to the same surface of α-hemoglobin as does β-hemoglobin. [Drawn from 1Y01.pdb.]

Additional globins are encoded in the human genome

In addition to the gene for myoglobin, the two genes for α-hemoglobin, and the one for β-hemoglobin, the human haploid genome contains other globin genes. We have already encountered fetal hemoglobin, which contains the γ chain in place of the β chain. Several other genes encode other hemoglobin subunits that are expressed during development, including the δ chain, the ϵ chain, and the ζ chain.

Examination of the human genome sequence has revealed two additional globins. Both of these proteins are monomeric proteins, more similar to myoglobin than to hemoglobin. The first, *neuroglobin,* is expressed primarily in the brain and at especially high levels in the retina. Neuroglobin may play a role in protecting neural tissues from hypoxia (insufficient oxygen). The second, *cytoglobin,* is expressed more widely throughout the body. Structural and spectroscopic studies reveal that, in both neuroglobin and cytoglobin, the proximal and the distal histidines are coordinated to the iron atom in the deoxy form. Oxygen binding displaces the distal histidine. Future studies should more completely elucidate the functions of these members of the globin family.

Summary

7.1 Myoglobin and Hemoglobin Bind Oxygen at Iron Atoms in Heme

Myoglobin is a largely α-helical protein that binds the prosthetic group heme. Heme consists of protoporphyrin, an organic component with four linked pyrrole rings, and a central iron ion in the Fe^{2+} state. The iron ion is coordinated to the side chain of a histidine residue in myoglobin, referred to as the proximal histidine. One of the oxygen atoms in O_2 binds to an open coordination site on the iron. Because of partial electron transfer from the iron to the oxygen, the iron ion moves into the plane of the porphyrin on oxygen binding. Hemoglobin consists of four polypeptide chains, two α chains and two β chains. Each of these chains is similar in amino acid sequence to myoglobin and folds into a very similar three-dimensional structure. The hemoglobin tetramer is best described as a pair of αβ dimers.

7.2 Hemoglobin Binds Oxygen Cooperatively

The oxygen-binding curve for myoglobin reveals a simple equilibrium binding process. Myoglobin is half-saturated with oxygen at an oxygen concentration of approximately 2 torr. The oxygen-binding curve for hemoglobin has an "S"-like (sigmoid) shape, indicating that the oxygen binding is cooperative. The binding of oxygen at one site within the hemoglobin tetramer affects the affinities of the other sites for oxygen. Cooperative oxygen binding and release significantly increase the efficiency of oxygen transport. The amount of the potential oxygen-carrying capacity utilized in transporting oxygen from the lungs (with a partial pressure of oxygen of 100 torr) to tissues (with a partial pressure of oxygen of 20 torr) is 66% compared with 7% if myoglobin had been used as the oxygen carrier.

The quaternary structure of hemoglobin changes on oxygen binding. The structure of deoxyhemoglobin is referred to as the T state. The structure of oxyhemoglobin is referred to as the R state. The two αβ dimers rotate by approximately 15 degrees with respect to one another in the transition from the T to the R state. Cooperative binding can be potentially explained by concerted and sequential models. In the concerted model, each hemoglobin adopts either the T state or the R state; the equilibrium between these two states is determined by the number of occupied oxygen-binding sites. Sequential models allow intermediate structures. Structural changes at the iron sites in response to oxygen binding are transmitted to the interface between αβ dimers, influencing the T-to-R equilibrium.

Red blood cells contain 2,3-bisphosphoglycerate in concentrations approximately equal to that for hemoglobin. 2,3-BPG binds tightly to the T state but not to the R state, stabilizing the T state and lowering the oxygen affinity of hemoglobin. Fetal hemoglobin binds oxygen more tightly than does adult hemoglobin owing to weaker 2,3-BPG binding. This difference allows oxygen transfer from maternal to fetal blood.

7.3 Hydrogen Ions and Carbon Dioxide Promote the Release of Oxygen

The oxygen-binding properties of hemoglobin are markedly affected by pH and by the presence of carbon dioxide, a phenomenon known as the Bohr effect. Increasing the concentration of hydrogen ions—that is, decreasing pH—decreases the oxygen affinity of hemoglobin, owing to the protonation of the amino termini and certain histidine residues. The protonated residues help stabilize the T state. Increasing concentrations of carbon dioxide decrease the oxygen affinity of hemoglobin by two mechanisms. First, carbon dioxide is converted into carbonic acid, which lowers the oxygen affinity of hemoglobin by decreasing the pH inside the red blood cell. Second, carbon dioxide adds to the amino termini of hemoglobin to form carbamates. These negatively charged groups stabilize deoxyhemoglobin through ionic interactions. Because hydrogen ions and carbon dioxide are produced in rapidly metabolizing tissues, the Bohr effect helps deliver oxygen to sites where it is most needed.

7.4 Mutations in Genes Encoding Hemoglobin Subunits Can Result in Disease

Sickle-cell disease is caused by a mutation in the β chain of hemoglobin that substitutes a valine residue for a glutamate residue. As a result, a hydrophobic patch forms on the surface of deoxy (T-state) hemoglobin that leads to the formation of fibrous polymers. These fibers distort red blood cells into sickle shapes. Sickle-cell disease was the first disease to be associated with a change in the amino acid sequence of a protein. Thalassemias are diseases caused by the reduced production of either the α or the β chain, yielding hemoglobin tetramers that contain only one type of hemoglobin chain. Such hemoglobin molecules are characterized by poor oxygen release and low solubility, leading to the destruction of red blood cells in the course of their development. Red-blood-cell precursors normally produce a slight excess of hemoglobin α chains compared with β chains. To prevent the aggregation of the excess α chains, they produce α-hemoglobin stabilizing protein, which binds specifically to newly synthesized α-chain monomers to form a soluble complex.

APPENDIX: Binding Models Can Be Formulated in Quantitative Terms: The Hill Plot and the Concerted Model

The Hill Plot

A useful way of quantitatively describing cooperative binding processes such as that for hemoglobin was developed by Archibald Hill in 1913. Consider the *hypothetical* equilibrium for a protein X binding a ligand S:

$$X + nS \rightleftharpoons X(S)_n \qquad (1)$$

where n is a variable that can take on both integral and fractional values. The parameter n is a measure of the degree of cooperativity in ligand binding, although it does not have deeper significance because equation 1 does not represent an actual physical process. For X = hemoglobin and S = O_2, the maximum value of n is 4. The value of $n = 4$ would apply if oxygen binding by

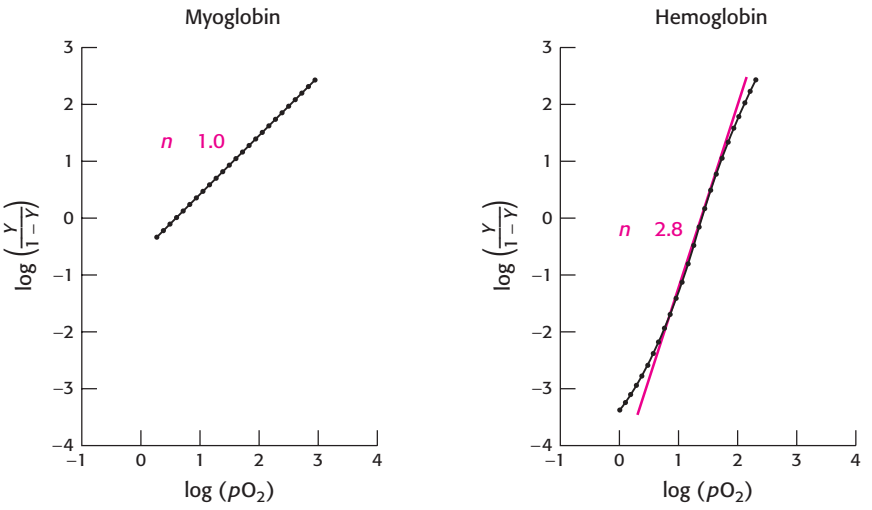


Figure 7.29 Hill plots for myoglobin and hemoglobin.

hemoglobin were completely cooperative. If oxygen binding were completely noncooperative, then n would be 1.

Analysis of the equilibrium in equation 1 yields the following expression for the fractional saturation, Y:

$$Y = \frac{[S]^n}{[S]^n + [S_{50}]^n}$$

where $[S_{50}]$ is the concentration at which X is half-saturated. For hemoglobin, this expression becomes

$$Y = \frac{pO_2{}^n}{pO_2{}^n + P_{50}{}^n}$$

where P_{50} is the partial pressure of oxygen at which hemoglobin is half-saturated. This expression can be rearranged to:

$$\frac{Y}{1 - Y} = \frac{pO_2{}^n}{P_{50}{}^n}$$

and so

$$\log\left(\frac{Y}{1 - Y}\right) = \log\left(\frac{pO_2{}^n}{P_{50}{}^n}\right) = n\log(pO_2) - n\log(P_{50})$$

This equation predicts that a plot of $\log(Y/1 - Y)$ versus $\log(P_{50})$, called a *Hill plot*, should be linear with a slope of n.

Hill plots for myoglobin and hemoglobin are shown in Figure 7.29. For myoglobin, the Hill plot is linear with a slope of 1. For hemoglobin, the Hill plot is not completely linear, because the equilibrium on which the Hill plot is based is not entirely correct. However, the plot is approximately linear in the center with a slope of 2.8. The slope, often referred to as the *Hill coefficient*, is a measure of the cooperativity of oxygen binding. The utility of the Hill plot is that it provides a simply derived quantitative assessment of the degree of cooperativity in binding. With the use of the Hill equation and the derived Hill coefficient, a binding curve that closely resembles that for hemoglobin is produced (Figure 7.30).

The Concerted Model

The concerted model can be formulated in quantitative terms. Only four parameters are required: (1) the number of binding sites (assumed to be equivalent) in the protein, (2) the ratio of the concentrations of the T and R states in the absence of bound ligands, (3) the affinity of sites in proteins in the R state for ligand binding, and (4) a measure of how much more tightly subunits in proteins in the R state bind ligands compared with subunits in the T state. The number of binding sites, n, is usually known from other information. For hemoglobin,

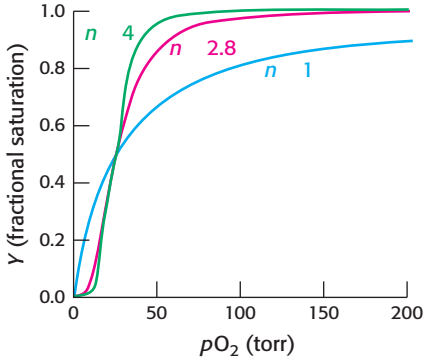


Figure 7.30 Oxygen-binding curves for several Hill coefficients. The curve labeled $n = 2.8$ closely resembles the curve for hemoglobin.

$n = 4$. The ratio of the concentrations of the T and R states with no ligands bound is a constant:

$$L = [T_0]/[R_0]$$

where the subscript refers to the number of ligands bound (in this case, zero). The affinity of subunits in the R state is defined by the dissociation constant for a ligand binding to a single site in the R state, K_R. Similarly, the dissociation constant for a ligand binding to a single site in the T state is K_T. We can define the ratio of these two dissociation constant as

$$c = K_R/K_T$$

This is the measure of how much more tightly a subunit for a protein in the R state binds a ligand compared with a subunit for a protein in the T state. Note that $c < 1$ because K_R and K_T are dissociation constants and tight binding corresponds to a small dissociation constant.

What is the ratio of the concentration of T-state proteins with one ligand bound to the concentration of R-state proteins with one ligand bound? The dissociation constant for a single site in the R state is K_R. For a protein with n sites, there are n possible sites for the first ligand to bind. This statistical factor favors ligand binding compared with a single-site protein. Thus, $[R_1] = n[R_0][S]/K_R$. Similarly, $[T_1] = n[T_0][S]/K_T$. Thus,

$$[T_1]/[R_1] = \frac{n[T_0][S]/K_T}{n[R_0][S]/K_R} = \frac{[T_0]}{[R_0](K_R/K_T)} = cL$$

Similar analysis reveals that, for states with i ligands bound, $[T_i]/[R_i] = c^iL$. In other words, the ratio of the concentrations of the T state to the R state is reduced by a factor of c for each ligand that binds.

Let us define a convenient scale for the concentration of S:

$$\alpha = [S]/K_R$$

This definition is useful because it is the ratio of the concentration of S to the dissociation constant that determines the extent of binding. Using this definition, we see that

$$[R_1] = \frac{n[R_0][S]}{K_R} = n[R_0]\alpha$$

Similarly,

$$[T_1] = \frac{n[T_0][S]}{K_T} = ncL[R_0]\alpha$$

What is the concentration of R-state molecules with two ligands bound? Again, we must consider the statistical factor—that is, the number of ways in which a second ligand can bind to a molecule with one site occupied. The number of ways is $n - 1$. However, because which ligand is the "first" and which is the "second" does not matter, we must divide by a factor of 2. Thus,

$$\begin{aligned}[R_2] &= \frac{\left(\frac{n-1}{2}\right)[R_1][S]}{K_R} \\ &= \left(\frac{n-1}{2}\right)[R_1]\alpha \\ &= \left(\frac{n-1}{2}\right)(n[R_0]\alpha)\alpha \\ &= n\left(\frac{n-1}{2}\right)[K_0]\alpha^2\end{aligned}$$

We can derive similar equations for the case with i ligands bound and for T states.

We can now calculate the fractional saturation, Y. This is the total concentration of sites with ligands bound divided by the total concentration of potential binding sites. Thus,

$$Y = \frac{([R_1] + [T_1]) + 2([R_2] + [T_2]) + \cdots + n([R_n] + [T_n])}{n([R_0] + [T_0] + [R_1] + [T_1] + \cdots + [R_n] + [T_n])}$$

Substituting into this equation, we find

$$Y = \frac{n[R_0]\alpha + nc[T_0]\alpha + 2(n(n-1)/2)[R_0]\alpha^2 + 2(n(n-1)/2)c^2[T_0]\alpha^2 + \cdots + n[R_0]\alpha^n + nc^n[T^0])\alpha^n}{n([R_0] + [T_0] + n[R_0]\alpha + nc[T_0]\alpha + \cdots + [R_0]\alpha^n + c^n[T_0]\alpha^n)}$$

Substituting $[T_0] = L[R_0]$ and summing these series yields

$$Y = \frac{\alpha(1 + \alpha)^{n-1} + Lc\alpha(1 + c\alpha)^{n-1}}{(1 + \alpha)^n + L(1 + c\alpha)^n}$$

We can now use this equation to fit the observed data for hemoglobin by varying the parameters L, c, and K_R (with $n = 4$). An excellent fit is obtained with $L = 9000$, $c = 0.014$, and $K_R = 2.5$ torr (Figure 7.31).

In addition to the fractional saturation, the concentrations of the species T_0, T_1, T_2, R_2, R_3, and R_4 are shown. The concentrations of all other species are very low. The addition of concentrations is a major difference between the analysis using the Hill equation and this analysis of the concerted model. The Hill equation gives only the fractional saturation, whereas the

analysis of the concerted model yields concentrations for all species. In the present case, this analysis yields the expected ratio of T-state proteins to R-state proteins at each stage of binding. This ratio changes from 9000 to 126 to 1.76 to 0.025 to 0.00035 with zero, one, two, three, and four oxygen molecules bound. This ratio provides a quantitative measure of the switching of the population of hemoglobin molecules from the T state to the R state.

The sequential model can also be formulated in quantitative terms. However, the formulation entails many more parameters, and many different sets of parameters often yield similar fits to the experimental data.

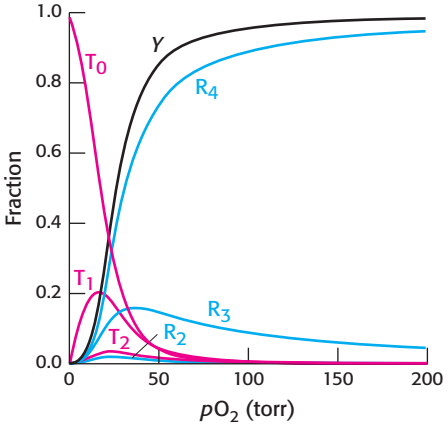


Figure 7.31 Modeling oxygen binding with the concerted model. The fractional saturation (Y) as a function pO_2: $L = 9000$, $c = 0.014$, and $K_R = 2.5$ torr. The fraction of molecules in the T state with zero, one, and two oxygen molecules bound (T_0, T_1, and T_2) and the fraction of molecules in the R state with two, three, and four oxygen molecules bound (R_2, R_3, and R_4) are shown. The fractions of molecules in other forms are too low to be shown.

Key Terms

heme (p. 196)
protoporphyrin (p. 196)
proximal histidine (p. 197)
functional magnetic resonance imaging (fMRI) (p. 197)
superoxide anion (p. 198)
metmyoglobin (p. 198)
distal histidine (p. 198)
α chain (p. 199)
β chain (p. 199)
globin fold (p. 199)
αβ dimer (p. 199)
oxygen-binding curve (p. 199)
fractional saturation (p. 199)
partial pressure (p. 200)
sigmoid (p. 200)
cooperative binding (p. 200)
T state (p. 202)
R state (p. 202)
concerted model (MWC model) (p. 202)
sequential model (p. 203)
2,3-bisphosphoglycerate (p. 204)
fetal hemoglobin (p. 205)
carbon monoxide (p. 205)
carboxyhemoglobin (p. 205)
Bohr effect (p. 206)
carbonic anhydrase (p. 207)
carbamate (p. 208)
sickle-cell anemia (p. 209)
hemoglobin S (p. 209)
malaria (p. 210)
thalassemia (p. 210)
hemoglobin H (p. 210)
thalassemia major (Cooley anemia) (p. 210)
α-hemoglobin stabilizing protein (AHSP) (p. 211)
neuroglobin (p. 212)
cytoglobin (p. 212)
Hill plot (p. 214)
Hill coefficient (p. 214)

Problems

1. *Screening the biosphere.* The first protein structure to have its structure determined was myoglobin from sperm whale. Propose an explanation for the observation that sperm whale muscle is a rich source of this protein.

2. *Hemoglobin content.* The average volume of a red blood cell is 87 μm^3. The mean concentration of hemoglobin in red cells is 0.34 g ml^{-1}.

(a) What is the weight of the hemoglobin contained in an average red cell?

(b) How many hemoglobin molecules are there in an average red cell? Assume that the molecular weight of the human hemoglobin tetramer is 65 kd.

(c) Could the hemoglobin concentration in red cells be much higher than the observed value? (Hint: Suppose that

a red cell contained a crystalline array of hemoglobin molecules in a cubic lattice with 65 Å sides.)

3. *Iron content.* How much iron is there in the hemoglobin of a 70-kg adult? Assume that the blood volume is 70 ml kg^{-1} of body weight and that the hemoglobin content of blood is 0.16 g ml^{-1}.

4. *Oxygenating myoglobin.* The myoglobin content of some human muscles is about 8 g kg^{-1}. In sperm whale, the myoglobin content of muscle is about 80 g kg^{-1}.

(a) How much O_2 is bound to myoglobin in human muscle and in sperm whale muscle? Assume that the myoglobin is saturated with O_2, and that the molecular weights of human and sperm whale myoglobin are the same.

(b) The amount of oxygen dissolved in tissue water (in equilibrium with venous blood) at 37°C is about 3.5×10^{-5} M. What is the ratio of oxygen bound to myoglobin to that directly dissolved in the water of sperm whale muscle?

5. *Tuning proton affinity.* The pK_a of an acid depends partly on its environment. Predict the effect of each of the following environmental changes on the pK_a of a glutamic acid side chain.

(a) A lysine side chain is brought into proximity.

(b) The terminal carboxyl group of the protein is brought into proximity.

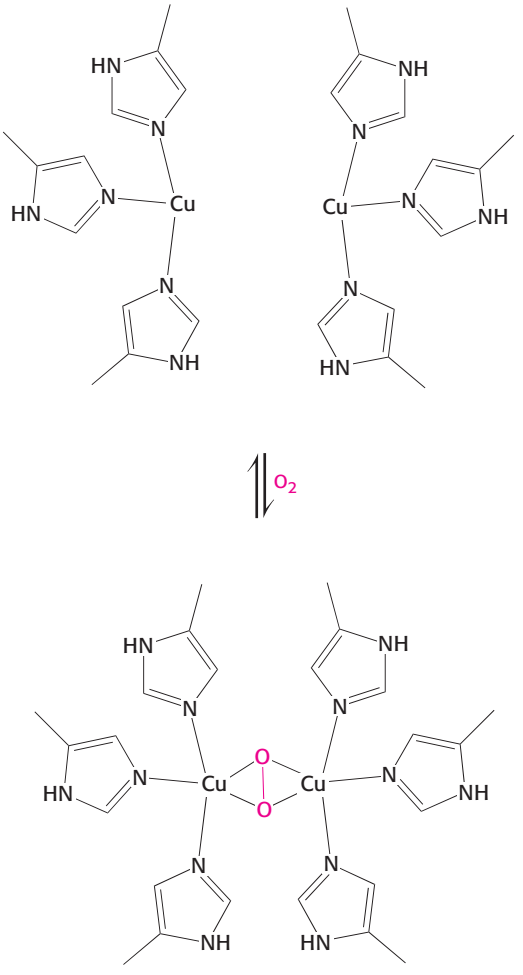
(c) The glutamic acid side chain is shifted from the outside of the protein to a nonpolar site inside.

6. *Saving grace.* Hemoglobin A inhibits the formation of the long fibers of hemoglobin S and the subsequent sickling of the red cell on deoxygenation. Why does hemoglobin A have this effect?

7. *Carrying a load.* Suppose that you are climbing a high mountain and the oxygen partial pressure in the air is reduced to 75 torr. Estimate the percentage of the oxygen-carrying capacity that will be utilized, assuming that the pH of both tissues and lungs is 7.4 and that the oxygen concentration in the tissues is 20 torr.

8. *High-altitude adaptation.* After spending a day or more at high altitude (with an oxygen partial pressure of 75 torr), the concentration of 2,3-bisphosphoglycerate (2,3-BPG) in red blood cells increases. What effect would an increased concentration of 2,3-BPG have on the oxygen-binding curve for hemoglobin? Why would this adaptation be beneficial for functioning well at high altitude?

9. *I'll take the lobster.* Arthropods such as lobsters have oxygen carriers quite different from hemoglobin. The oxygen-binding sites do not contain heme but, instead, are based on two copper(I) ions. The structural changes that accompany oxygen binding are shown below. How might these changes be used to facilitate cooperative oxygen binding?



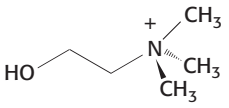
10. *A disconnect.* With the use of site-directed mutagenesis, hemoglobin has been prepared in which the proximal histidine residues in both the α and the β subunits have been replaced by glycine. The imidazole ring from the histidine residue can be replaced by adding free imidazole in solution. Would you expect this modified hemoglobin to show cooperativity in oxygen binding? Why or why not?

Imidazole

11. *Successful substitution.* Blood cells from some birds do not contain 2,3-bisphosphoglycerate but, instead, contain one of the compounds in parts *a* through *d*, which plays an

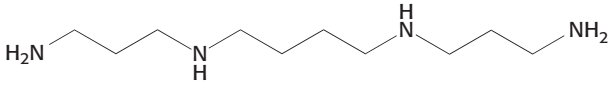
analogous functional role. Which compound do you think is most likely to play this role? Explain briefly.

(a)



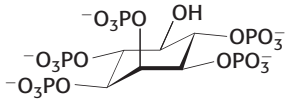
Choline

(b)



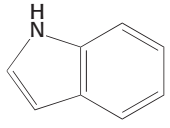
Spermine

(c)



Inositol pentaphosphate

(d)



Indole

12. *Theoretical curves.* (a) Using the Hill equation, plot an oxygen-binding curve for a hypothetical two-subunit hemoglobin with $n = 1.8$ and $P_{50} = 10$ torr. (b) Repeat, using the concerted model with $n = 2$, $L = 1000$, $c = 0.01$, and $K_R = 1$ torr.

13. *Parasitic effect.* When *P. falciparum* lives inside red blood cells, the metabolism of the parasite tends to release acid. What effect is the presence of acid likely to have on the oxygen-carrying capacity of the red blood cells? On the likelihood that these cells sickle?

Data Interpretation Problems

14. *Primitive oxygen binding.* Lampreys are primitive organisms whose ancestors diverged from the ancestors of fish and mammals approximately 400 million years ago. Lamprey blood contains a hemoglobin related to mammalian hemoglobin. However, lamprey hemoglobin is monomeric in the oxygenated state. Oxygen-binding data for lamprey hemoglobin are as follows:

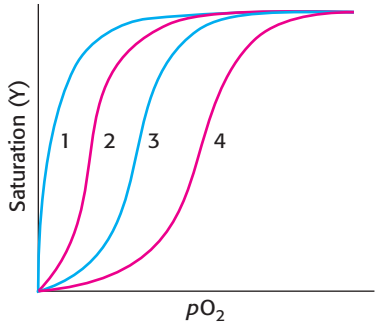
pO_2	Y	pO_2	Y	pO_2	Y
0.1	.0060	2.0	.112	50.0	.889
0.2	.0124	3.0	.170	60.0	.905
0.3	.0190	4.0	.227	70.0	.917
0.4	.0245	5.0	.283	80.0	.927
0.5	.0307	7.5	.420	90.0	.935
0.6	.0380	10.0	.500	100	.941
0.7	.0430	15.0	.640	150	.960
0.8	.0481	20.0	.721	200	.970
0.9	.0530	30.0	.812		
1.0	.0591	40.0	.865		

(a) Plot these data to produce an oxygen-binding curve. At what oxygen partial pressure is this hemoglobin half-saturated? On the basis of the appearance of this curve, does oxygen binding seem to be cooperative?

(b) Construct a Hill plot using these data. Does the Hill plot show any evidence for cooperativity? What is the Hill coefficient?

(c) Further studies revealed that lamprey hemoglobin forms oligomers, primarily dimers, in the deoxygenated state. Propose a model to explain any observed cooperativity in oxygen binding by lamprey hemoglobin.

15. *Leaning to the left or to the right.* The illustration below shows several oxygen-dissociation curves. Assume that curve 3 corresponds to hemoglobin with physiological concentrations of CO_2 and 2,3-BPG at pH 7. Which curves represent each of the following perturbations?



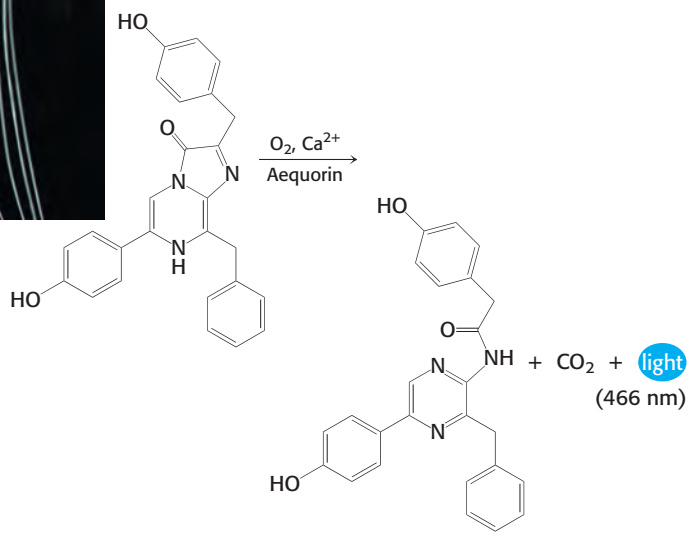
(a) Decrease in CO_2
(b) Increase in 2,3-BPG
(c) Increase in pH
(d) Loss of quaternary structure

Chapter Integration Problem

16. *Location is everything.* 2,3-Bisphosphoglycerate lies in a central cavity within the hemoglobin tetramer, stabilizing the T state. What would be the effect of mutations that placed the BPG-binding site on the surface of hemoglobin?

Enzymes: Basic Concepts and Kinetics

The activity of an enzyme is responsible for the glow of the luminescent jellyfish at left. The enzyme aequorin catalyzes the oxidation of a compound by oxygen in the presence of calcium to release CO_2 and light. [(Left) Lesya Castillo/Featurepics.]



Enzymes, the catalysts of biological systems, are remarkable molecular devices that determine the patterns of chemical transformations. They also mediate the transformation of one form of energy into another. About a quarter of the genes in the human genome encode enzymes, a testament to their importance to life. The most striking characteristics of enzymes are their *catalytic power* and *specificity*. Catalysis takes place at a particular site on the enzyme called the *active site*.

Nearly all known enzymes are proteins. However, proteins do not have an absolute monopoly on catalysis; the discovery of catalytically active RNA molecules provides compelling evidence that RNA was a biocatalyst early in evolution. Proteins as a class of macromolecules are highly effective catalysts for an enormous diversity of chemical reactions because of their capacity *to specifically bind a very wide range of molecules.* By utilizing the full repertoire of intermolecular forces, enzymes bring substrates together in an optimal orientation, the prelude to making and breaking chemical bonds. They catalyze reactions *by stabilizing transition states,* the highest-energy species in reaction pathways. By selectively stabilizing a transition state, an enzyme determines which one of several potential chemical reactions actually takes place.

OUTLINE

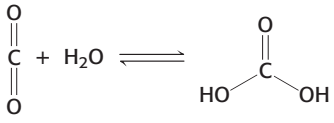
Table 8.1 Rate enhancement by selected enzymes

Enzyme	Nonenzymatic half-life		Uncatalyzed rate (k_{un} s^{-1})	Catalyzed rate (k_{cat} s^{-1})	Rate enhancement (k_{cat} s^{-1}/k_{un} s^{-1})
OMP decarboxylase	78,000,000	years	2.8×10^{-16}	39	1.4×10^{17}
Staphylococcal nuclease	130,000	years	1.7×10^{-13}	95	5.6×10^{14}
AMP nucleosidase	69,000	years	1.0×10^{-11}	60	6.0×10^{12}
Carboxypeptidase A	7.3	years	3.0×10^{-9}	578	1.9×10^{11}
Ketosteroid isomerase	7	weeks	1.7×10^{-7}	66,000	3.9×10^{11}
Triose phosphate isomerase	1.9	days	4.3×10^{-6}	4,300	1.0×10^{9}
Chorismate mutase	7.4	hours	2.6×10^{-5}	50	1.9×10^{6}
Carbonic anhydrase	5	seconds	1.3×10^{-1}	1×10^{6}	7.7×10^{6}

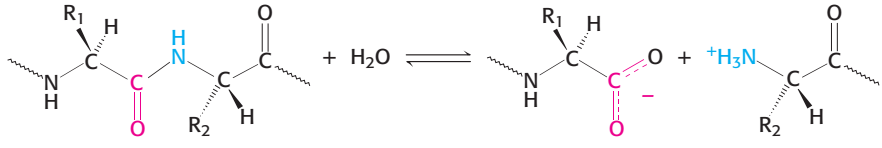
Abbreviations: OMP, orotidine monophosphate; AMP, adenosine monophosphate.
Source: After A. Radzicka and R. Wolenden. *Science* 267:90–93, 1995.

8.1 Enzymes Are Powerful and Highly Specific Catalysts

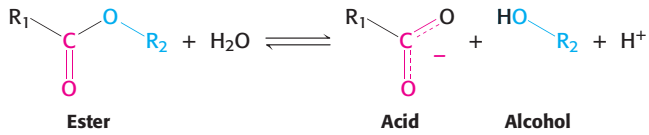
Enzymes accelerate reactions by factors of as much as a million or more (Table 8.1). Indeed, most reactions in biological systems do not take place at perceptible rates in the absence of enzymes. Even a reaction as simple as the hydration of carbon dioxide is catalyzed by an enzyme—namely, carbonic anhydrase (Section 9.2). The transfer of CO_2 from the tissues to the blood and then to the air in the alveolae of the lungs would be less complete in the absence of this enzyme. In fact, carbonic anhydrase is one of the fastest enzymes known. Each enzyme molecule can hydrate 10^6 molecules of CO_2 *per second.* This catalyzed reaction is 10^7 times as fast as the uncatalyzed one. We will consider the mechanism of carbonic anhydrase catalysis in Chapter 9.



Enzymes are highly specific both in the reactions that they catalyze and in their choice of reactants, which are called *substrates*. An enzyme usually catalyzes a single chemical reaction or a set of closely related reactions. Let us consider *proteolytic enzymes* as an example. In vivo, these enzymes catalyze *proteolysis,* the hydrolysis of a peptide bond.



Most proteolytic enzymes also catalyze a different but related reaction in vitro—namely, the hydrolysis of an ester bond. Such reactions are more easily monitored than is proteolysis and are useful in experimental investigations of these enzymes.



Proteolytic enzymes differ markedly in their degree of substrate specificity. Papain, which is found in papaya plants, is quite undiscriminating: it will cleave any peptide bond with little regard to the identity of the adjacent side chains. This lack of specificity accounts for its use in meat-tenderizing sauces. The digestive enzyme trypsin, on the other hand, is quite specific and

catalyzes the splitting of peptide bonds only on the carboxyl side of lysine and arginine residues (Figure 8.1A). Thrombin, an enzyme that participates in blood clotting, is even more specific than trypsin. It catalyzes the hydrolysis of Arg–Gly bonds in particular peptide sequences only (Figure 8.1B).

DNA polymerase I, a template-directed enzyme (Section 28.3), is another highly specific catalyst. To a DNA strand that is being synthesized, it adds nucleotides in a sequence determined by the sequence of nucleotides in another DNA strand that serves as a template. DNA polymerase I is remarkably precise in carrying out the instructions given by the template. It inserts the wrong nucleotide into a new DNA strand less than one in a thousand times. *The specificity of an enzyme is due to the precise interaction of the substrate with the enzyme. This precision is a result of the intricate three-dimensional structure of the enzyme protein.*

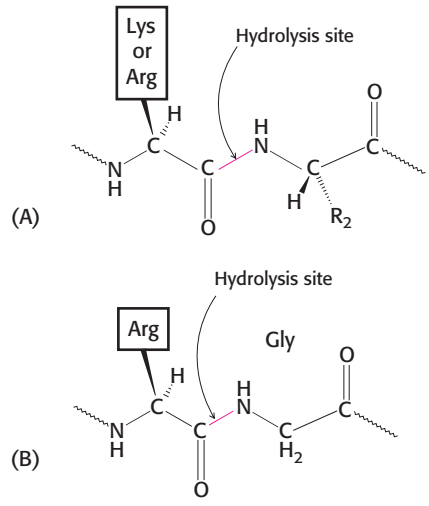


Figure 8.1 Enzyme specificity. (A) Trypsin cleaves on the carboxyl side of arginine and lysine residues, whereas (B) thrombin cleaves Arg–Gly bonds in particular sequences only.

Many enzymes require cofactors for activity

The catalytic activity of many enzymes depends on the presence of small molecules termed *cofactors,* although the precise role varies with the cofactor and the enzyme. Generally, these cofactors are able to execute chemical reactions that cannot be performed by the standard set of twenty amino acids. An enzyme without its cofactor is referred to as an *apoenzyme;* the complete, catalytically active enzyme is called a *holoenzyme*.

$$\text{Apoenzyme} + \text{cofactor} = \text{holoenzyme}$$

Cofactors can be subdivided into two groups: (1) metals and (2) small organic molecules called *coenzymes* (Table 8.2). Often derived from vitamins, coenzymes can be either tightly or loosely bound to the enzyme. Tightly bound coenzymes are called *prosthetic groups*. Loosely associated coenzymes are more like cosubstrates because, like substrates and products, they bind to the enzyme and are released from it. The use of the same coenzyme by a variety of enzymes sets coenzymes apart from normal substrates, however, as does their source in vitamins (Section 15.4). Enzymes that use the same coenzyme usually perform catalysis by similar mechanisms. In Chapter 9, we will examine the importance of metals to enzyme activity and, throughout the book, we will see how coenzymes and their enzyme partners operate in their biochemical context.

Table 8.2 Enzyme cofactors

Cofactor	Enzyme
Coenzyme	
Thiamine pyrophosphate	Pyruvate dehydrogenase
Flavin adenine nucleotide	Monoamine oxidase
Nicotinamide adenine dinucleotide	Lactate dehydrogenase
Pyridoxal phosphate	Glycogen phosphorylase
Coenzyme A (CoA)	Acetyl CoA carboxylase
Biotin	Pyruvate carboxylase
5′-Deoxyadenosyl cobalamin	Methylmalonyl mutase
Tetrahydrofolate	Thymidylate synthase
Metal	
Zn^{2+}	Carbonic anhydrase
Zn^{2+}	Carboxypeptidase
Mg^{2+}	*Eco*RV
Mg^{2+}	Hexokinase
Ni^{2+}	Urease
Mo	Nitrate reductase
Se	Glutathione peroxidase
Mn	Superoxide dismutase
K^{+}	Propionyl CoA carboxylase

Enzymes can transform energy from one form into another

A key activity in all living systems is the ability to convert one form of energy into another. For example, in photosynthesis, light energy is converted into chemical-bond energy. In cellular respiration, which takes place in mitochondria, the free energy contained in small molecules derived from food is converted first into the free energy of an ion gradient and then into a different currency—the free energy of adenosine triphosphate. Given their centrality to life, it should come as no surprise that enzymes play vital roles in energy transformation. As we will see, enzymes play fundamental roles in photosynthesis and cellular respiration. Other enzymes can then use the chemical-bond energy of ATP in diverse ways. For instance, the enzyme myosin converts the energy of ATP into the mechanical energy of contracting muscles

(Chapter 35). Pumps in the membranes of cells and organelles, which can be thought of as enzymes that move substrates rather than chemically alter them, use the energy of ATP to transport molecules and ions across the membrane (Chapter 13). The chemical and electrical gradients resulting from the unequal distribution of these molecules and ions are themselves forms of energy that can be used for a variety of purposes, such as sending nerve impulses.

The molecular mechanisms of these energy-transducing enzymes are being unraveled. We will see in subsequent chapters how unidirectional cycles of discrete steps—binding, chemical transformation, and release—lead to the conversion of one form of energy into another.

8.2 Free Energy Is a Useful Thermodynamic Function for Understanding Enzymes

Enzymes speed up the rate of chemical reactions, but the properties of the reaction—whether it can take place at all and the degree to which the enzyme accelerates the reaction—depend on energy differences between reactants and products. *Free energy (G),* which was touched on in Chapter 1, is a thermodynamic property that is a measure of useful energy, or the energy that is capable of doing work. To understand how enzymes operate, we need to consider only two thermodynamic properties of the reaction: (1) the free-energy difference (ΔG) between the products and reactants and (2) the energy required to initiate the conversion of reactants into products. The former determines whether the reaction will take place spontaneously, whereas the latter determines the rate of the reaction. Enzymes affect only the latter. Let us review some of the principles of thermodynamics as they apply to enzymes.

The free-energy change provides information about the spontaneity but not the rate of a reaction

As discussed in Chapter 1, the free-energy change of a reaction (ΔG) tells us if the reaction can take place spontaneously:

1. *A reaction can take place spontaneously only if* ΔG *is negative.* Such reactions are said to be *exergonic*.

2. A system is at equilibrium and no *net* change can take place if ΔG is zero.

3. A reaction cannot take place spontaneously if ΔG is positive. An input of free energy is required to drive such a reaction. These reactions are termed *endergonic*.

4. The ΔG of a reaction depends only on the free energy of the products (the final state) minus the free energy of the reactants (the initial state). *The* ΔG *of a reaction is independent of the path (or molecular mechanism) of the transformation.* The mechanism of a reaction has no effect on ΔG. For example, the ΔG for the oxidation of glucose to CO_2 and H_2O is the same whether it takes place by combustion or by a series of enzyme-catalyzed steps in a cell.

5. *The* ΔG *provides no information about the rate of a reaction.* A negative ΔG indicates that a reaction *can* take place spontaneously, but it does not signify whether it will proceed at a perceptible rate. As will be discussed shortly (Section 8.3), the rate of a reaction depends on the *free energy of activation* ($\Delta G^{\ddagger}$), which is largely unrelated to the ΔG of the reaction.

The standard free-energy change of a reaction is related to the equilibrium constant

As for any reaction, we need to be able to determine ΔG for an enzyme-catalyzed reaction to know whether the reaction is spontaneous or an input of energy is required. To determine this important thermodynamic parameter, we need to take into account the nature of both the reactants and the products as well as their concentrations.

Consider the reaction

$$A + B \rightleftharpoons C + D$$

The ΔG of this reaction is given by

$$\Delta G = \Delta G^\circ + RT \ln \frac{[C][D]}{[A][B]} \quad (1)$$

in which ΔG° is the *standard free-energy change*, R is the gas constant, T is the absolute temperature, and [A], [B], [C], and [D] are the molar concentrations (more precisely, the activities) of the reactants. ΔG° is the free-energy change for this reaction under standard conditions—that is, when each of the reactants A, B, C, and D is present at a concentration of 1.0 M (for a gas, the standard state is usually chosen to be 1 atmosphere). Thus, the ΔG of a reaction depends on the *nature* of the reactants (expressed in the ΔG° term of equation 1) and on their *concentrations* (expressed in the logarithmic term of equation 1).

A convention has been adopted to simplify free-energy calculations for biochemical reactions. The standard state is defined as having a pH of 7. Consequently, when H^+ is a reactant, its activity has the value 1 (corresponding to a pH of 7) in equations 1 and 3 (below). The activity of water also is taken to be 1 in these equations. The *standard free-energy change at pH 7*, denoted by the symbol $\Delta G^{\circ\prime}$, will be used throughout this book. The *kilojoule* (abbreviated *kJ*) and the *kilocalorie* (*kcal*) will be used as the units of energy. One kilojoule is equivalent to 0.239 kilocalorie.

Units of energy

A *kilojoule* (kJ) is equal to 1000 J.
A *joule* (J) is the amount of energy needed to apply a 1-newton force over a distance of 1 meter.
A *kilocalorie* (kcal) is equal to 1000 cal.
A *calorie* (cal) is equivalent to the amount of heat required to raise the temperature of 1 gram of water from 14.5°C to 15.5°C.
1 kJ = 0.239 kcal.

A simple way to determine $\Delta G^{\circ\prime}$ is to measure the concentrations of reactants and products when the reaction has reached equilibrium. At equilibrium, there is no net change in reactants and products; in essence, the reaction has stopped and $\Delta G = 0$. At equilibrium, equation 1 then becomes

$$0 = \Delta G^{\circ\prime} + RT \ln \frac{[C][D]}{[A][B]} \quad (2)$$

and so

$$\Delta G^{\circ\prime} = -RT \ln \frac{[C][D]}{[A][B]} \quad (3)$$

The equilibrium constant under standard conditions, K'_{eq}, is defined as

$$K'_{eq} = \frac{[C][D]}{[A][B]} \quad (4)$$

Substituting equation 4 into equation 3 gives

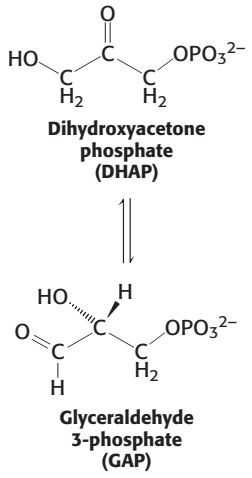
$$\Delta G^{\circ\prime} = -RT \ln K'_{eq} \quad (5)$$

which can be rearranged to give

$$K'_{eq} = 10^{-\Delta G^{\circ\prime}/RT} \quad (6)$$

Table 8.3 Relation between $\Delta G^{\circ\prime}$ and K'_{eq} (at 25°C)

K'_{eq}	$\Delta G^{\circ\prime}$ kJ mol^{-1}	$\Delta G^{\circ\prime}$ kcal mol^{-1}
10^{-5}	28.53	6.82
10^{-4}	22.84	5.46
10^{-3}	17.11	4.09
10^{-2}	11.42	2.73
10^{-1}	5.69	1.36
1	0.00	0.00
10	−5.69	−1.36
10^{2}	−11.42	−2.73
10^{3}	−17.11	−4.09
10^{4}	−22.84	−5.46
10^{5}	−28.53	−6.82



Substituting $R = 8.315 \times 10^{-3}$ kJ mol^{-1} deg^{-1} and $T = 298$ K (corresponding to 25°C) gives

$$K'_{eq} = 10^{-\Delta G^{\circ\prime}/2.47} \tag{7}$$

where $\Delta G^{\circ\prime}$ is here expressed in kilojoules per mole because of the choice of the units for R in equation 7. Thus, the standard free energy and the equilibrium constant of a reaction are related by a simple expression. For example, an equilibrium constant of 10 gives a standard free-energy change of −5.69 kJ mol^{-1} (−1.36 kcal mol^{-1}) at 25°C (Table 8.3). Note that, for each 10-fold change in the equilibrium constant, the $\Delta G^{\circ\prime}$ changes by 5.69 kJ mol^{-1} (1.36 kcal mol^{-1}).

As an example, let us calculate $\Delta G^{\circ\prime}$ and ΔG for the isomerization of dihydroxyacetone phosphate (DHAP) to glyceraldehyde 3-phosphate (GAP). This reaction takes place in glycolysis (Chapter 16). At equilibrium, the ratio of GAP to DHAP is 0.0475 at 25°C (298 K) and pH 7. Hence, $K'_{eq} = 0.0475$. The standard free-energy change for this reaction is then calculated from equation 5:

$$\begin{aligned}\Delta G^{\circ\prime} &= -RT \ln K'_{eq} \\ &= -8.315 \times 10^{-3} \times 298 \times \ln (0.0475) \\ &= +7.53 \text{ kJ mol}^{-1} (+1.80 \text{ kcal mol}^{-1})\end{aligned}$$

Under these conditions, the reaction is endergonic. DHAP will not spontaneously convert into GAP.

Now let us calculate ΔG for this reaction when the initial concentration of DHAP is 2×10^{-4} M and the initial concentration of GAP is 3×10^{-6} M. Substituting these values into equation 1 gives

$$\begin{aligned}\Delta G &= 7.53 \text{ kJ mol}^{-1} + RT \ln \frac{3 \times 10^{-6} \text{ M}}{2 \times 10^{-4} \text{ M}} \\ &= 7.53 \text{ kJ mol}^{-1} - 10.42 \text{ kJ mol}^{-1} \\ &= -2.89 \text{ kJ mol}^{-1} (-0.69 \text{ kcal mol}^{-1})\end{aligned}$$

This negative value for the ΔG indicates that the isomerization of DHAP to GAP is exergonic and can take place spontaneously when these species are present at the preceding concentrations. Note that ΔG for this reaction is negative, although $\Delta G^{\circ\prime}$ is positive. *It is important to stress that whether the* ΔG *for a reaction is larger, smaller, or the same as* ΔG$^{\circ\prime}$ *depends on the concentrations of the reactants and products*. The criterion of spontaneity for a reaction is ΔG, not $\Delta G^{\circ\prime}$. This point is important because reactions that are not spontaneous based on $\Delta G^{\circ\prime}$ can be made spontaneous by adjusting the concentrations of reactants and products. This principle is the basis of the coupling of reactions to form metabolic pathways (Chapter 15).

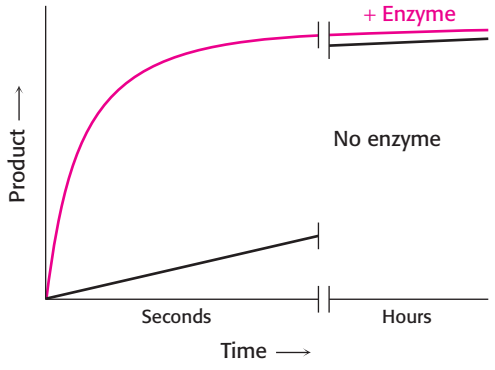


Figure 8.2 Enzymes accelerate the reaction rate. The same equilibrium point is reached but much more quickly in the presence of an enzyme.

Enzymes alter only the reaction rate and not the reaction equilibrium

Because enzymes are such superb catalysts, it is tempting to ascribe to them powers that they do not have. An enzyme cannot alter the laws of thermodynamics and *consequently cannot alter the equilibrium of a chemical reaction*. Consider an enzyme-catalyzed reaction, the conversion of substrate, S, into product, P. Figure 8.2 shows the rate of product formation with time in the presence and absence of enzyme. Note that the amount of product formed is the same whether or not the enzyme

is present but, in the present example, the amount of product formed in seconds when the enzyme is present might take hours (or centuries, see Table 8.1) to form if the enzyme were absent.

Why does the rate of product formation level off with time? The reaction has reached equilibrium. Substrate S is still being converted into product P, but P is being converted into S at a rate such that the amount of P present stays the same.

Let us examine the equilibrium in a more quantitative way. Suppose that, in the absence of enzyme, the forward rate constant (k_F) for the conversion of S into P is $10^{-4}\ s^{-1}$ and the reverse rate constant (k_R) for the conversion of P into S is $10^{-6}\ s^{-1}$. The equilibrium constant K is given by the ratio of these rate constants:

$$S \underset{10^{-6}\ s^{-1}}{\overset{10^{-4}\ s^{-1}}{\rightleftharpoons}} P$$

$$K = \frac{[P]}{[S]} = \frac{k_F}{k_R} = \frac{10^{-4}}{10^{-6}} = 100$$

The equilibrium concentration of P is 100 times that of S, whether or not enzyme is present. However, it might take a very long time to approach this equilibrium without enzyme, whereas equilibrium would be attained rapidly in the presence of a suitable enzyme (see Table 8.1). *Enzymes accelerate the attainment of equilibria but do not shift their positions. The equilibrium position is a function only of the free-energy difference between reactants and products.*

8.3 Enzymes Accelerate Reactions by Facilitating the Formation of the Transition State

The free-energy difference between reactants and products accounts for the equilibrium of the reaction, but enzymes accelerate how quickly this equilibrium is attained. How can we explain the rate enhancement in terms of thermodynamics? To do so, we have to consider not the end points of the reaction but the chemical pathway between the end points.

A chemical reaction of substrate S to form product P goes through a *transition state* $X^{\ddagger}$ that has a higher free energy than does either S or P.

$$S \longrightarrow X^{\ddagger} \longrightarrow P$$

The double dagger denotes the transition state. The transition state is a transitory molecular structure that is no longer the substrate but is not yet the product. The transition state is the least-stable and most-seldom-occupied species along the reaction pathway because it is the one with the highest free energy. The difference in free energy between the transition state and the substrate is called the *Gibbs free energy of activation* or simply the *activation energy,* symbolized by $\Delta G^{\ddagger}$ (Figure 8.3).

$$\Delta G^{\ddagger} = G_X{}^{\ddagger} - G_S$$

Note that the energy of activation, or $\Delta G^{\ddagger}$, does not enter into the final ΔG calculation for the reaction, because the energy required to generate the transition state is released when the transition state forms the product. The activation-energy barrier immediately suggests how an enzyme enhances the reaction rate without altering ΔG of the reaction: enzymes function to lower the activation energy, or, in other words, *enzymes facilitate the formation of the transition state.*

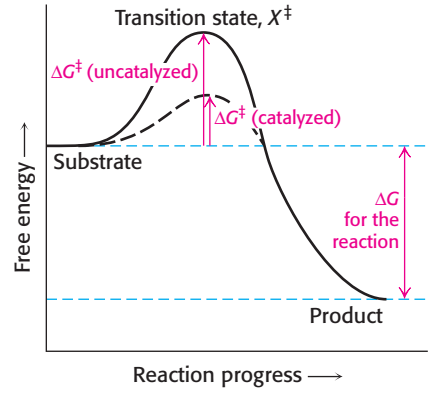


Figure 8.3 Enzymes decrease the activation energy. Enzymes accelerate reactions by decreasing $\Delta G^{\ddagger}$, the free energy of activation.

One approach to understanding the increase in reaction rates achieved by enzymes is to assume that the transition state ($X^{\ddagger}$) and the substrate (S) are in equilibrium.

$$S \overset{K^{\ddagger}}{\rightleftharpoons} X^{\ddagger} \xrightarrow{v} P$$

in which $K^{\ddagger}$ is the equilibrium constant for the formation of $X^{\ddagger}$ and v is the rate of formation of product from $X^{\ddagger}$. The rate of the reaction v is proportional to the concentration of $X^{\ddagger}$,

$$v \propto [X^{\ddagger}],$$

because only $X^{\ddagger}$ can be converted into product. The concentration of $X^{\ddagger}$ at equilibrium is in turn related to the free-energy difference $\Delta G^{\ddagger}$ between $X^{\ddagger}$ and S; the greater the difference in free energy between these two states, the smaller the amount of $X^{\ddagger}$. Thus, the overall rate of reaction V depends on $\Delta G^{\ddagger}$. Specifically,

$$V = v[X^{\ddagger}] = \frac{kT}{h}[S]e^{-\Delta G^{\ddagger}/RT}$$

In this equation, k is Boltzmann's constant, and h is Planck's constant. The value of kT/h at 25°C is $6.6 \times 10^{12}\ s^{-1}$. Suppose that the free energy of activation is 28.53 kJ mol^{-1} (6.82 kcal mol^{-1}). If we were to substitute this value of ΔG in equation 7 (as shown in Table 8.3), this free-energy difference will result when the ratio $[X^{\ddagger}]/[S]$ is 10^{-5}. If we assume for simplicity's sake that [S] = 1 M, then the reaction rate V is $6.2 \times 10^{7}\ s^{-1}$. If $\Delta G^{\ddagger}$ were lowered by 5.69 kJ mol^{-1} (1.36 kcal mol^{-1}), the ratio $[X^{\ddagger}]/[S]$ would then be 10^{-4}, and the reaction rate would be $6.2 \times 10^{8}\ s^{-1}$. A decrease of 5.69 kJ mol^{-1} in $\Delta G^{\ddagger}$ yields a 10-fold larger V. A relatively small decrease in $\Delta G^{\ddagger}$ (20% in this particular reaction) results in a much greater increase in V.

Thus, we see the key to how enzymes operate: *enzymes accelerate reactions by decreasing $\Delta G^{\ddagger}$, the activation energy*. The combination of substrate and enzyme creates a reaction pathway whose transition-state energy is lower than that of the reaction in the absence of enzyme (see Figure 8.3). Because the activation energy is lower, more molecules have the energy required to reach the transition state. Decreasing the activation barrier is analogous to lowering the height of a high-jump bar; more athletes will be able to clear the bar. *The essence of catalysis is stabilization of the transition state*.

> "I think that enzymes are molecules that are complementary in structure to the activated complexes of the reactions that they catalyze, that is, to the molecular configuration that is intermediate between the reacting substances and the products of reaction for these catalyzed processes. The attraction of the enzyme molecule for the activated complex would thus lead to a decrease in its energy and hence to a decrease in the energy of activation of the reaction and to an increase in the rate of reaction."
>
> —Linus Pauling
> *Nature* 161:707, 1948

The formation of an enzyme–substrate complex is the first step in enzymatic catalysis

Much of the catalytic power of enzymes comes from their bringing substrates together in favorable orientations to promote the formation of the transition states. Enzymes bring together substrates in *enzyme–substrate* (ES) complexes. The substrates are bound to a specific region of the enzyme called the *active site*. Most enzymes are highly selective in the substrates that they bind. Indeed, the catalytic specificity of enzymes depends in part on the specificity of binding.

What is the evidence for the existence of an enzyme–substrate complex?

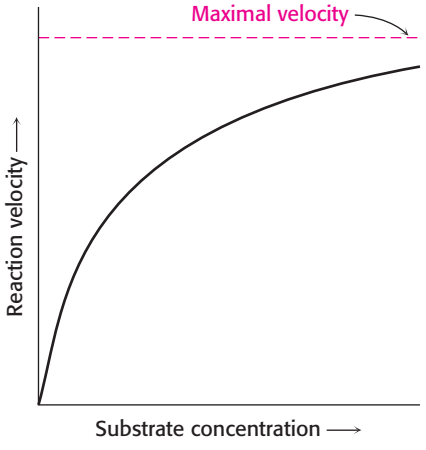


Figure 8.4 Reaction velocity versus substrate concentration in an enzyme-catalyzed reaction. An enzyme-catalyzed reaction approaches a maximal velocity.

1. The first clue was the observation that, at a constant concentration of enzyme, the reaction rate increases with increasing substrate concentration until a maximal velocity is reached (Figure 8.4). In contrast, uncatalyzed reactions do not show this saturation effect. *The fact that an enzyme-catalyzed reaction has a maximal velocity suggests the formation of a discrete ES complex*. At a sufficiently high substrate concentration, all the catalytic sites are filled, or saturated, and so the reaction rate cannot increase. Although

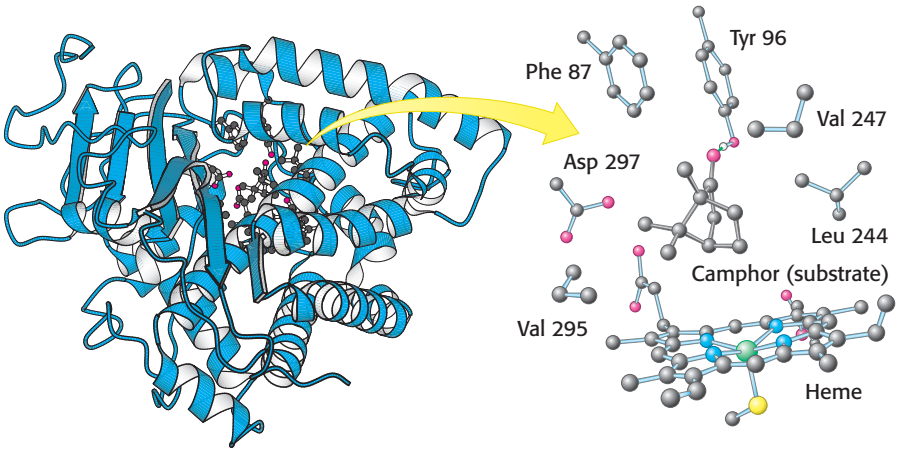


Figure 8.5 Structure of an enzyme–substrate complex. (Left) The enzyme cytochrome P450 is illustrated bound to its substrate camphor. (Right) *Notice* that, in the active site, the substrate is surrounded by residues from the enzyme. Note also the presence of a heme cofactor. [Drawn from 2CPP.pdb.]

indirect, the ability to saturate an enzyme with substrate is the most general evidence for the existence of ES complexes.

2. *X-ray crystallography* has provided high-resolution images of substrates and substrate analogs bound to the active sites of many enzymes (Figure 8.5). In Chapter 9, we will take a close look at several of these complexes.

3. The *spectroscopic characteristics* of many enzymes and substrates change on the formation of an ES complex. These changes are particularly striking if the enzyme contains a colored prosthetic group (see Problem 31).

The active sites of enzymes have some common features

The *active site* of an enzyme is the region that binds the substrates (and the cofactor, if any). It also contains the residues that directly participate in the making and breaking of bonds. These residues are called the *catalytic groups*. In essence, *the interaction of the enzyme and substrate at the active site promotes the formation of the transition state*. The active site is the region of the enzyme that most directly lowers the $\Delta G^{\ddagger}$ of the reaction, thus providing the rate-enhancement characteristic of enzyme action. Although enzymes differ widely in structure, specificity, and mode of catalysis, a number of generalizations concerning their active sites can be stated:

1. *The active site is a three-dimensional cleft, or crevice,* formed by groups that come from different parts of the amino acid sequence: indeed, residues far apart in the amino acid sequence may interact more strongly than adjacent residues in the sequence, which may be sterically constrained from interacting with one another. In lysozyme, an enzyme that degrades the cell walls of some bacteria, the important groups in the active site are contributed by residues numbered 35, 52, 62, 63, 101, and 108 in the sequence of 129 amino acids (Figure 8.6).

2. *The active site takes up a small part of the total volume of an enzyme*. Most of the amino acid residues in an enzyme are not in contact with the substrate, which raises the intriguing question of why enzymes are so big. Nearly all enzymes are made up of more than 100 amino acid residues, which gives them a mass greater than 10 kd and a diameter of more than 25 Å. The "extra" amino acids serve as a scaffold to create the three-dimensional active site. In many proteins, the remaining amino acids also

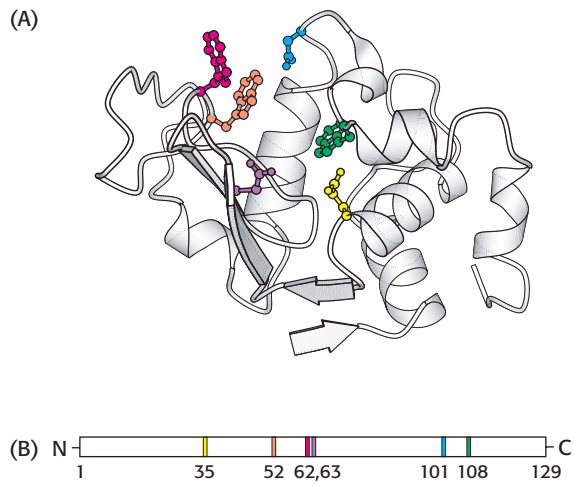


Figure 8.6 Active sites may include distant residues. (A) Ribbon diagram of the enzyme lysozyme with several components of the active site shown in color. (B) A schematic representation of the primary structure of lysozyme shows that the active site is composed of residues that come from different parts of the polypeptide chain. [Drawn from 6LYZ.pdb.]

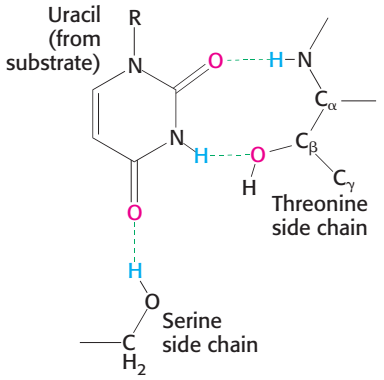


Figure 8.7 Hydrogen bonds between an enzyme and substrate. The enzyme ribonuclease forms hydrogen bonds with the uridine component of the substrate. [After F. M. Richards, H. W. Wyckoff, and N. Allewell. In *The Neurosciences: Second Study Program,* F. O. Schmidt, Ed. (Rockefeller University Press, 1970), p. 970.]

constitute regulatory sites, sites of interaction with other proteins, or channels to bring the substrates to the active sites.

3. *Active sites are unique microenvironments*. In all enzymes of known structure, active sites are shaped like a cleft, or crevice, to which the substrates bind. Water is usually excluded unless it is a reactant. The nonpolar microenvironment of the cleft enhances the binding of substrates as well as catalysis. Nevertheless, the cleft may also contain polar residues. In the nonpolar microenvironment of the active site, certain of these polar residues acquire special properties essential for substrate binding or catalysis. The internal positions of these polar residues are biologically crucial exceptions to the general rule that polar residues are exposed to water.

4. *Substrates are bound to enzymes by multiple weak attractions*. The noncovalent interactions in ES complexes are much weaker than covalent bonds, which have energies between -210 and -460 kJ mol^{-1} (between -50 and -110 kcal mol^{-1}). In contrast, ES complexes usually have equilibrium constants that range from 10^{-2} to 10^{-8} M, corresponding to free energies of interaction ranging from about -13 to -50 kJ mol^{-1} (from -3 to -12 kcal mol^{-1}). As discussed in Section 1.3, these weak reversible interactions are mediated by electrostatic interactions, hydrogen bonds, and van der Waals forces. Van der Waals forces become significant in binding only when numerous substrate atoms simultaneously come close to many enzyme atoms through the hydrophobic effect. Hence, the enzyme and substrate should have complementary shapes. The directional character of hydrogen bonds between enzyme and substrate often enforces a high degree of specificity, as seen in the RNA-degrading enzyme ribonuclease (Figure 8.7).

5. *The specificity of binding depends on the precisely defined arrangement of atoms in an active site*. Because the enzyme and the substrate interact by means of short-range forces that require close contact, a substrate must have a matching shape to fit into the site. Emil Fischer proposed the lock-and-key analogy in 1890 (Figure 8.8), which was the model for enzyme–substrate interaction for several decades. However, we now know that enzymes are flexible and that the shapes of the active sites can be markedly modified by the binding of substrate, as was postulated by Daniel E. Koshland, Jr., in 1958. The active site of some enzymes assumes a shape that is complementary to that of the substrate only *after* the substrate has been bound. This process of dynamic recognition is called *induced fit* (Figure 8.9).

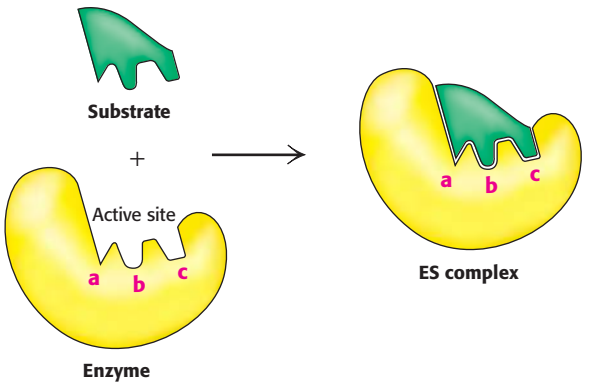


Figure 8.8 Lock-and-key model of enzyme–substrate binding. In this model, the active site of the unbound enzyme is complementary in shape to the substrate.

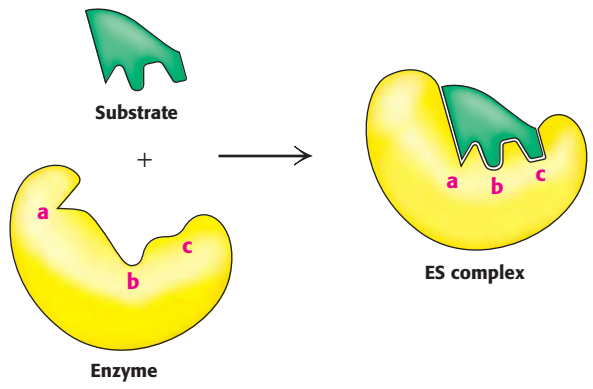


Figure 8.9 Induced-fit model of enzyme–substrate binding. In this model, the enzyme changes shape on substrate binding. The active site forms a shape complementary to the substrate only after the substrate has been bound.

The binding energy between enzyme and substrate is important for catalysis

Enzymes lower the activation energy, but where does the energy to lower the activation energy come from? Free energy is released by the formation of a large number of weak interactions between a complementary enzyme and its substrate. The free energy released on binding is called the *binding energy*. Only the correct substrate can participate in most or all of the interactions with the enzyme and thus maximize binding energy, accounting for the exquisite substrate specificity exhibited by many enzymes. Furthermore, *the full complement of such interactions is formed only when the substrate is converted into the transition state*. Thus, the maximal binding energy is released when the enzyme facilitates the formation of the transition state. The energy released by the interactions between the enzyme and the substrate can be thought of as lowering the activation energy. Paradoxically, the most-stable interaction (maximum binding energy) takes place between the enzyme and the transition state, the least-stable reaction intermediate. However, the transition state is too unstable to exist for long. It collapses to either substrate or product, but which of the two accumulates is determined only by the energy difference between the substrate and the product—that is, by the ΔG of the reaction.

8.4 The Michaelis–Menten Equation Describes the Kinetic Properties of Many Enzymes

The study of the rates of chemical reactions is called *kinetics,* and the study of the rates of enzyme-catalyzed reactions is called *enzyme kinetics*. A kinetic description of enzyme activity will help us understand how enzymes function. We begin by briefly examining some of the basic principles of reaction kinetics.

Kinetics is the study of reaction rates

What do we mean when we say the "rate" of a chemical reaction? Consider a simple reaction:

$$\mathrm{A} \longrightarrow \mathrm{P}$$

The rate V is the quantity of A that disappears in a specified unit of time. It is equal to the rate of the appearance of P, or the quantity of P that appears in a specified unit of time.

$$V = -\Delta \mathrm{A}/\Delta T = \Delta \mathrm{P}/\Delta T \tag{8}$$

If A is yellow and P is colorless, we can follow the decrease in the concentration of A by measuring the decrease in the intensity of yellow color with time. Consider only the change in the concentration of A for now. The rate of the reaction is directly related to the concentration of A by a proportionality constant, k, called the *rate constant*.

$$V = k[\mathrm{A}] \tag{9}$$

Reactions that are directly proportional to the reactant concentration are called *first-order reactions*. First-order rate constants have the units of s^{-1}.

Many important biochemical reactions include two reactants. For example,

$$2\,\mathrm{A} \longrightarrow \mathrm{P}$$

or

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{P}$$

They are called *bimolecular reactions* and the corresponding rate equations often take the form

$$V = k[\mathrm{A}]^2 \tag{10}$$

and

$$V = k[\mathrm{A}][\mathrm{B}] \tag{11}$$

The rate constants, called second-order rate constants, have the units $\mathrm{M}^{-1}\ \mathrm{s}^{-1}$.

Sometimes, second-order reactions can appear to be first-order reactions. For instance, in reaction 11, if B is present in excess and A is present at low concentrations, the reaction rate will be first order with respect to A and will not appear to depend on the concentration of B. These reactions are called *pseudo-first-order reactions,* and we will see them a number of times in our study of biochemistry.

Interestingly enough, under some conditions, a reaction can be zero order. In these cases, the rate is independent of reactant concentrations. Enzyme-catalyzed reactions can approximate zero-order reactions under some circumstances (p. 232).

The steady-state assumption facilitates a description of enzyme kinetics

The simplest way to investigate the reaction rate is to follow the increase in reaction product as a function of time. The extent of product formation is determined as a function of time for a series of substrate concentrations (Figure 8.10A). As expected, in each case, the amount of product formed increases with time, although eventually a time is reached when there is *no net change* in the concentration of S or P. The enzyme is still actively converting substrate into product and visa versa, but the reaction equilibrium has been attained. However, enzyme kinetics is more readily comprehended if we consider only the forward reaction. We can define the rate of catalysis V_0 as the number of moles of product formed per second when the reaction is just beginning—that is, when $t \approx 0$ (see Figure 8.10A). On the time scale of enzyme-catalyzed reactions, the amount of enzyme present is constant. When we plot V_0 versus the substrate concentration [S], assuming a constant amount of enzyme, many enzymes yield the results shown in Figure 8.10B. The rate of catalysis rises linearly as substrate concentration increases and then begins to level off and approach a maximum at higher substrate concentrations.

In 1913, Leonor Michaelis and Maud Menten proposed a simple model to account for these kinetic characteristics. The critical feature in their treatment is that a specific ES complex is a necessary intermediate in catalysis. The model proposed is

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{E} + \mathrm{P}$$

An enzyme E combines with substrate S to form an ES complex, with a rate constant k_1. The ES complex has two possible fates. It can dissociate to E and S, with a rate constant k_{-1}, or it can proceed to form product P, with a rate constant k_2. The ES complex can also be re-formed from E and P by the reverse reaction with a rate constant k_{-2}. However, as before, we can simplify these reactions by considering the rate of reaction at times close to zero (hence, V_0) when there is negligible product formation and thus no back reaction ($k_{-2}\,[\mathrm{E}][\mathrm{P}] \approx 0$).

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P} \tag{12}$$

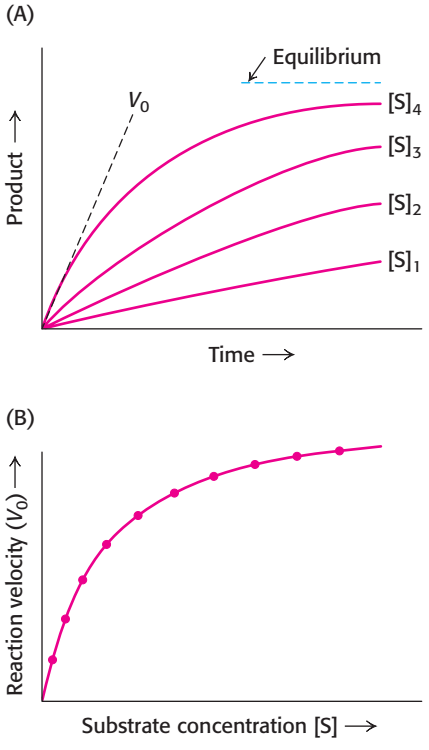


Figure 8.10 Determining the relation between initial velocity and substrate concentration. (A) The amount of product formed at different substrate concentrations is plotted as a function of time. The initial velocity (V_0) for each substrate concentration is determined from the slope of the curve at the beginning of a reaction, when the reverse reaction is insignificant. (B) The values for initial velocity determined in part A are then plotted against substrate concentration.

Thus, for the graph in Figure 8.11, V_0 is determined for each substrate concentration by measuring the rate of product formation at early times before P accumulates (see Figure 8.10A).

We want an expression that relates the rate of catalysis to the concentrations of substrate and enzyme and the rates of the individual steps. Our starting point is that the catalytic rate is equal to the product of the concentration of the ES complex and k_2.

$$V_0 = k_2[\mathrm{ES}] \tag{13}$$

Now we need to express [ES] in terms of known quantities. The rates of formation and breakdown of ES are given by

$$\text{Rate of formation of ES} = k_1[\mathrm{E}][\mathrm{S}] \tag{14}$$

$$\text{Rate of formation of ES} = (k_{-1} + k_2)[\mathrm{ES}] \tag{15}$$

To simplify matters, George Briggs and John Haldane suggested the *steady-state assumption* in 1924. In a steady state, the concentrations of intermediates—in this case, [ES]—stay the same even if the concentrations of starting materials and products are changing. This steady state is reached when the rates of formation and breakdown of the ES complex are equal. Setting the right-hand sides of equations 14 and 15 equal gives

$$k_1[\mathrm{E}][\mathrm{S}] = (k_{-1} + k_2)[\mathrm{ES}] \tag{16}$$

By rearranging equation 16, we obtain

$$[\mathrm{E}][\mathrm{S}]/[\mathrm{ES}] = (k_{-1} + k_2)/k_1 \tag{17}$$

Equation 17 can be simplified by defining a new constant, K_M, called the *Michaelis constant:*

$$K_\mathrm{M} = \frac{k_{-1} + k_2}{k_1} \tag{18}$$

Note that K_M has the units of concentration and is independent of enzyme and substrate concentrations. As will be explained, K_M is an important characteristic of enzyme–substrate interactions.

Inserting equation 18 into equation 17 and solving for [ES] yields

$$[\mathrm{ES}] = \frac{[\mathrm{E}][\mathrm{S}]}{K_\mathrm{M}} \tag{19}$$

Now let us examine the numerator of equation 19. Because the substrate is usually present at a much higher concentration than that of the enzyme, the concentration of uncombined substrate [S] is very nearly equal to the total substrate concentration. The concentration of uncombined enzyme [E] is equal to the total enzyme concentration $[\mathrm{E}]_\mathrm{T}$ minus the concentration of the ES complex:

$$[\mathrm{E}] = [\mathrm{E}]_\mathrm{T} - [\mathrm{ES}] \tag{20}$$

Substituting this expression for [E] in equation 19 gives

$$[\mathrm{ES}] = \frac{([\mathrm{E}]_\mathrm{T} - [\mathrm{ES}])[\mathrm{S}]}{K_\mathrm{M}} \tag{21}$$

Solving equation 21 for [ES] gives

$$[\mathrm{ES}] = \frac{[\mathrm{E}]_\mathrm{T}[\mathrm{S}]/K_\mathrm{M}}{1 + [\mathrm{S}]/K_\mathrm{M}} \tag{22}$$

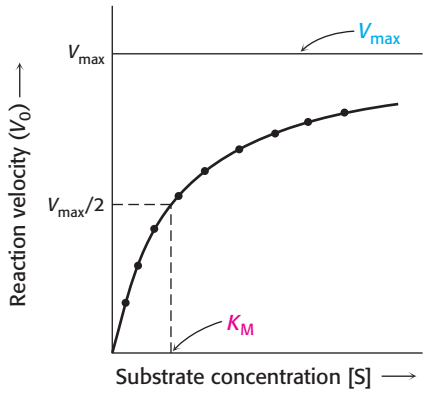


Figure 8.11 Michaelis–Menten kinetics. A plot of the reaction velocity (V_0) as a function of the substrate concentration [S] for an enzyme that obeys Michaelis–Menten kinetics shows that the maximal velocity (V_{max}) is approached asymptotically. The Michaelis constant (K_M) is the substrate concentration yielding a velocity of $V_{max}/2$.

or

$$[ES] = [E]_T \frac{[S]}{[S] + K_M} \tag{23}$$

By substituting this expression for [ES] into equation 13, we obtain

$$V_0 = k_2[E]_T \frac{[S]}{[S] + K_M} \tag{24}$$

The *maximal rate,* V_{max}, is attained when the catalytic sites on the enzyme are saturated with substrate—that is, when $[ES] = [E]_T$. Thus,

$$V_{max} = k_2[E]_T \tag{25}$$

Substituting equation 25 into equation 24 yields the *Michaelis–Menten equation:*

$$V_0 = V_{max} \frac{[S]}{[S] + K_M} \tag{26}$$

This equation accounts for the kinetic data given in Figure 8.11. At very low substrate concentration, when [S] is much less than K_M, $V_0 = (V_{max}/K_M)[S]$; that is, the reaction is first order with the rate directly proportional to the substrate concentration. At high substrate concentration, when [S] is much greater than K_M, $V_0 = V_{max}$; that is, the rate is maximal. The reaction is zero order, independent of substrate concentration.

The meaning of K_M is evident from equation 26. When $[S] = K_M$, then $V_0 = V_{max}/2$. Thus, *K_M is equal to the substrate concentration at which the reaction rate is half its maximal value.* K_M is an important characteristic of an enzyme-catalyzed reaction and is significant for its biological function.

Variations in K_M can have physiological consequences

The physiological consequence of K_M is illustrated by the sensitivity of some persons to ethanol. Such persons exhibit facial flushing and rapid heart rate (tachycardia) after ingesting even small amounts of alcohol. In the liver, alcohol dehydrogenase converts ethanol into acetaldehyde.

$$CH_3CH_2OH + NAD^+ \xrightleftharpoons{\text{Alcohol dehydrogenase}} CH_3CHO + NADH + H^+$$

Normally, the acetaldehyde, which is the cause of the symptoms when present at high concentrations, is processed to acetate by aldehyde dehydrogenase.

$$CH_3CHO + NAD^+ + H_2O \xrightleftharpoons{\text{Aldehyde dehydrogenase}} CH_3COO^- + NADH + 2H^+$$

Most people have two forms of the aldehyde dehydrogenase, a low K_M mitochondrial form and a high K_M cytoplasmic form. In susceptible persons, the mitochondrial enzyme is less active owing to the substitution of a single amino acid, and acetaldehyde is processed only by the cytoplasmic enzyme. Because this enzyme has a high K_M, it achieves a high rate of catalysis only at very high concentrations of acetaldehyde. Consequently, less acetaldehyde is converted into acetate; excess acetaldehyde escapes into the blood and accounts for the physiological effects.

K_M and V_{max} values can be determined by several means

K_M is equal to the substrate concentration that yields $V_{max}/2$; however V_{max}, like perfection, is only approached but never attained. How, then, can

we experimentally determine K_M and V_{max}, and how do these parameters enhance our understanding of enzyme-catalyzed reactions? The Michaelis constant, K_M, and the maximal rate, V_{max}, can be readily derived from rates of catalysis measured at a variety of substrate concentrations if an enzyme operates according to the simple scheme given in equation 26. The derivation of K_M and V_{max} is most commonly achieved with the use of curve-fitting programs on a computer. However, an older method, although rarely used because the data points at high and low concentrations are weighted differently and thus sensitive to errors, is a source of further insight into the meaning of K_M and V_{max}.

Before the availability of computers, the determination of K_M and V_{max} values required algebraic manipulation of the basic Michaelis–Menten equation. Because V_{max} is approached asymptotically (see Figure 8.11), it is impossible to obtain a definitive value from a Michaelis–Menten curve. Because K_M is the concentration of substrate at $V_{max}/2$, it is likewise impossible to determine an accurate value of K_M. However, V_{max} can be accurately determined if the Michaelis–Menten equation is transformed into one that gives a straight-line plot. Taking the reciprocal of both sides of equation 26 gives

$$\frac{1}{V_0} = \frac{K_M}{V_{max}} \cdot \frac{1}{S} + \frac{1}{V_{max}} \qquad (27)$$

A plot of $1/V_0$ versus $1/[S]$, called a *Lineweaver–Burk* or *double-reciprocal plot,* yields a straight line with a y-intercept of $1/V_{max}$ and a slope of K_M/V_{max} (Figure 8.12). The intercept on the x-axis is $-1/K_M$.

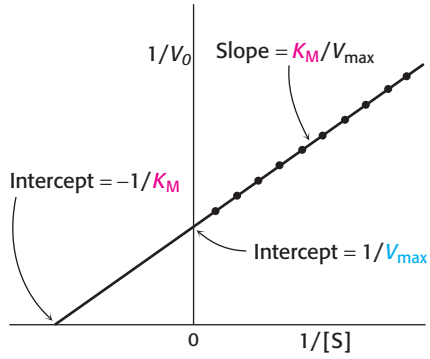


Figure 8.12 A double-reciprocal or Lineweaver–Burk plot. A double-reciprocal plot of enzyme kinetics is generated by plotting $1/V_0$ as a function $1/[S]$. The slope is K_M/V_{max}, the intercept on the vertical axis is $1/V_{max}$, and the intercept on the horizontal axis is $-1/K_M$.

K_M and V_{max} values are important enzyme characteristics

The K_M values of enzymes range widely (Table 8.4). For most enzymes, K_M lies between 10^{-1} and 10^{-7} M. The K_M value for an enzyme depends on the particular substrate and on environmental conditions such as pH, temperature, and ionic strength. The Michaelis constant, K_M, has two meanings. First, K_M is the concentration of substrate at which half the active sites are filled. Thus, K_M provides a measure of the substrate concentration required for significant catalysis to take place. For many enzymes, experimental evidence suggests that K_M provides an approximation of substrate concentration in vivo.

Second, K_M is related to the rate constants of the individual steps in the catalytic scheme given in equation 12. In equation 18, K_M is defined as $(k_{-1} + k_2)/k_1$. Consider a case in which k_{-1} is much greater than k_2. Under such circumstances, the ES complex dissociates to E and S much more rapidly than product is formed. Under these conditions ($k_{-1} \gg k_2$),

$$K_M \approx \frac{k_{-1}}{k_1} \qquad (28)$$

Equation 28 describes the dissociation constant of the ES complex.

$$K_{ES} = \frac{[E][S]}{[ES]} = \frac{k_{-1}}{k_1} \qquad (29)$$

In other words, *K_M is equal to the dissociation constant of the ES complex if k_2 is much smaller than k_{-1}*. When this condition is met, K_M is a measure of the strength of the ES complex: a high

Table 8.4 K_M values of some enzymes

Enzyme	Substrate	K_M (μM)
Chymotrypsin	Acetyl-L-tryptophanamide	5000
Lysozyme	Hexa-*N*-acetylglucosamine	6
β-Galactosidase	Lactose	4000
Threonine deaminase	Threonine	5000
Carbonic anhydrase	CO_2	8000
Penicillinase	Benzylpenicillin	50
Pyruvate carboxylase	Pyruvate	400
	HCO_3^-	1000
	ATP	60
Arginine-tRNA synthetase	Arginine	3
	tRNA	0.4
	ATP	300

K_M indicates weak binding; a low K_M indicates strong binding. It must be stressed that K_M indicates the affinity of the ES complex only when k_{-1} is much greater than k_2.

The maximal rate, V_{max}, reveals the *turnover number* of an enzyme, which is *the number of substrate molecules converted into product by an enzyme molecule in a unit time when the enzyme is fully saturated with substrate.* It is equal to the rate constant k_2, which is also called k_{cat}. The maximal rate, V_{max}, reveals the turnover number of an enzyme if the concentration of active sites $[E]_T$ is known, because

$$V_{max} = k_2[E]_T \tag{30}$$

and thus

$$k_2 = V_{max}/[E]_T \tag{31}$$

Table 8.5 Turnover numbers of some enzymes

Enzyme	Turnover number (per second)
Carbonic anhydrase	600,000
3-Ketosteroid isomerase	280,000
Acetylcholinesterase	25,000
Penicillinase	2,000
Lactate dehydrogenase	1,000
Chymotrypsin	100
DNA polymerase I	15
Tryptophan synthetase	2
Lysozyme	0.5

For example, a 10^{-6} M solution of carbonic anhydrase catalyzes the formation of 0.6 M H_2CO_3 per second when the enzyme is fully saturated with substrate. Hence, k_2 is $6 \times 10^5\ s^{-1}$. This turnover number is one of the largest known. Each catalyzed reaction takes place in a time equal to, on average, $1/k_2$, which is 1.7 μs for carbonic anhydrase. The turnover numbers of most enzymes with their physiological substrates range from 1 to 10^4 per second (Table 8.5).

K_M and V_{max} also permit the determination of f_{ES}, the fraction of active sites filled. This relation of f_{ES} to K_M and V_{max} is given by the following equation:

$$f_{ES} = \frac{V}{V_{max}} = \frac{[S]}{[S] + K_M} \tag{32}$$

k_{cat}/K_M is a measure of catalytic efficiency

When the substrate concentration is much greater than K_M, the rate of catalysis is equal to V_{max}, which is a function of k_{cat}, the turnover number, as already described. However, most enzymes are not normally saturated with substrate. Under physiological conditions, the $[S]/K_M$ ratio is typically between 0.01 and 1.0. When $[S] \ll K_M$, the enzymatic rate is much less than k_{cat} because most of the active sites are unoccupied. Is there a number that characterizes the kinetics of an enzyme under these more typical cellular conditions? Indeed there is, as can be shown by combining equations 13 and 19 to give

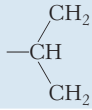
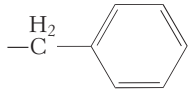
$$V_0 = \frac{k_{cat}}{K_M}[E][S] \tag{33}$$

When $[S] \ll K_M$, the concentration of free enzyme [E], is nearly equal to the total concentration of enzyme $[E]_T$; so

$$V_0 = \frac{k_{cat}}{K_M}[S][E]_T \tag{34}$$

Thus, when $[S] \ll K_M$, the enzymatic velocity depends on the values of k_{cat}/K_M, [S], and $[E]_T$. Under these conditions, k_{cat}/K_M is the rate constant for the interaction of S and E. The rate constant k_{cat}/K_M is a measure of catalytic efficiency because it takes into account both the rate of catalysis with a particular substrate (k_{cat}) and the strength of the enzyme–substrate interaction (K_M). For instance, by using k_{cat}/K_M values, we can compare an enzyme's preference for different substrates. Table 8.6 shows the k_{cat}/K_M values for several different substrates of chymotrypsin.

Table 8.6 Substrate preferences of chymotrypsin

Amino acid in ester	Amino acid side chain	k_{cat}/K_M ($s^{-1}\ M^{-1}$)
Glycine	—H	1.3×10^{-1}
Valine		2.0
Norvaline	$—CH_2CH_2CH_3$	3.6×10^{2}
Norleucine	$—CH_2CH_2CH_2CH_3$	3.0×10^{3}
Phenylalanine		1.0×10^{5}

Source: After A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman and Company, 1999), Table 7.3.

Chymotrypsin clearly has a preference for cleaving next to bulky, hydrophobic side chains.

How efficient can an enzyme be? We can approach this question by determining whether there are any physical limits on the value of k_{cat}/K_M. Note that the $k_{cat}\ K_M$ *ratio* depends on k_1, k_{-1}, and k_{cat}, as can be shown by substituting for K_M.

$$k_{cat}/K_M = \frac{k_{cat}k_1}{k_{-1} + k_{cat}} = \left(\frac{k_{cat}}{k_{-1} + k_{cat}}\right)k_1 < k_1 \qquad (35)$$

Suppose that the rate of formation of product (k_{cat}) is much faster than the rate of dissociation of the ES complex (k_{-1}). The value of k_{cat}/K_M then approaches k_1. Thus, the ultimate limit on the value of k_{cat}/K_M is set by k_1, the rate of formation of the ES complex. *This rate cannot be faster than the diffusion-controlled encounter of an enzyme and its substrate.* Diffusion limits the value of k_1 and so it cannot be higher than between 10^8 and $10^9\ s^{-1}\ M^{-1}$. Hence, the upper limit on k_{cat}/K_M is between 10^8 and $10^9\ s^{-1}\ M^{-1}$.

The k_{cat}/K_M ratios of the enzymes superoxide dismutase, acetylcholinesterase, and triose phosphate isomerase are between 10^8 and $10^9\ s^{-1}\ M^{-1}$. Enzymes that have k_{cat}/K_M ratios at the upper limits have attained *kinetic perfection. Their catalytic velocity is restricted only by the rate at which they encounter substrate in the solution* (Table 8.7). Any further gain in catalytic rate can come only by decreasing the time for diffusion of the substrate into the enzyme's immediate environment. Remember that the active site is only a small part of the total enzyme structure. Yet, for catalytically perfect enzymes, every encounter between enzyme and substrate is productive. In these cases, there may be attractive electrostatic forces on the enzyme that entice the substrate to the active site. These forces are sometimes referred to poetically as *Circe effects*.

The diffusion of a substrate throughout a solution can also be partly overcome by confining substrates and products in the limited volume of a multienzyme complex. Indeed, some series of enzymes are organized into complexes so that the product of one enzyme is very rapidly found by the next enzyme. In effect, products are channeled from one enzyme to the next, much as in an assembly line.

Table 8.7 Enzymes for which k_{cat}/K_M is close to the diffusion-controlled rate of encounter

Enzyme	k_{cat}/K_M ($s^{-1}\ M^{-1}$)
Acetylcholinesterase	1.6×10^{8}
Carbonic anhydrase	8.3×10^{7}
Catalase	4×10^{7}
Crotonase	2.8×10^{8}
Fumarase	1.6×10^{8}
Triose phosphate isomerase	2.4×10^{8}
β-Lactamase	1×10^{8}
Superoxide dismutase	7×10^{9}

Source: After A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman and Company, 1999), Table 4.5.

Circe effect

The utilization of attractive forces to lure a substrate into a site in which it undergoes a transformation of structure, as defined by William P. Jencks, an enzymologist, who coined the term.

A goddess of Greek mythology, Circe lured Odysseus's men to her house and then transformed them into pigs.

Most biochemical reactions include multiple substrates

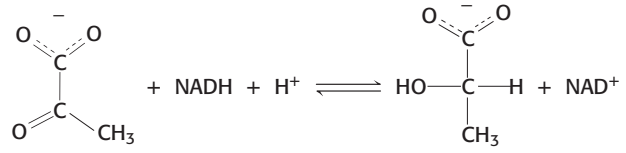
Most reactions in biological systems start with two substrates and yield two products. They can be represented by the bisubstrate reaction:

$$A + B \rightleftharpoons P + Q$$

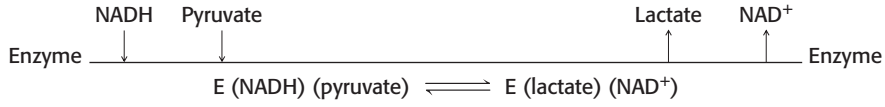
Many such reactions transfer a functional group, such as a phosphoryl or an ammonium group, from one substrate to the other. Those that are oxidation–reduction reactions transfer electrons between substrates. Multiple substrate reactions can be divided into two classes: *sequential* reactions and *double-displacement* reactions.

Sequential reactions. In *sequential reactions,* all substrates must bind to the enzyme before any product is released. Consequently, in a bisubstrate reaction, a *ternary complex* of the enzyme and both substrates forms. Sequential mechanisms are of two types: ordered, in which the substrates bind the enzyme in a defined sequence, and random.

Many enzymes that have NAD^+ or NADH as a substrate exhibit the ordered sequential mechanism. Consider lactate dehydrogenase, an important enzyme in glucose metabolism (Section 16.1). This enzyme reduces pyruvate to lactate while oxidizing NADH to NAD^+.

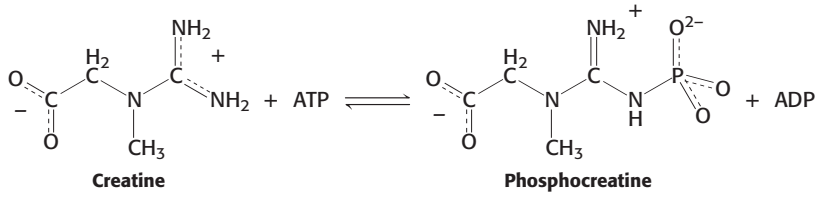


In the ordered sequential mechanism, the coenzyme always binds first and the lactate is always released first. This sequence can be represented by using a notation developed by W. Wallace Cleland:

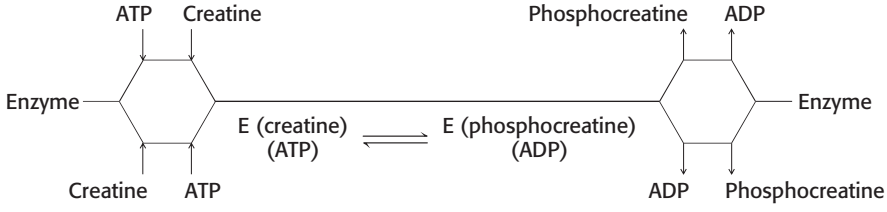


The enzyme exists as a ternary complex consisting of, first, the enzyme and substrates and, after catalysis, the enzyme and products.

In the random sequential mechanism, the order of the addition of substrates and the release of products is random. An example of a random sequential reaction is the formation of phosphocreatine and ADP from ATP and creatine, which is catalyzed by creatine kinase (p. 8).

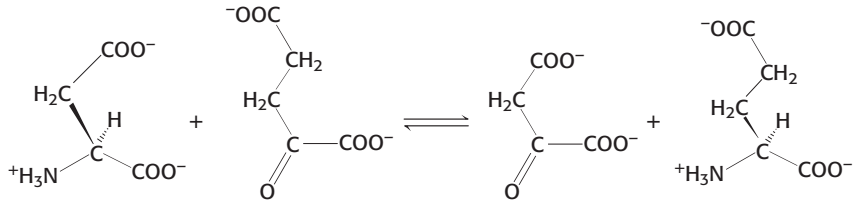


Either creatine or ATP may bind first, and either phosphocreatine or ADP may be released first. Phosphocreatine is an important energy source in muscle. Sequential random reactions also can be depicted in the Cleland notation.

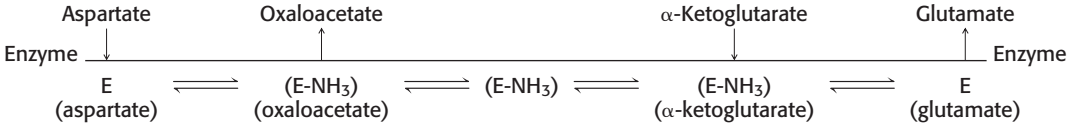


Although the order of certain events is random, the reaction still passes through the ternary complexes including, first, substrates and, then, products.

Double-displacement (ping-pong) reactions. In *double-displacement,* or *ping-pong, reactions,* one or more products are released before all substrates bind the enzyme. The defining feature of double-displacement reactions is the existence of a *substituted enzyme intermediate,* in which the enzyme is temporarily modified. Reactions that shuttle amino groups between amino acids and α-ketoacids are classic examples of double-displacement mechanisms. The enzyme aspartate aminotransferase catalyzes the transfer of an amino group from aspartate to α-ketoglutarate.



The sequence of events can be portrayed as the following Cleland notation:



After aspartate binds to the enzyme, the enzyme accepts aspartate's amino group to form the substituted enzyme intermediate. The first product, oxaloacetate, subsequently departs. The second substrate, α-ketoglutarate, binds to the enzyme, accepts the amino group from the modified enzyme, and is then released as the final product, glutamate. In the Cleland notation, the substrates appear to bounce on and off the enzyme much as a Ping-Pong ball bounces on a table.

Allosteric enzymes do not obey Michaelis–Menten kinetics

The Michaelis–Menten model has greatly assisted the development of enzymology. Its virtues are simplicity and broad applicability. However, the Michaelis–Menten model cannot account for the kinetic properties of many enzymes. An important group of enzymes that do not obey Michaelis–Menten kinetics are the *allosteric enzymes*. These enzymes consist of multiple subunits and multiple active sites.

Allosteric enzymes often display sigmoidal plots (Figure 8.13) of the reaction velocity V_0 versus substrate concentration [S], rather than the hyperbolic plots predicted by the Michaelis–Menten equation (equation 26). In allosteric enzymes, the binding of substrate to one active site can alter the properties of other active sites in the same enzyme molecule. A possible outcome of this interaction between subunits is that the binding of substrate becomes *cooperative;* that is, the binding of substrate to one active site facilitates the binding of substrate to the other active sites. As it does for hemoglobin (Chapter 7), such cooperativity results in a sigmoidal plot of V_0 versus [S]. In addition, the activity of an allosteric enzyme may be altered by regulatory molecules that are reversibly bound to specific sites other than the catalytic sites. The catalytic properties of allosteric enzymes can thus be adjusted to meet the immediate needs of a cell (Chapter 10). For this reason, allosteric enzymes are key regulators of metabolic pathways.

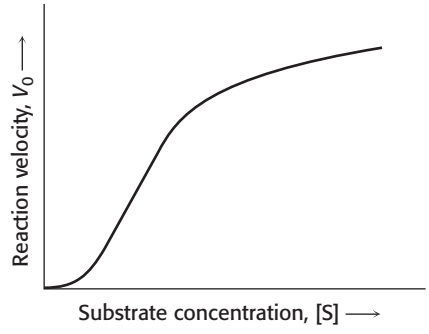


Figure 8.13 Kinetics for an allosteric enzyme. Allosteric enzymes display a sigmoidal dependence of reaction velocity on substrate concentration.

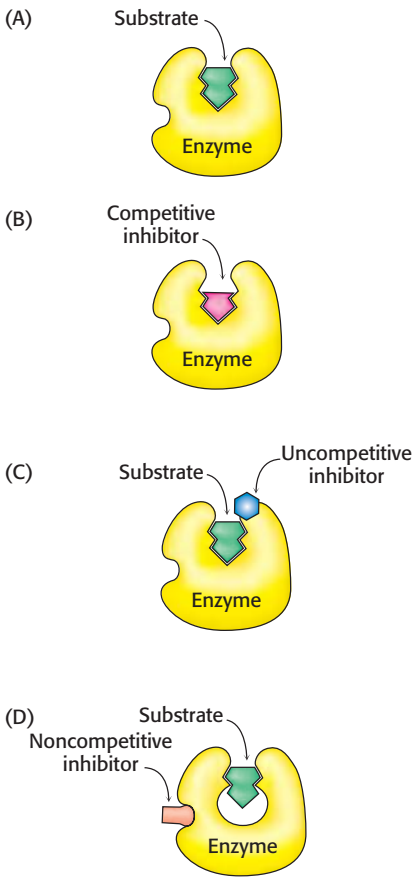
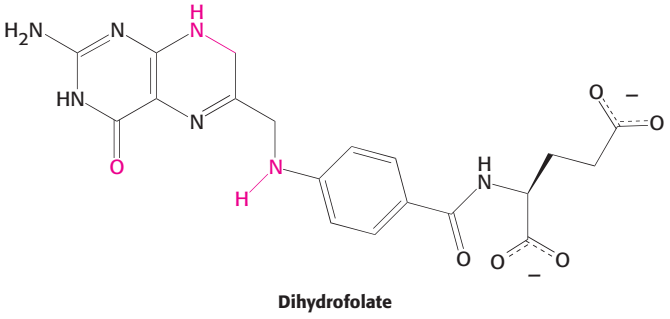


Figure 8.14 Distinction between reversible inhibitors. (A) Enzyme–substrate complex; (B) a competitive inhibitor binds at the active site and thus prevents the substrate from binding; (C) an uncompetitive inhibitor binds only to the enzyme–substrate complex; (D) a noncompetitive inhibitor does not prevent the substrate from binding.



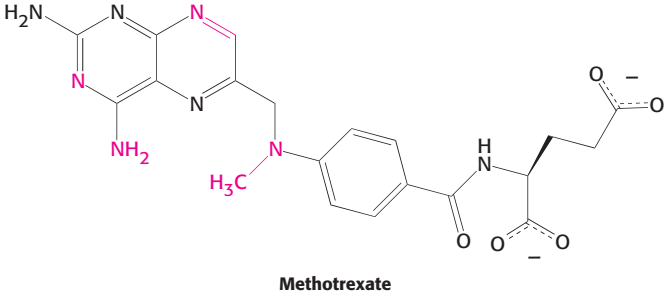


Figure 8.15 Enzyme inhibitors. The substrate dihydrofolate and its structural analog methotrexate. Regions with structural differences are shown in red.

8.5 Enzymes Can Be Inhibited by Specific Molecules

The activity of many enzymes can be inhibited by the binding of specific small molecules and ions. This means of inhibiting enzyme activity serves as a major control mechanism in biological systems, typified by the regulation of allosteric enzymes. In addition, many drugs and toxic agents act by inhibiting enzymes (Chapter 36). Inhibition can be a source of insight into the mechanism of enzyme action: specific inhibitors can often be used to identify residues critical for catalysis. Transition-state analogs are especially potent inhibitors.

Enzyme inhibition can be either irreversible or reversible. An *irreversible inhibitor* dissociates very slowly from its target enzyme because it has become tightly bound to the enzyme, either covalently or noncovalently. Some irreversible inhibitors are important drugs. Penicillin acts by covalently modifying the enzyme transpeptidase, thereby preventing the synthesis of bacterial cell walls and thus killing the bacteria (p. 244). Aspirin acts by covalently modifying the enzyme cyclooxygenase, reducing the synthesis of signaling molecules in inflamation.

Reversible inhibition, in contrast with irreversible inhibition, is characterized by a rapid dissociation of the enzyme–inhibitor complex. In the type of reversible inhibition called *competitive inhibition,* an enzyme can bind substrate (forming an ES complex) or inhibitor (EI) but not both (ESI, enzyme–substrate–inhibitor complex). The competitive inhibitor often resembles the substrate and binds to the active site of the enzyme (Figure 8.14). The substrate is thereby prevented from binding to the same active site. *A competitive inhibitor diminishes the rate of catalysis by reducing the proportion of enzyme molecules bound to a substrate.* At any given inhibitor concentration, competitive inhibition can be relieved by increasing the substrate concentration. Under these conditions, the substrate successfully competes with the inhibitor for the active site. Methotrexate is an especially potent competitive inhibitor of the enzyme dihydrofolate reductase, which plays a role in the biosynthesis of purines and pyrimidines. Methotrexate is a structural analog of dihydrofolate, a substrate for dihydrofolate reductase (Figure 8.15). What makes it such a potent competitive inhibitor is that it binds to the enzyme 1000 times as tightly as the natural substrate binds, and it inhibits nucleotide base synthesis. It is used to treat cancer.

Uncompetitive inhibition is distinguished by the fact that the inhibitor binds only to the enzyme–substrate complex. The uncompetitive inhibitor's binding site is created only on interaction of the enzyme and substrate (see Figure 8.14C). Uncompetitive inhibition cannot be overcome by the addition of more substrate.

In *noncompetitive inhibition,* the inhibitor and substrate can bind simultaneously to an enzyme molecule at different binding sites (see Figure 8.14D). Unlike uncompetitive inhibition, a noncompetitive inhibitor can bind free enzyme or the enzyme–substrate complex. A noncompetitive inhibitor acts by decreasing the concentration of functional enzyme rather than by diminishing the proportion of enzyme molecules that are bound to substrate. The net effect is to decrease the turnover number. Noncompetitive inhibition, like uncompetitive inhibition, cannot be overcome by increasing the substrate concentration. A more complex pattern, called *mixed inhibition,* is produced

when a single inhibitor both hinders the binding of substrate and decreases the turnover number of the enzyme.

Reversible inhibitors are kinetically distinguishable

How can we determine whether a reversible inhibitor acts by competitive, uncompetitive, or noncompetitive inhibition? Let us consider only enzymes that exhibit Michaelis–Menten kinetics. Measurements of the rates of catalysis at different concentrations of substrate and inhibitor serve to distinguish the three types of inhibition. In *competitive inhibition*, the inhibitor competes with the substrate for the active site. The dissociation constant for the inhibitor is given by

$$K_i = [E][I]/[EI]$$

The smaller the K_i, the more potent the inhibition. *The hallmark of competitive inhibition is that it can be overcome by a sufficiently high concentration of substrate* (Figure 8.16). The effect of a competitive inhibitor is to increase the apparent value of K_M, meaning that more substrate is needed to obtain the same reaction rate. This new value of K_M, called K_M^{app}, is numerically equal to

$$K_M^{app} = K_M(1 + [I]/K_i)$$

where [I] is the concentration of inhibitor and K_i is the dissociation constant for the enzyme–inhibitor complex. In the presence of a competitive inhibitor, an enzyme will have the same V_{max} as in the absence of an inhibitor. At a sufficiently high concentration, virtually all the active sites are filled by substrate, and the enzyme is fully operative.

Competitive inhibitors are commonly used as drugs. Drugs such as ibuprofen are competitive inhibitors of enzymes that participate in signaling pathways in the inflammatory response. Statins are drugs that reduce high cholesterol levels by competitively inhibiting a key enzyme in cholesterol biosynthesis.

In *uncompetitive inhibition*, the inhibitor binds only to the ES complex. This enzyme–substrate–inhibitor complex, ESI, does not go on to form any product. Because some unproductive ESI complex will always be present, V_{max} will be lower in the presence of inhibitor than in its absence (Figure 8.17). The uncompetitive inhibitor lowers the apparent value of K_M because the inhibitor binds to ES to form ESI, depleting ES. To maintain the equilibrium between E and ES, more S binds to E. Thus, a lower concentration of S is required to form half of the maximal concentration of ES and the apparent value of K_M is reduced. The herbicide glyphosate, also known as Roundup, is an uncompetitive inhibitor of an enzyme in the biosynthetic pathway for aromatic amino acids.

In *noncompetitive inhibition* (Figure 8.18), substrate can still bind to

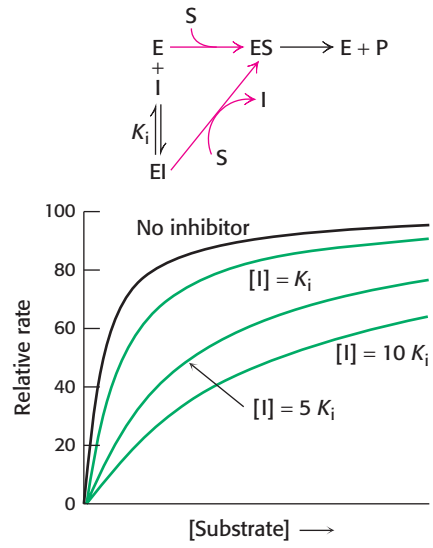


Figure 8.16 Kinetics of a competitive inhibitor. As the concentration of a competitive inhibitor increases, higher concentrations of substrate are required to attain a particular reaction velocity. The reaction pathway suggests how sufficiently high concentrations of substrate can completely relieve competitive inhibition.

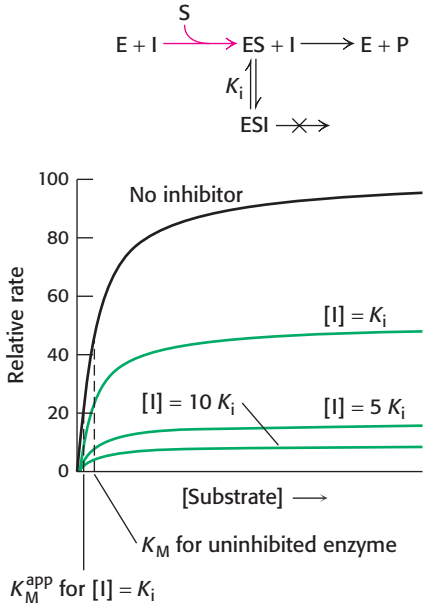


Figure 8.17 Kinetics of an uncompetitive inhibitor. The reaction pathway shows that the inhibitor binds only to the enzyme–substrate complex. Consequently, V_{max} cannot be attained, even at high substrate concentrations. The apparent value for K_M is lowered, becoming smaller as more inhibitor is added.

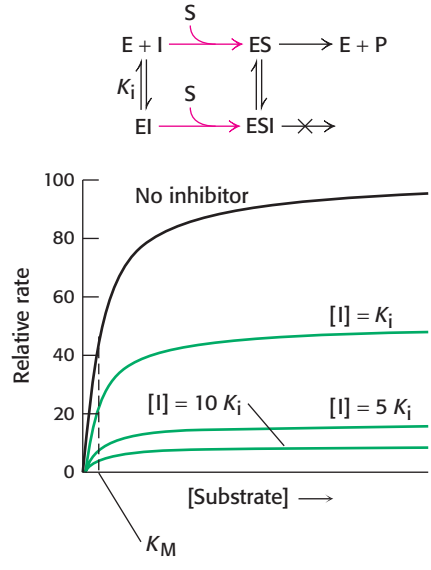


Figure 8.18 Kinetics of a noncompetitive inhibitor. The reaction pathway shows that the inhibitor binds both to free enzyme and to an enzyme–substrate complex. Consequently, as with uncompetitive competition, V_{max} cannot be attained. K_M remains unchanged, and so the reaction rate increases more slowly at low substrate concentrations than is the case for uncompetitive competition.

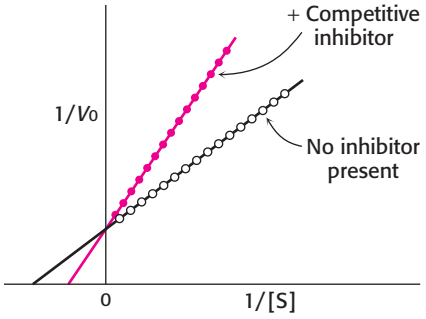


Figure 8.19 Competitive inhibition illustrated on a double-reciprocal plot. A double-reciprocal plot of enzyme kinetics in the presence and absence of a competitive inhibitor illustrates that the inhibitor has no effect on V_{max} but increases K_M.

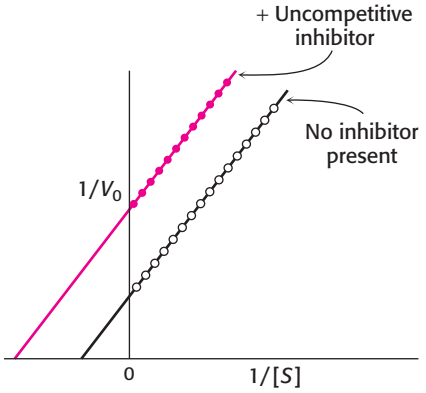


Figure 8.20 Uncompetitive inhibition illustrated by a double-reciprocal plot. An uncompetitive inhibitor does not effect the slope of the double-reciprocal plot. V_{max} and K_M are reduced by equivalent amounts.

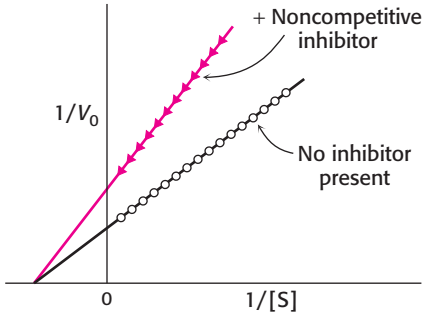


Figure 8.21 Noncompetitive inhibition illustrated on a double-reciprocal plot. A double-reciprocal plot of enzyme kinetics in the presence and absence of a noncompetitive inhibitor shows that K_M is unaltered and V_{max} is decreased.

the enzyme–inhibitor complex. However, the enzyme–inhibitor–substrate complex *does not* proceed to form product. The value of V_{max} is decreased to a new value called V_{max}^{app}, whereas the value of K_M is unchanged. The maximal velocity in the presence of a pure noncompetitive inhibitor, V_{max}^{app}, is given by

$$V_{max}^{app} = \frac{V_{max}}{1 + [I]/K_i} \tag{36}$$

Why is V_{max} lowered though K_M remains unchanged? In essence, the inhibitor simply lowers the concentration of functional enzyme. The resulting solution behaves as a more dilute solution of enzyme does. *Noncompetitive inhibition cannot be overcome by increasing the substrate concentration.* Deoxycycline, an antibiotic, functions at low concentrations as a noncompetitive inhibitor of a proteolytic enzyme (collagenase). It is used to treat periodontal disease. Some of the toxic effects of lead poisoning may be due to lead's ability to act as a noncompetitive inhibitor of a host of enzymes. Lead reacts with crucial sulfhydryl groups in these enzymes.

Double-reciprocal plots are especially useful for distinguishing between competitive, uncompetitive, and noncompetitive inhibitors. In competitive inhibition, the intercept on the y-axis of the plot of $1/V_0$ versus $1/[S]$ is the same in the presence and in the absence of inhibitor, although the slope is increased (Figure 8.19). The intercept is unchanged because a competitive inhibitor does not alter V_{max}. The increase in the slope of the $1/V_0$ versus $1/[S]$ plot indicates the strength of binding of a competitive inhibitor. In the presence of a competitive inhibitor, equation 27 is replaced by

$$\frac{1}{V_0} = \frac{1}{V_{max}} + \frac{K_M}{V_{max}}\left(1 + \frac{[I]}{K_i}\right)\left(\frac{1}{[S]}\right) \tag{37}$$

In other words, the slope of the plot is increased by the factor $(1 + [I]/K_i)$ in the presence of a competitive inhibitor. Consider an enzyme with a K_M of 10^{-4} M. In the absence of inhibitor, $V_0 = V_{max}/2$ when $[S] = 10^{-4}$ M. In the presence of a 2×10^{-3} M competitive inhibitor that is bound to the enzyme with a K_i of 10^{-3} M, the apparent K_M (K_M^{app}) will be equal to K_M $(1 + [I]/K_i)$, or 3×10^{-4} M. Substitution of these values into equation 37 gives $V_0 = V_{max}/4$, when $[S] = 10^{-4}$ M. The presence of the competitive inhibitor thus cuts the reaction rate in half at this substrate concentration.

In uncompetitive inhibition (Figure 8.20), the inhibitor combines only with the enzyme–substrate complex. The equation that describes the double–reciprocal plot for an uncompetitive inhibitor is

$$\frac{1}{V_0} = \frac{K_M}{V_{max}}\frac{1}{[S]} + \frac{1}{V_{max}}\left(1 + \frac{[I]}{K_i}\right) \tag{38}$$

The slope of the line, K_M/V_{max}, is the same as that for the uninhibited enzyme, but the intercept on the y-axis will be increased by $1 + [I]/K_i$. Consequently, the lines in double-reciprocal plots will be parallel.

In noncompetitive inhibition (Figure 8.21), the inhibitor can combine with either the enzyme or the enzyme–substrate complex. In pure noncompetitive inhibition, the values of the dissociation constants of the inhibitor and enzyme and of the inhibitor and enzyme–substrate complex are equal. The value of V_{max} is decreased to the new value V_{max}^{app}, and so the intercept on the vertical axis is increased. The new slope, which is equal to K_M/V_{max}^{app}, is larger by the same factor. In contrast with V_{max}, K_M is not affected by pure noncompetitive inhibition.

Irreversible inhibitors can be used to map the active site

In Chapter 9, we will examine the chemical details of how enzymes function. The first step in obtaining the chemical mechanism of an enzyme is to determine what functional groups are required for enzyme activity. How can we ascertain what these functional groups are? X-ray crystallography of the enzyme bound to its substrate or substrate analog provides one approach. Irreversible inhibitors that covalently bond to the enzyme provide an alternative and often complementary approach: the inhibitors modify the functional groups, which can then be identified. Irreversible inhibitors can be divided into three categories: group-specific reagents, reactive substrate analogs (also called affinity labels), and suicide inhibitors.

Group-specific reagents react with specific side chains of amino acids. An example of a group-specific reagent is diisopropylphosphofluoridate (DIPF). DIPF modifies only 1 of the 28 serine residues in the proteolytic enzyme chymotrypsin, implying that this serine residue is especially reactive. We will see in Chapter 9 that this serine residue is indeed located at the active site. DIPF also revealed a reactive serine residue in acetylcholinesterase, an enzyme important in the transmission of nerve impulses (Figure 8.22). Thus, DIPF and similar compounds that bind and inactivate acetylcholinesterase are potent nerve gases. Many group-specific reagents do not display the exquisite specificity shown by DIPF. Consequently, more specific means of modifying the active site are required.

Affinity labels, or *reactive substrate analogs*, are molecules that are structurally similar to the substrate for an enzyme and that covalently bind to active-site residues. They are thus more specific for the enzyme's active site than are group-specific reagents. Tosyl-L-phenylalanine chloromethyl ketone (TPCK) is a substrate analog for chymotrypsin (Figure 8.23). TPCK

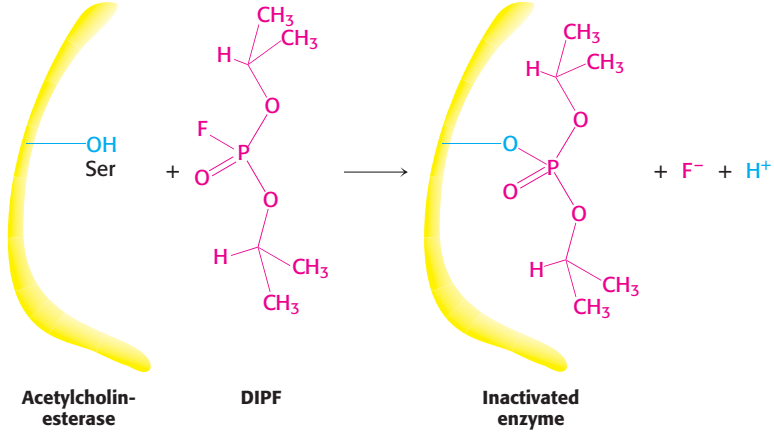


Figure 8.22 Enzyme inhibition by diisopropylphosphofluoridate (DIPF), a group-specific reagent. DIPF can inhibit an enzyme by covalently modifying a crucial serine residue.

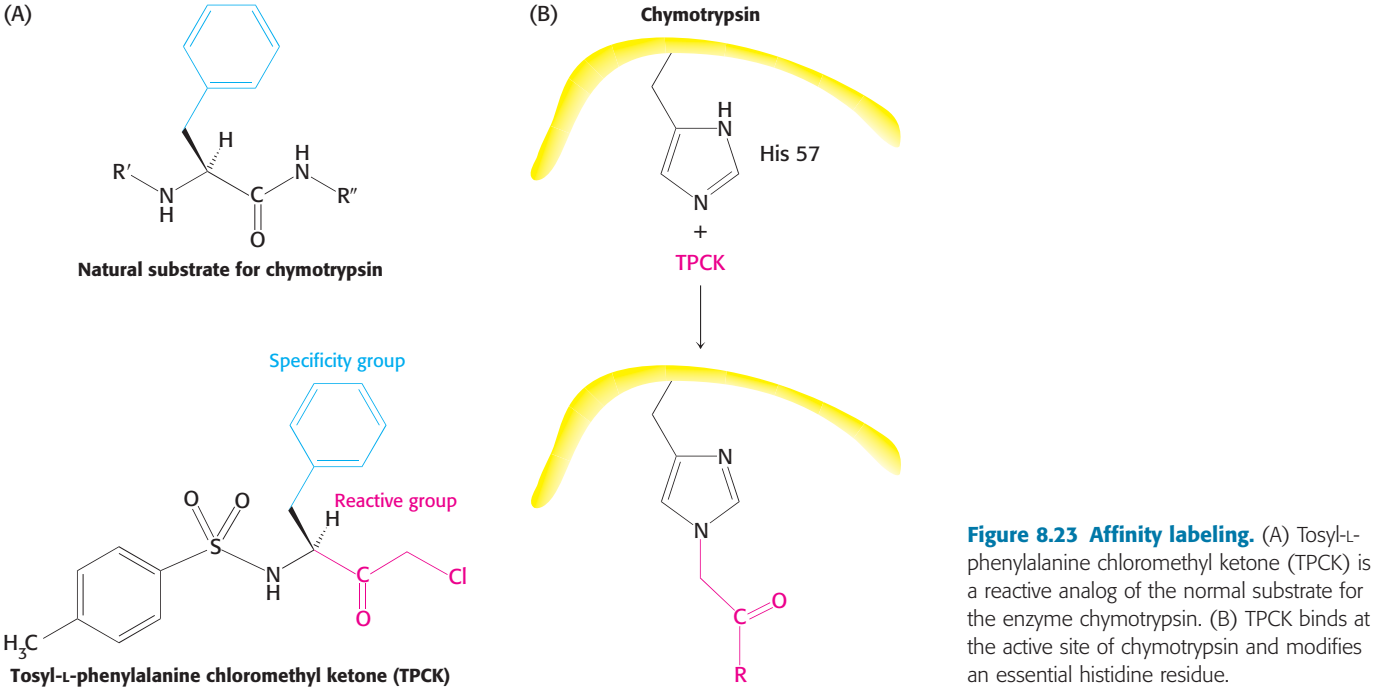


Figure 8.23 Affinity labeling. (A) Tosyl-L-phenylalanine chloromethyl ketone (TPCK) is a reactive analog of the normal substrate for the enzyme chymotrypsin. (B) TPCK binds at the active site of chymotrypsin and modifies an essential histidine residue.

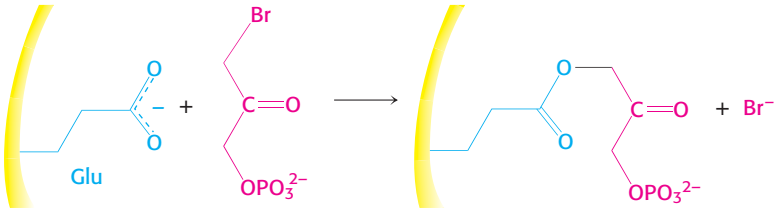
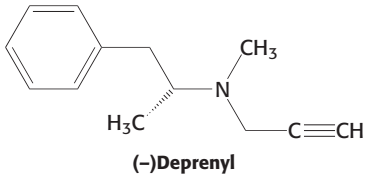


Figure 8.24 Bromoacetol phosphate, an affinity label for triose phosphate isomerase (TPI). Bromoacetol phosphate, an analog of dihydroxyacetone phosphate, binds at the active site of the enzyme and covalently modifies a glutamic acid residue required for enzyme activity.

binds at the active site and then reacts irreversibly with a histidine residue at that site, inhibiting the enzyme. The compound 3-bromoacetol phosphate is an affinity label for the enzyme triose phosphate isomerase (TPI). It mimics the normal substrate, dihydroxyacetone phosphate, by binding at the active site; then it covalently modifies the enzyme such that the enzyme is irreversibly inhibited (Figure 8.24).

Suicide inhibitors, or *mechanism-based inhibitors,* are modified substrates that provide the most specific means for modifying an enzyme's active site. The inhibitor binds to the enzyme as a substrate and is initially processed by the normal catalytic mechanism. The mechanism of catalysis then generates a chemically reactive intermediate that inactivates the enzyme through covalent modification. The fact that the enzyme participates in its own irreversible inhibition strongly suggests that the covalently modified group on the enzyme is vital for catalysis. One example of such an inhibitor is *N*,*N*-dimethylpropargylamine, an inhibitor of the enzyme monoamine oxidase (MAO). A flavin prosthetic group of monoamine oxidase oxidizes the *N*,*N*-dimethylpropargylamine, which in turn inactivates the enzyme by binding to N-5 of the flavin prosthetic group (Figure 8.25). Monoamine oxidase deaminates neurotransmitters such as dopamine and serotonin, lowering their levels in the brain. Parkinson disease is associated with low levels of dopamine, and depression is associated with low levels of serotonin. The drug (–)deprenyl, which is used to treat Parkinson disease and depression, is a suicide inhibitor of monoamine oxidase.



(–)Deprenyl

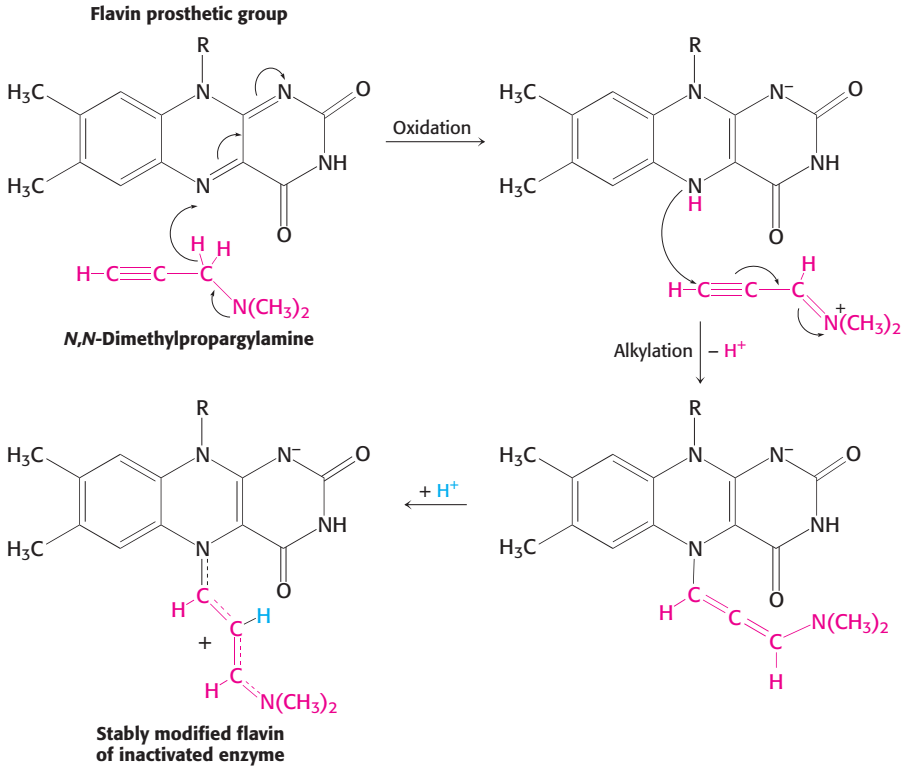


Figure 8.25 Mechanism-based (suicide) inhibition. Monoamine oxidase, an enzyme important for neurotransmitter synthesis, requires the cofactor FAD (flavin adenine dinucleotide). *N*,*N*-Dimethylpropargylamine inhibits monoamine oxidase by covalently modifying the flavin prosthetic group only after the inhibitor has been oxidized. The N-5 flavin adduct is stabilized by the addition of a proton. R represents the remainder of the flavin prosthetic group.

Transition-state analogs are potent inhibitors of enzymes

We turn now to compounds that provide the most intimate views of the catalytic process itself. Linus Pauling proposed in 1948 that compounds resembling the transition state of a catalyzed reaction should be very effective inhibitors of enzymes. These mimics are called *transition-state analogs*. The inhibition of proline racemase is an instructive example. The racemization of proline proceeds through a transition state in which the tetrahedral α-carbon atom has become trigonal (Figure 8.26). In the trigonal form, all three bonds are in the same plane; C_α also carries a net negative charge.

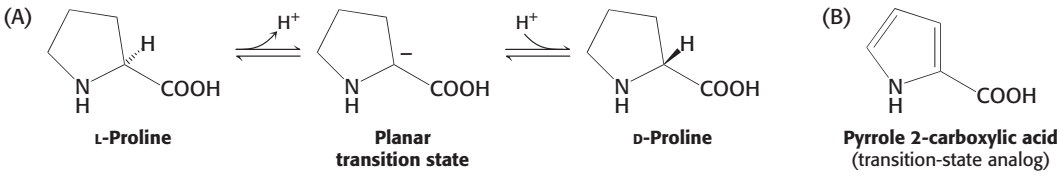


Figure 8.26 Inhibition by transition-state analogs. (A) The isomerization of L-proline to D-proline by proline racemase, a bacterial enzyme, proceeds through a planar transition state in which the α-carbon atom is trigonal rather than tetrahedral. (B) Pyrrole 2-carboxylic acid, a transition-state analog because of its trigonal geometry, is a potent inhibitor of proline racemase.

This symmetric carbanion can be reprotonated on one side to give the L isomer or on the other side to give the D isomer. This picture is supported by the finding that the inhibitor pyrrole 2-carboxylate binds to the racemase 160 times as tightly as does proline. *The α-carbon atom of this inhibitor, like that of the transition state, is trigonal.* An analog that also carries a negative charge on C_α would be expected to bind even more tightly. In general, highly potent and specific inhibitors of enzymes can be produced by synthesizing compounds that more closely resemble the transition state than the substrate itself. The inhibitory power of transition-state analogs underscores the essence of catalysis: *selective binding of the transition state*.

Catalytic antibodies demonstrate the importance of selective binding of the transition state to enzymatic activity

Antibodies that recognize transition states should function as catalysts, if our understanding of the importance of the transition state to catalysis is correct. The preparation of an antibody that catalyzes the insertion of a metal ion into a porphyrin nicely illustrates the validity of this approach. Ferrochelatase, the final enzyme in the biosynthetic pathway for the production of heme, catalyzes the insertion of Fe^{2+} into protoporphyrin IX. The nearly planar porphyrin must be bent for iron to enter.

The challenge was to find a transition-state analog for this metallation reaction that could be used as an antigen (immunogen) to generate an antibody. The solution came from studies showing that an alkylated porphyrin, *N*-methylmesoporphyrin, is a potent inhibitor of ferrochelatase. This compound resembles the transition state because N-*alkylation forces the porphyrin to be bent*. Moreover, *N*-alkylporphyrins were known to chelate metal ions 10^4 times as fast as their unalkylated counterparts do. Bending increases the exposure of the pyrrole nitrogen lone pairs of electrons to solvent, which enables the binding of the iron ion.

An antibody catalyst was produced with the use of an *N*-alkylporphyrin as the antigen. The resulting antibody presumably distorts a planar porphyrin to facilitate the entry of a metal ion (Figure 8.27). On average, an antibody molecule metallated 80 porphyrin molecules per hour, a rate only

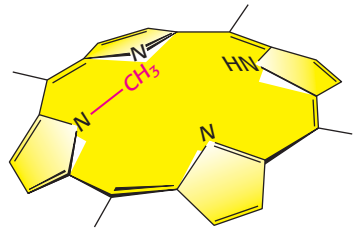


Figure 8.27 *N*-Methylmesoporphyrin is a transition-state analog used to generate catalytic antibodies. The insertion of a metal ion into a porphyrin by ferrochelatase proceeds through a transition state in which the porphyrin is bent. *N*-Methylmesoporphyrin, a bent porphyrin that resembles the transition state of the ferrochelatase-catalyzed reaction, was used to generate an antibody that also catalyzes the insertion of a metal ion into a porphyrin ring.

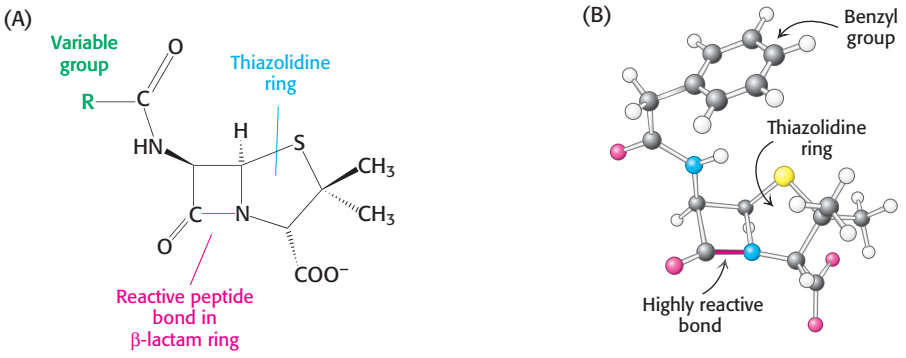


Figure 8.28 The reactive site of penicillin is the peptide bond of its β-lactam ring.
(A) Structural formula of penicillin.
(B) Representation of benzylpenicillin.

10-fold less than that of ferrochelatase, and 2500-fold faster than the uncatalyzed reaction. *Catalytic antibodies (abzymes) can indeed be produced by using transition-state analogs as antigens.* Antibodies catalyzing many other kinds of chemical reactions—exemplified by ester and amide hydrolysis, amide-bond formation, transesterification, photoinduced cleavage, photoinduced dimerization, decarboxylation, and oxidization—have been produced with the use of similar strategies. Studies with transition-state analogs provide strong evidence that enzymes can function by assuming a conformation in the active site that is complementary in structure to the transition state. *The power of transition-state analogs is now evident: (1) they are sources of insight into catalytic mechanisms, (2) they can serve as potent and specific inhibitors of enzymes, and (3) they can be used as immunogens to generate a wide range of novel catalysts.*

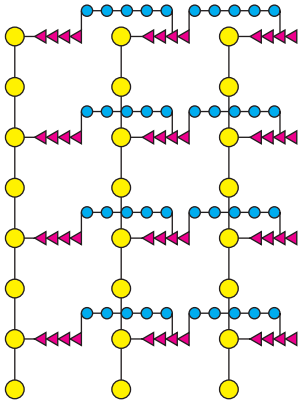


Figure 8.29 Schematic representation of the peptidoglycan in *Staphylococcus aureus*. The sugars are shown in yellow, the tetrapeptides in red, and the pentaglycine bridges in blue. The cell wall is a single, enormous, bag-shaped macromolecule because of extensive cross-linking.

Penicillin irreversibly inactivates a key enzyme in bacterial cell-wall synthesis

Penicillin, the first antibiotic discovered, provides us with an example of a clinically useful suicide inhibitor. Penicillin consists of a thiazolidine ring fused to a *β-lactam* ring to which a variable R group is attached by a peptide bond (Figure 8.28A). In benzylpenicillin, for example, R is a benzyl group (Figure 8.28B). This structure can undergo a variety of rearrangements, and, in particular, the β-lactam ring is very labile. Indeed, this instability is closely tied to the antibiotic action of penicillin, as will be evident shortly.

How does penicillin inhibit bacterial growth? Let us consider *Staphylococcus aureus,* the most common cause of staph infections. Penicillin works by interfering with the synthesis of the *S. aureus* cell walls. The *S. aureus* cell wall is made up of a macromolecule, called a *peptidoglycan* (Figure 8.29), which consists of linear polysaccharide chains that are cross-linked by short peptides (pentaglycines and tetrapeptides). The enormous bag-shaped peptidoglycan confers mechanical support and prevents bacteria from bursting in response to their high internal osmotic pressure.

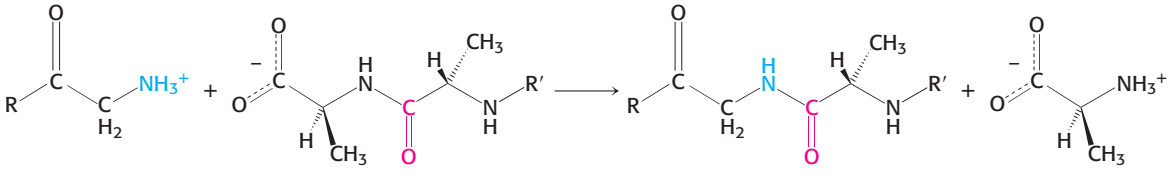


Figure 8.30 Formation of cross-links in *S. aureus* peptidoglycan. The terminal amino group of the pentaglycine bridge in the cell wall attacks the peptide bond between two D-alanine residues to form a cross-link.

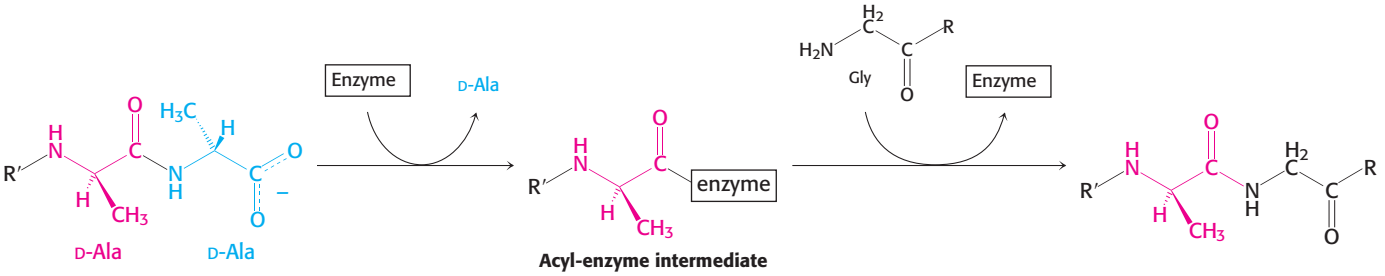


Figure 8.31 Transpeptidation reaction. An acyl-enzyme intermediate is formed in the transpeptidation reaction leading to cross-link formation.

Glycopeptide transpeptidase catalyzes the formation of the cross-links that make the peptidoglycan so stable (Figure 8.30). Bacterial cell walls are unique in containing D amino acids, which form cross-links by a mechanism different from that used to synthesize proteins.

Penicillin inhibits the cross-linking transpeptidase by the Trojan horse stratagem. The transpeptidase normally forms an *acyl intermediate* with the penultimate D-alanine residue of the D-Ala-D-Ala peptide (Figure 8.31). This covalent acyl-enzyme intermediate then reacts with the amino group of the terminal glycine in another peptide to form the cross-link. Penicillin is welcomed into the active site of the transpeptidase because it mimics the D-Ala-D-Ala moiety of the normal substrate (Figure 8.32). Bound penicillin then forms a covalent bond with a serine residue at the active site of the enzyme. *This penicilloyl-enzyme does not react further. Hence, the transpeptidase is irreversibly inhibited and cell-wall synthesis cannot take place.*

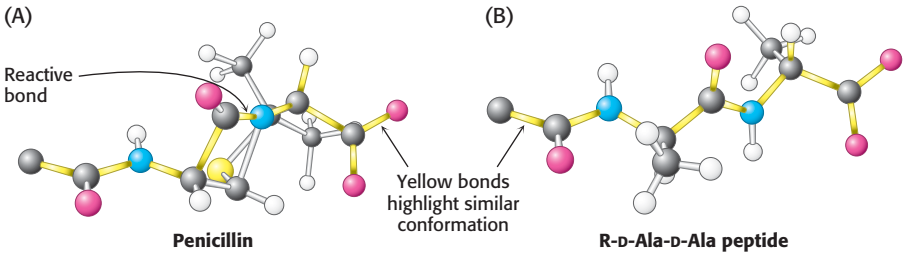


Figure 8.32 Conformations of penicillin and a normal substrate. The conformation of penicillin in the vicinity of its reactive peptide bond (A) resembles the postulated conformation of the transition state of R-D-Ala-D-Ala (B) in the transpeptidation reaction. [After B. Lee. *J. Mol. Biol.* 61:463–469, 1971.]

Why is penicillin such an effective inhibitor of the transpeptidase? The highly strained, four-membered β-lactam ring of penicillin makes it especially reactive. On binding to the transpeptidase, the serine residue at the active site attacks the carbonyl carbon atom of the lactam ring to form the penicilloyl-serine derivative (Figure 8.33). Because the peptidase participates in its own inactivation, penicillin acts as a suicide inhibitor.

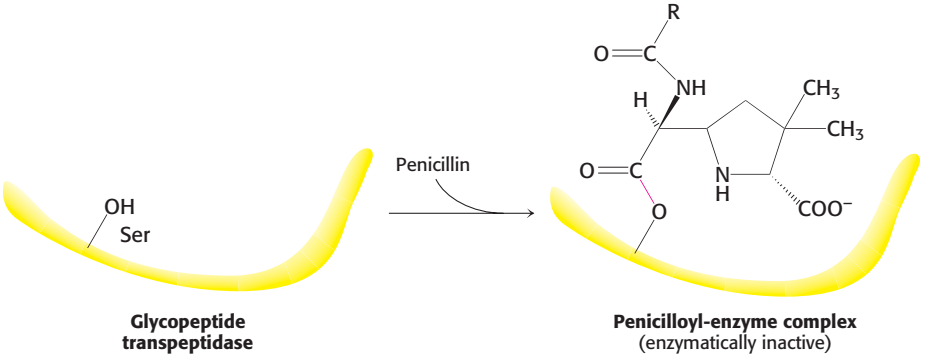


Figure 8.33 Formation of a penicilloyl-enzyme complex. Penicillin reacts with the transpeptidase to form an inactive complex, which is indefinitely stable.

8.6 Enzymes Can Be Studied One Molecule at a Time

Most experiments that are preformed to determine an enzyme characteristic require an enzyme preparation in a buffered solution. Even a few microliters of such a solution will contain millions of enzyme molecules. Much that we have learned about enzymes thus far has come from such experiments, called *ensemble studies*. A basic assumption of ensemble studies is that all of the enzymes are the same or very similar. When we determine an enzyme property such as the value of K_M in ensemble studies, that value is of necessity an average value of all of the enzymes present. However, we know that molecular heterogeneity, the ability of a molecule, over time, to assume several different structures that differ slightly in stability, is an inherent property of all large biomolecules. Recall that prions can exist in two different structures, one of which is prone to aggregation (pp. 55–56). How can we tell if this molecular heterogeneity affects enzyme activity?

By way of example, consider a hypothetical situation. A Martian visits Earth to learn about higher education. The spacecraft hovers high above a university, and our Martian meticulously records how the student population moves about campus. Much information can be gathered from such studies: where students are likely to be at certain times on certain days, which buildings are used when and by how many. Now, suppose our visitor developed a high-magnification camera that could follow one student throughout the day. Such data would provide a much different perspective on college life: What does this student eat? To whom does she talk? How much time does she spend studying? This new *in singulo* method, examining one individual at a time, yields a host of new information but also illustrates a potential pitfall of studying individuals, be they students or enzymes: How can we be certain that the student or molecule is representative and not an outlier? This pitfall can be overcome by studying enough individuals to satisfy statistical analysis for validity.

Let us leave our Martian to his observations, and consider a more biochemical situation. Figure 8.34A shows an enzyme that displays molecular heterogeneity, with three active forms that catalyze the same reaction but at different rates. These forms have slightly different stabilities, but thermal noise is sufficient to interconvert the forms. Each form is present as a fraction of the total enzyme population as indicated. If we were to perform an experiment to determine enzyme activity under a particular set of conditions with the use of ensemble methods, we would get a single value, which would represent the average of the heterogeneous assembly (Figure 8.34B). However, were we to perform a sufficient number of single-molecule experiments, we would discover that the enzyme has three different molecular forms with very different activities (Figure 8.34C). Moreover, these different forms would most likely correspond to important biochemical differences.

The development of powerful techniques—such as patch-clamp recording, single-molecule fluorescence, and optical tweezers—has enabled biochemists to look into the workings of individual molecules. We will examine single-molecule studies of membrane channels with the use of patch-clamp recording (Section 13.4), ATP-synthesizing complexes with the use of single-molecule fluorescence and molecular motors with the use of an optical trap (Section 34.2). Single-molecule studies open a new vista on the function of enzymes in particular and on all large biomolecules in general.

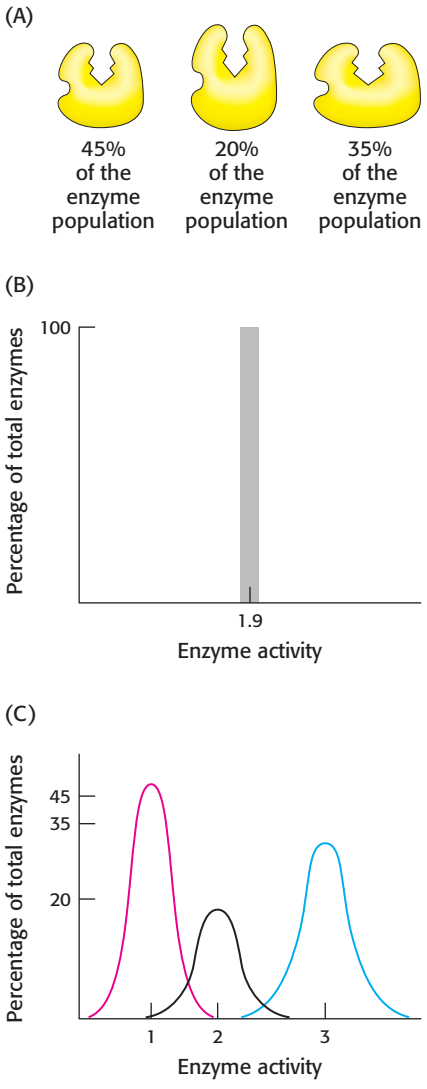


Figure 8.34 Single molecule studies can reveal molecular heterogeneity. (A) Complex biomolecules, such as enzymes, display molecular heterogeneity. (B) When measuring an enzyme property using ensemble methods, an average value of the all of the enzymes present is the result. (C) Single enzyme studies reveal molecular heterogeneity, with the various forms showing different properties.

Summary

8.1 Enzymes Are Powerful and Highly Specific Catalysts

Most catalysts in biological systems are enzymes, and nearly all enzymes are proteins. Enzymes are highly specific and have great catalytic power. They can enhance reaction rates by factors of 10^6 or more. Many enzymes require cofactors for activity. Such cofactors can be metal ions or small, vitamin-derived organic molecules called coenzymes.

8.2 Free Energy Is a Useful Thermodynamic Function for Understanding Enzymes

Free energy (G) is the most valuable thermodynamic function for determining whether a reaction can take place and for understanding the energetics of catalysis. A reaction can take place spontaneously only if the change in free energy (ΔG) is negative. The free-energy change of a reaction that takes place when reactants and products are at unit activity is called the standard free-energy change ($\Delta G°$). Biochemists usually use $\Delta G°'$, the standard free-energy change at pH 7. Enzymes do not alter reaction equilibria; rather, they increase reaction rates.

8.3 Enzymes Accelerate Reactions by Facilitating the Formation of the Transition State

Enzymes serve as catalysts by decreasing the free energy of activation of chemical reactions. Enzymes accelerate reactions by providing a reaction pathway in which the transition state (the highest-energy species) has a lower free energy and hence is more rapidly formed than in the uncatalyzed reaction.

The first step in catalysis is the formation of an enzyme–substrate complex. Substrates are bound to enzymes at active-site clefts from which water is largely excluded when the substrate is bound. The specificity of enzyme–substrate interactions arises mainly from hydrogen bonding, which is directional, and from the shape of the active site, which rejects molecules that do not have a sufficiently complementary shape. The recognition of substrates by enzymes is often accompanied by conformational changes at active sites, and such changes facilitate the formation of the transition state.

8.4 The Michaelis–Menten Model Accounts for the Kinetic Properties of Many Enzymes

The kinetic properties of many enzymes are described by the Michaelis–Menten model. In this model, an enzyme (E) combines with a substrate (S) to form an enzyme–substrate (ES) complex, which can proceed to form a product (P) or to dissociate into E and S.

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{E} + \mathrm{P}$$

The rate V_0 of formation of product is given by the Michaelis–Menten equation:

$$V_0 = V_{\max}\frac{[S]}{[S] + K_M}$$

in which V_{max} is the reaction rate when the enzyme is fully saturated with substrate and K_M, the Michaelis constant, is the substrate concentration at which the reaction rate is half maximal. The maximal rate, V_{max}, is equal to the product of k_2, or k_{cat}, and the total concentration of enzyme. The kinetic constant k_{cat}, called the turnover number, is

the number of substrate molecules converted into product per unit time at a single catalytic site when the enzyme is fully saturated with substrate. Turnover numbers for most enzymes are between 1 and 10^4 per second. The ratio of k_{cat}/K_M provides a penetrating probe into enzyme efficiency.

Allosteric enzymes constitute an important class of enzymes whose catalytic activity can be regulated. These enzymes, which do not conform to Michaelis–Menten kinetics, have multiple active sites. These active sites display cooperativity, as evidenced by a sigmoidal dependence of reaction velocity on substrate concentration.

8.5 Enzymes Can Be Inhibited by Specific Molecules

Specific small molecules or ions can inhibit even nonallosteric enzymes. In irreversible inhibition, the inhibitor is covalently linked to the enzyme or bound so tightly that its dissociation from the enzyme is very slow. Covalent inhibitors provide a means of mapping the enzyme's active site. In contrast, reversible inhibition is characterized by a more rapid equilibrium between enzyme and inhibitor. A competitive inhibitor prevents the substrate from binding to the active site. It reduces the reaction velocity by diminishing the proportion of enzyme molecules that are bound to substrate. Competitive inhibition can be overcome by raising the substrate concentration. In uncompetitive inhibition, the inhibitor combines only with the enzyme–substrate complex. In noncompetitive inhibition, the inhibitor decreases the turnover number. Uncompetitive and noncompetitive inhibition cannot be overcome by raising the substrate concentration.

The essence of catalysis is selective stabilization of the transition state. Hence, an enzyme binds the transition state more tightly than it binds the substrate. Transition-state analogs are stable compounds that mimic key features of this highest-energy species. They are potent and specific inhibitors of enzymes. Proof that transition-state stabilization is a key aspect of enzyme activity comes from the generation of catalytic antibodies. Transition-state analogs are used as antigens, or immunogens, in generating catalytic antibodies.

8.6 Enzymes Can Be Studied One Molecule at a Time

Many enzymes are now being studied *in singulo,* at the level of a single molecule. Such studies are important because they yield information that is difficult to obtain in studies of populations of molecules. Single-molecule methods reveal a distribution of enzyme characteristics rather than an average value as is acquired with the use of ensemble methods.

APPENDIX: Enzymes Are Classified on the Basis of the Types of Reactions That They Catalyze

Many enzymes have common names that provide little information about the reactions that they catalyze. For example, a proteolytic enzyme secreted by the pancreas is called trypsin. Most other enzymes are named for their substrates and for the reactions that they catalyze, with the suffix "ase" added. Thus, a peptide hydrolase is an enzyme that hydrolyzes peptide bonds, whereas ATP synthase is an enzyme that synthesizes ATP.

To bring some consistency to the classification of enzymes, in 1964 the International Union of Biochemistry established an Enzyme Commission to develop a nomenclature for enzymes. Reactions were divided into six major groups numbered 1 through 6 (Table 8.8). These groups were subdivided and further subdivided so that a four-digit number preceded by the letters *EC* for Enzyme Commission could precisely identify all enzymes.

Consider as an example nucleoside monophosphate (NMP) kinase, an enzyme that we will examine in detail in Section 9.4. It catalyzes the following reaction:

$$\text{ATP} + \text{NMP} \rightleftharpoons \text{ADP} + \text{NDP}$$

Table 8.8 Six major classes of enzymes

Class	Type of reaction	Example	Chapter
1. Oxidoreductases	Oxidation–reduction	Lactate dehydrogenase	16
2. Transferases	Group transfer	Nucleoside monophosphate kinase (NMP kinase)	9
3. Hydrolases	Hydrolysis reactions (transfer of functional groups to water)	Chymotrypsin	9
4. Lyases	Addition or removal of groups to form double bonds	Fumarase	17
5. Isomerases	Isomerization (intramolecular group transfer)	Triose phosphate isomerase	16
6. Ligases	Ligation of two substrates at the expense of ATP hydrolysis	Aminoacyl-tRNA synthetase	30

NMP kinase transfers a phosphoryl group from ATP to NMP to form a nucleoside diphosphate (NDP) and ADP. Consequently, it is a transferase, or member of group 2. Many groups other than phosphoryl groups, such as sugars and single-carbon units, can be transferred. Transferases that shift a phosphoryl group are designated 2.7. Various functional groups can accept the phosphoryl group. If a phosphate is the acceptor, the transferase is designated 2.7.4. The final number designates the acceptor more precisely. In regard to NMP kinase, a nucleoside monophosphate is the acceptor, and the enzyme's designation is EC 2.7.4.4. Although the common names are used routinely, the classification number is used when the precise identity of the enzyme might be ambiguous.

Key Terms

enzyme (p. 220)
substrate (p. 220)
cofactor (p. 221)
apoenzyme (p. 221)
holoenzyme (p. 221)
coenzyme (p. 221)
prosthetic group (p. 221)
free energy (p. 222)
free energy of activation (p. 222)
transition state (p. 225)
active site (p. 227)
induced fit (p. 228)
K_M (the Michaelis constant) (p. 231)
V_{max} (maximal rate) (p. 232)
Michaelis–Menten equation (p. 232)
Lineweaver–Burk equation (double-reciprocal plot) (p. 233)
turnover number (p. 234)
k_{cat}/K_M ratio (p. 235)
sequential reaction (p. 236)
double-displacement (ping-pong) reaction (p. 237)
allosteric enzyme (p. 237)
competitive inhibition (p. 238)
uncompetitive inhibition (p. 238)
noncompetitive inhibition (p. 238)
group-specific reagent (p. 241)
affinity label (reactive substrate analog) (p. 241)
mechanism-based (suicide) inhibition (p. 242)
transition-state analog (p. 243)
catalytic antibody (abzyme) (p. 244)

Problems

1. *Raisons d'etre*. What are the two properties of enzymes that make them especially useful catalysts?

2. *Partners*. What does an apoenzyme require to become a holoenzyme?

3. *Different partners*. What are the two main types of cofactors?

4. *One a day*. Why are vitamins necessary for good health?

5. *A function of state*. What is the fundamental mechanism by which enzymes enhance the rate of chemical reactions?

6. *Nooks and crannies*. What is the structural basis for enzyme specificity?

7. *Give with one hand, take with the other*. Why does the activation energy of a reaction not appear in the final ΔG of the reaction?

8. *Mountain climbing*. Proteins are thermodynamically unstable. The ΔG of the hydrolysis of proteins is quite negative, yet proteins can be quite stable. Explain this apparent paradox. What does it tell you about protein synthesis?

9. *Protection.* Suggest why the enzyme lysozyme, which degrades cell walls of some bacteria, is present in tears.

10. *Stability matters.* Transition-state analogs, which can be used as enzyme inhibitors and to generate catalytic antibodies, are often difficult to synthesize. Suggest a reason.

11. *Match'em.* Match the K'_{eq} values with the appropriate $\Delta G^{\circ\prime}$ values.

	K'_{eq}	$\Delta G^{\circ\prime}$ (kJ mol^{-1})
(a)	1	28.53
(b)	10^{-5}	−11.42
(c)	10^{4}	5.69
(d)	10^{2}	0
(e)	10^{-1}	−22.84

12. *Free energy!* Assume that you have a solution of 0.1 M glucose 6-phosphate. To this solution, you add the enzyme phosphoglucomutase, which catalyzes the following reaction:

$$\text{Glucose 6-phosphate} \xrightleftharpoons{\text{Phosphoglucomutase}} \text{glucose 1-phosphate}$$

The $\Delta G^{\circ\prime}$ for the reaction is +7.5 kJ mol^{-1} (+1.8 kcal mol^{-1}).

(a) Does the reaction proceed as written? If so, what are the final concentrations of glucose 6-phosphate and glucose 1-phosphate?

(b) Under what cellular conditions could you produce glucose 1-phosphate at a high rate?

13. *Free energy, too!* Consider the following reaction:

$$\text{Glucose 1-phosphate} \rightleftharpoons \text{glucose 6-phosphate}$$

After reactant and product were mixed and allowed to reach equilibrium at 25°C, the concentration of each compound was measured:

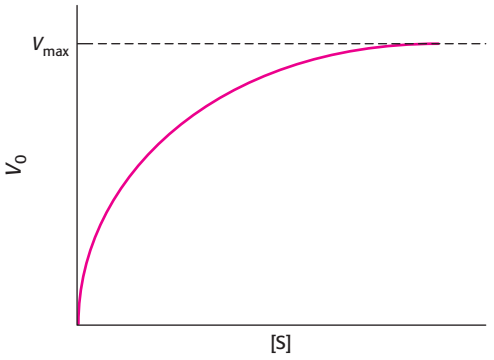
$$[\text{Glucose 1-phosphate}]_{eq} = 0.01 \text{ M}$$
$$[\text{Glucose 6-phosphate}]_{eq} = 0.19 \text{ M}$$

Calculate K_{eq} and $\Delta G^{\circ\prime}$.

14. *Keeping busy.* Many isolated enzymes, if incubated at 37°C, will be denatured. However, if the enzymes are incubated at 37°C in the presence of substrate, the enzymes are catalytically active. Explain this apparent paradox.

15. *Active yet responsive.* What is the biochemical advantage of having a K_M approximately equal to the substrate concentration normally available to an enzyme?

16. *Angry biochemists.* Many biochemists go bananas, and justifiably, when they see a Michaelis–Menten plot like the one shown at the top of the next column. To see why, determine the V_0 as a fraction of V_{max} when the substrate concentration is equal to 10 K_M and 20 K_M. Please control your outrage.



17. *Hydrolytic driving force.* The hydrolysis of pyrophosphate to orthophosphate is important in driving forward biosynthetic reactions such as the synthesis of DNA. This hydrolytic reaction is catalyzed in *Escherichia coli* by a pyrophosphatase that has a mass of 120 kd and consists of six identical subunits. For this enzyme, a unit of activity is defined as the amount of enzyme that hydrolyzes 10 μmol of pyrophosphate in 15 minutes at 37°C under standard assay conditions. The purified enzyme has a V_{max} of 2800 units per milligram of enzyme.

(a) How many moles of substrate is hydrolyzed per second per milligram of enzyme when the substrate concentration is much greater than K_M?

(b) How many moles of active sites is there in 1 mg of enzyme? Assume that each subunit has one active site.

(c) What is the turnover number of the enzyme? Compare this value with others mentioned in this chapter.

18. *Destroying the Trojan horse.* Penicillin is hydrolyzed and thereby rendered inactive by penicillinase (also known as β-lactamase), an enzyme present in some penicillin-resistant bacteria. The mass of this enzyme in *Staphylococcus aureus* is 29.6 kd. The amount of penicillin hydrolyzed in 1 minute in a 10-ml solution containing 10^{-9} g of purified penicillinase was measured as a function of the concentration of penicillin. Assume that the concentration of penicillin does not change appreciably during the assay.

[Penicillin] μM	Amount hydrolyzed (nmol)
1	0.11
3	0.25
5	0.34
10	0.45
30	0.58
50	0.61

(a) Plot V_0 versus [S] and $1/V_0$ versus $1/[S]$ for these data. Does penicillinase appear to obey Michaelis–Menten kinetics? If so, what is the value of K_M?

(b) What is the value of V_{max}?

(c) What is the turnover number of penicillinase under these experimental conditions? Assume one active site per enzyme molecule.

19. *Counterpoint.* Penicillinase (β-lactamase) hydrolyzes penicillin. Compare penicillinase with glycopeptide transpeptidase.

20. *A different mode.* The kinetics of an enzyme are measured as a function of substrate concentration in the presence and absence of 100 μM inhibitor.

(a) What are the values of V_{max} and K_M in the presence of this inhibitor?

(b) What type of inhibition is it?

(c) What is the dissociation constant of this inhibitor?

	Velocity (μmol minute^{-1})	
[S] (μM)	No inhibitor	Inhibitor
3	10.4	2.1
5	14.5	2.9
10	22.5	4.5
30	33.8	6.8
90	40.5	8.1

(d) If [S] = 30 μM, what fraction of the enzyme molecules have a bound substrate in the presence and in the absence of 100 μM inhibitor?

21. *A fresh view.* The plot of $1/V_0$ versus $1/[S]$ is sometimes called a Lineweaver–Burk plot. Another way of expressing the kinetic data is to plot V_0 versus $V_0/[S]$, which is known as an Eadie–Hofstee plot.

(a) Rearrange the Michaelis–Menten equation to give V_0 as a function of $V_0/[S]$.

(b) What is the significance of the slope, the vertical intercept, and the horizontal intercept in a plot of V_0 versus $V_0/[S]$?

(c) Sketch a plot of V_0 versus $V_0/[S]$ in the absence of an inhibitor, in the presence of a competitive inhibitor, and in the presence of a noncompetitive inhibitor.

22. *Competing substrates.* Suppose that two substrates, A and B, compete for an enzyme. Derive an expression relating the ratio of the rates of utilization of A and B, V_A/V_B, to the concentrations of these substrates and their values of k_{cat} and K_M. (Hint: Express V_A as a function of k_{cat}/K_M for substrate A, and do the same for V_B.) Is specificity determined by K_M alone?

23. *A tenacious mutant.* Suppose that a mutant enzyme binds a substrate 100 times as tightly as does the native enzyme. What is the effect of this mutation on catalytic rate if the binding of the transition state is unaffected?

24. *More Michaelis–Menten.* For an enzyme that follows simple Michaelis–Menten kinetics, what is the value of V_{max} if V_0 is equal to 1 μmol minute^{-1} at 10 K_M?

25. *Controlled paralysis.* Succinylcholine is a fast-acting, short-duration muscle relaxant that is used when a tube is inserted into a patient's trachea or when a bronchoscope is used to examine the trachea and bronchi for signs of cancer. Within seconds of the administration of succinylcholine, the patient experiences muscle paralysis and is placed on a respirator while the examination proceeds. Succinylcholine is a competitive inhibitor of acetylcholinesterase, a nervous system enzyme, and this inhibition causes paralysis. However, succinylcholine is hydrolyzed by blood-serum cholinesterase, which shows a broader substrate specificity than does the nervous system enzyme. Paralysis lasts until the succinylcholine is hydrolyzed by the serum cholinesterase, usually several minutes later.

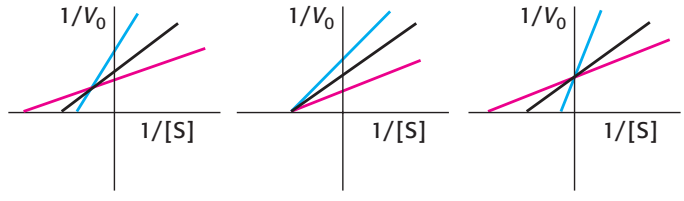
(a) As a safety measure, serum cholinesterase is measured before the examination takes place. Explain why this measurement is good idea.

(b) What would happen to the patient if the serum cholinesterase activity were only 10 units of activity per liter rather than the normal activity of about 80 units?

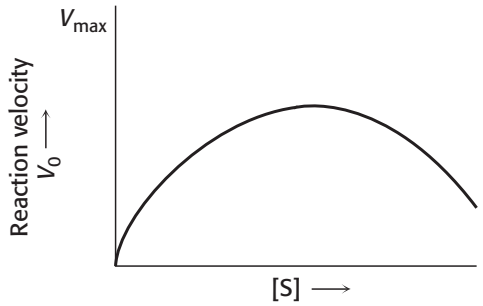
(c) Some patients have a mutant form of the serum cholinesterase that displays a K_M of 10 mM, rather than the normal 1.4 mM. What will be the effect of this mutation on the patient?

Data Interpretation Problems

26. *Varying the enzyme.* For a one-substrate, enzyme-catalyzed reaction, double-reciprocal plots were determined for three different enzyme concentrations. Which of the following three families of curve would you expect to be obtained? Explain.



27. *Too much of a good thing.* A simple Michaelis–Menten enzyme, in the absence of any inhibitor, displayed the following kinetic behavior. The expected value of V_{max} is shown on the y-axis in the graph on the following page.



(a) Draw a double-reciprocal plot that corresponds to the velocity-versus-substrate curve.

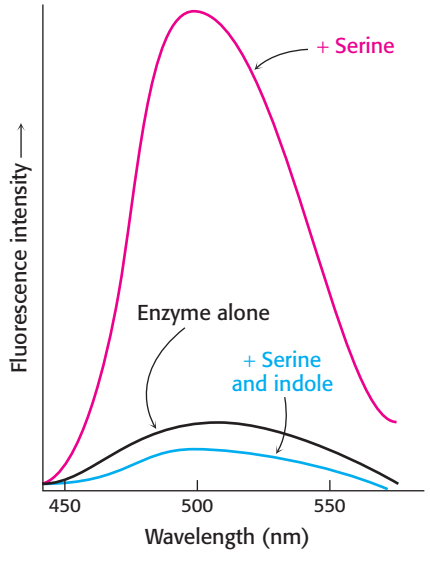
(b) Explain the kinetic results.

28. *Rate-limiting step.* In the conversion of A into D in the following biochemical pathway, enzymes E_A, E_B, and E_C have the K_M values indicated under each enzyme. If all of the substrates and products are present at a concentration of 10^{-4}M and the enzymes have approximately the same V_{max}, which step will be rate limiting and why?

$$A \underset{E_A}{\rightleftharpoons} B \underset{E_B}{\rightleftharpoons} C \underset{E_C}{\rightleftharpoons} D$$

$$K_M = \quad 10^{-2}\,M \qquad 10^{-4}\,M \qquad 10^{-4}\,M$$

29. *Colored luminosity* Tryptophan synthetase, a bacterial enzyme that contains a pyridoxal phosphate (PLP) prosthetic group, catalyzes the synthesis of L-tryptophan from L-serine and an indole derivative. The addition of L-serine to the enzyme produces a marked increase in the fluorescence of the PLP group, as the adjoining graph shows. The subsequent addition of indole, the second substrate, reduces this fluorescence to a level even lower than that produced by the enzyme alone. How do these changes in fluorescence support the notion that the enzyme interacts directly with its substrates?



Chapter Integration Problems

30. *Titration experiment.* The effect of pH on the activity of an enzyme was examined. At its active site, the enzyme has an ionizable group that must be negatively charged for substrate binding and catalysis to take place. The ionizable group has a pK_a of 6.0. The substrate is positively charged throughout the pH range of the experiment.

$$E + S^+ \rightleftharpoons E - S^+ \longrightarrow E^- + P^+$$

$$+$$

$$H^+$$

$$\Downarrow$$

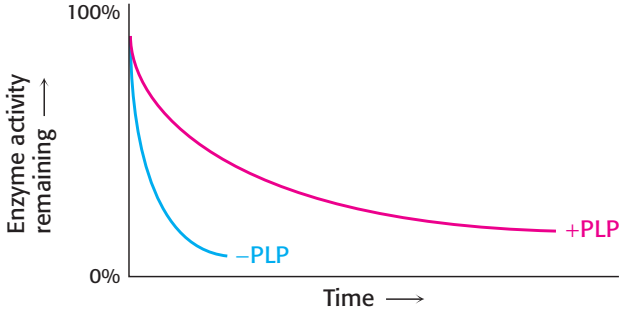
$$EH$$

(a) Draw the V_0-versus-pH curve when the substrate concentration is much greater than the enzyme K_M.

(b) Draw the V_0-versus-pH curve when the substrate concentration is much less than the enzyme K_M.

(c) At which pH will the velocity equal one-half of the maximal velocity attainable under these condition?

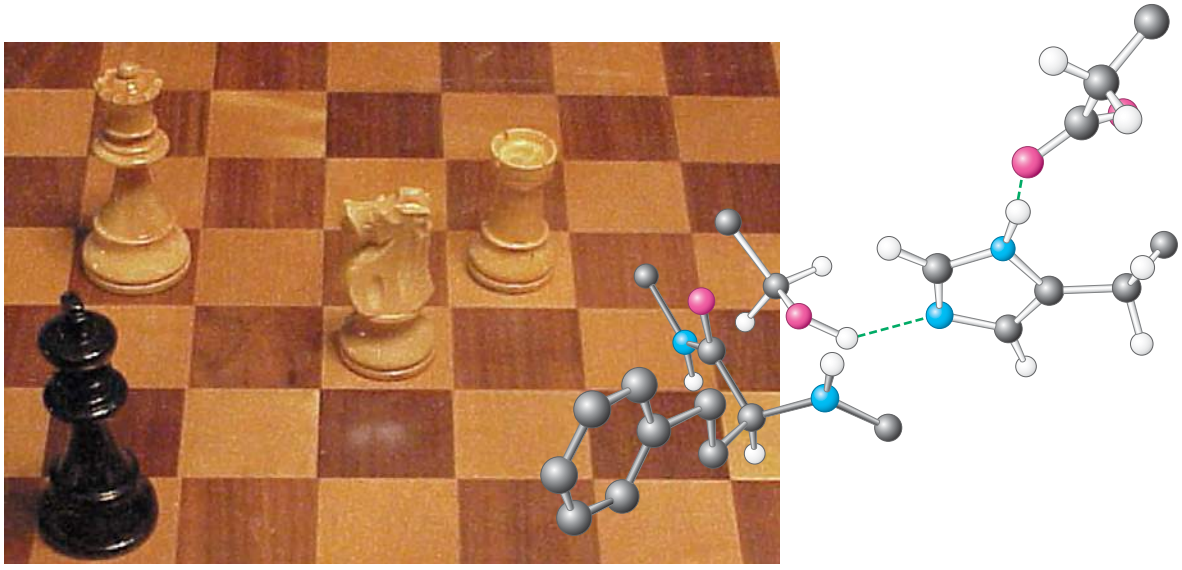
31. *A question of stability.* Pyridoxal phosphate (PLP) is a coenzyme for the enzyme ornithine aminotransferase. The enzyme was purified from cells grown in PLP-deficient media as well as from cells grown in media that contained pyridoxal phosphate. The stability of the enzyme was then measured by incubating the enzyme at 37°C and assaying for the amount of enzyme activity remaining. The following results were obtained.



(a) Why does the amount of active enzyme decrease with the time of incubation?

(b) Why does the amount of enzyme from the PLP-deficient cells decline more rapidly?

Catalytic Strategies



Chess and enzymes have in common the use of strategy, consciously thought out in the game of chess and selected by evolution for the action of an enzyme. The three amino acid residues at the right, denoted by the white bonds, constitute a catalytic triad found in the active site of a class of enzymes that cleave peptide bonds. The substrate, represented by the molecule with the black bonds, is as hopelessly trapped as the king in the photograph of a chess match at the left and is sure to be cleaved. [Photograph courtesy of Wendie Berg.]

What are the sources of the catalytic power and specificity of enzymes? This chapter presents the catalytic strategies used by four classes of enzymes: serine proteases, carbonic anhydrases, restriction endonucleases, and myosins. Each class catalyzes reactions that require the addition of water to a substrate. The mechanisms of these enzymes have been revealed through the use of incisive experimental probes, including the techniques of protein structure determination (Chapter 3) and site-directed mutagenesis (Chapter 5). The mechanisms illustrate many important principles of catalysis. We shall see how these enzymes facilitate the formation of the transition state through the use of binding energy and induced fit as well as classes of several specific catalytic strategies.

Each of the four classes of enzymes in this chapter illustrates the use of such strategies to solve a different problem. For serine proteases, exemplified by chymotrypsin, the challenge is to promote a reaction that is almost immeasurably slow at neutral pH in the absence of a catalyst. For carbonic anhydrases, the challenge is to achieve a high absolute rate of reaction, suitable for integration with other rapid physiological processes. For restriction endonucleases such as *Eco*RV, the challenge is to attain a high degree of specificity. Finally, for myosins, the challenge is to utilize the free

OUTLINE

energy associated with the hydrolysis of adenosine triphosphate (ATP) to drive other processes. Each of the examples selected is a member of a large protein class. For each of these classes, comparison between class members reveals how enzyme active sites have evolved and been refined. Structural and mechanistic comparisons of enzyme action are thus the sources of insight into the evolutionary history of enzymes. In addition, our knowledge of catalytic strategies has been used to develop practical applications, including potent drugs and specific enzyme inhibitors. Finally, although we shall not consider catalytic RNA molecules explicitly in this chapter, the principles also apply to these catalysts.

A few basic catalytic principles are used by many enzymes

In Chapter 8, we learned that enzymatic catalysis begins with substrate binding. The *binding energy* is the free energy released in the formation of a large number of weak interactions between the enzyme and the substrate. We can envision this binding energy as serving two purposes: it establishes substrate specificity and increases catalytic efficiency. Only the correct substrate can participate in most or all of the interactions with the enzyme and thus maximize binding energy, accounting for the exquisite substrate specificity exhibited by many enzymes. Furthermore, the full complement of such interactions is formed only when the combination of enzyme and substrate is in the transition state. Thus, interactions between the enzyme and the substrate stabilize the transition state, thereby lowering the free energy of activation. The binding energy can also promote structural changes in both the enzyme and the substrate that facilitate catalysis, a process referred to as *induced fit*.

Enzymes commonly employ one or more of the following strategies to catalyze specific reactions:

1. *Covalent Catalysis.* In covalent catalysis, the active site contains a reactive group, usually a powerful nucleophile, that becomes temporarily covalently attached to a part of the substrate in the course of catalysis. The proteolytic enzyme chymotrypsin provides an excellent example of this strategy (Section 9.1).

2. *General Acid–Base Catalysis.* In general acid–base catalysis, a molecule other than water plays the role of a proton donor or acceptor. Chymotrypsin uses a histidine residue as a base catalyst to enhance the nucleophilic power of serine (Section 9.1), whereas a histidine residue in carbonic anhydrase facilitates the removal of a hydrogen ion from a zinc-bound water molecule to generate hydroxide ion (Section 9.2). For myosins, a phosphate group of the ATP substrate serves as a base to promote its own hydrolysis (Section 9.3).

3. *Catalysis by Approximation.* Many reactions include two distinct substrates, including all four classes of hydrolases considered in detail in this chapter. In such cases, the reaction rate may be considerably enhanced by bringing the two substrates together along a single binding surface on an enzyme. For example, carbonic anhydrase binds carbon dioxide and water in adjacent sites to facilitate their reaction (Section 9.2).

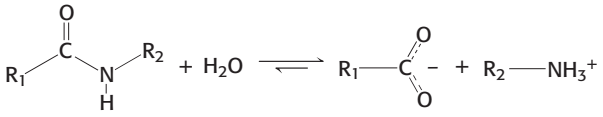
4. *Metal Ion Catalysis.* Metal ions can function catalytically in several ways. For instance, a metal ion may facilitate the formation of nucleophiles such as hydroxide ion by direct coordination. A zinc(II) ion serves this purpose in catalysis by carbonic anhydrase (Section 9.2). Alternatively, a metal ion may serve as an electrophile, stabilizing a negative charge on a reaction intermediate. A magnesium(II) ion plays this role in *Eco*RV (Section 9.3).

Finally, a metal ion may serve as a bridge between enzyme and substrate, increasing the binding energy and holding the substrate in a conformation appropriate for catalysis. This strategy is used by myosins (Section 9.4) and, indeed, by almost all enzymes that utilize ATP as a substrate.

9.1 Proteases Facilitate a Fundamentally Difficult Reaction

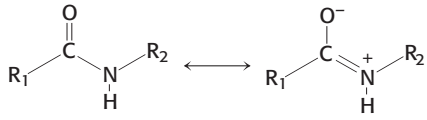
Protein turnover is an important process in living systems (Chapter 23). Proteins that have served their purpose must be degraded so that their constituent amino acids can be recycled for the synthesis of new proteins. Proteins ingested in the diet must be broken down into small peptides and amino acids for absorption in the gut. Furthermore, as described in detail in Chapter 10, proteolytic reactions are important in regulating the activity of certain enzymes and other proteins.

Proteases cleave proteins by a hydrolysis reaction—the addition of a molecule of water to a peptide bond:



Although the hydrolysis of peptide bonds is thermodynamically favored, such hydrolysis reactions are extremely slow. In the absence of a catalyst, the half-life for the hydrolysis of a typical peptide at neutral pH is estimated to be between 10 and 1000 years. Yet, peptide bonds must be hydrolyzed within milliseconds in some biochemical processes.

The chemical bonding in peptide bonds is responsible for their kinetic stability. Specifically, the resonance structure that accounts for the planarity of a peptide bond (Section 2.2) also makes such bonds resistant to hydrolysis. This resonance structure endows the peptide bond with partial double-bond character:



The carbon–nitrogen bond is strengthened by its double-bond character. Furthermore, the carbonyl carbon atom is less electrophilic and less susceptible to nucleophilic attack than are the carbonyl carbon atoms in more reactive compounds such as carboxylate esters. Consequently, to promote peptide-bond cleavage, an enzyme must facilitate nucleophilic attack at a normally unreactive carbonyl group.

Chymotrypsin possesses a highly reactive serine residue

A number of proteolytic enzymes participate in the breakdown of proteins in the digestive systems of mammals and other organisms. One such enzyme, chymotrypsin, cleaves peptide bonds selectively on the carboxyl-terminal side of the large hydrophobic amino acids such as tryptophan, tyrosine, phenylalanine, and methionine (Figure 9.1). Chymotrypsin is a good example of the use of *covalent catalysis*. The enzyme employs a powerful nucleophile to attack the unreactive carbonyl carbon atom of the substrate. This nucleophile becomes covalently attached to the substrate briefly in the course of catalysis.

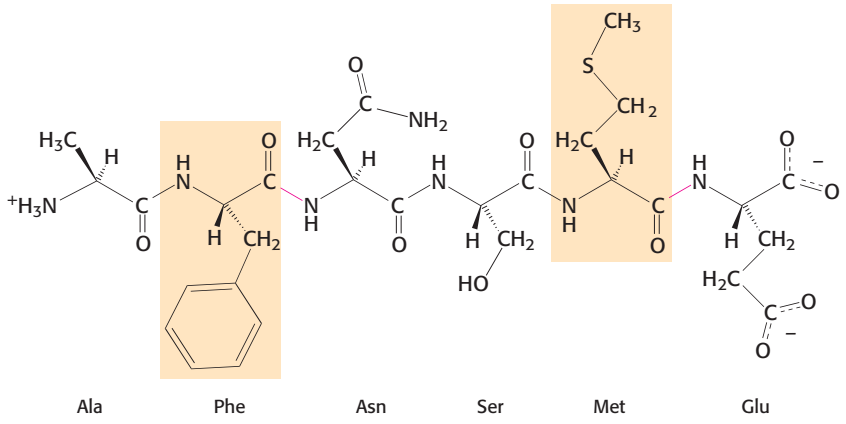


Figure 9.1 Specificity of chymotrypsin. Chymotrypsin cleaves proteins on the carboxyl side of aromatic or large hydrophobic amino acids (shaded orange). The likely bonds cleaved by chymotrypsin are indicated in red.

What is the nucleophile that chymotrypsin employs to attack the substrate carbonyl carbon atom? A clue came from the fact that chymotrypsin contains an extraordinarily reactive serine residue. Chymotrypsin molecules treated with organofluorophosphates such as diisopropylphosphofluoridate (DIPF) lost all activity irreversibly (Figure 9.2). Only a single residue, serine 195, was modified. This *chemical modification reaction* suggested that this unusually reactive serine residue plays a central role in the catalytic mechanism of chymotrypsin.

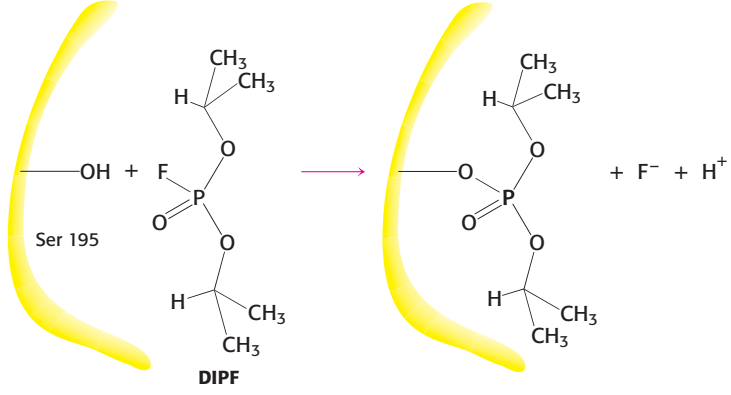


Figure 9.2 An unusually reactive serine residue in chymotrypsin. Chymotrypsin is inactivated by treatment with diisopropylphosphofluoridate (DIPF), which reacts only with serine 195 among 28 possible serine residues.

Chymotrypsin action proceeds in two steps linked by a covalently bound intermediate

A study of the enzyme's kinetics provided a second clue to chymotrypsin's catalytic mechanism. The kinetics of enzyme action are often easily monitored by having the enzyme act on a substrate analog that forms a colored product. For chymotrypsin, such a *chromogenic substrate* is *N*-acetyl-L-phenylalanine *p*-nitrophenyl ester. This substrate is an ester rather than an amide, but many proteases will also hydrolyze esters. One of the products formed by chymotrypsin's cleavage of this substrate is *p*-nitrophenolate, which has a yellow color (Figure 9.3). Measurements of the absorbance of light revealed the amount of *p*-nitrophenolate being produced.

Under steady-state conditions, the cleavage of this substrate obeys Michaelis–Menten kinetics with a K_M of 20 μM and a k_{cat} of 77 s^{-1}. The initial phase of the reaction was examined by using the stopped-flow method, which makes it possible to mix enzyme and substrate and monitor the results within a millisecond. This method revealed an initial rapid burst of colored product, followed by its slower formation as the reaction reached the steady state (Figure 9.4). These results suggest that hydrolysis proceeds

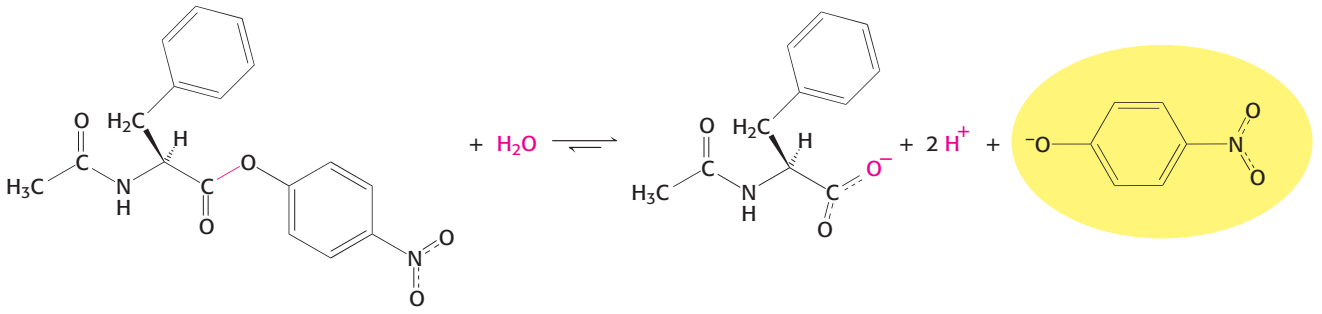


Figure 9.3 Chromogenic substrate. *N*-Acetyl-L-phenylalanine *p*-nitrophenyl ester yields a yellow product, *p*-nitrophenolate, on cleavage by chymotrypsin. *p*-Nitrophenolate forms by deprotonation of *p*-nitrophenol at pH 7.

in two phases. In the first reaction cycle that takes place immediately after mixing, only the first phase must take place before the colored product is released. In subsequent reaction cycles, both phases must take place. Note that the burst is observed because the first phase is substantially more rapid than the second phase for this substrate.

The two phases are explained by the formation of a covalently bound enzyme–substrate intermediate (Figure 9.5). First, the acyl group of the substrate becomes covalently attached to the enzyme as *p*-nitrophenolate (or an amine if the substrate is an amide rather than an ester) is released. The enzyme–acyl group complex is called the *acyl-enzyme intermediate*. Second, the acyl-enzyme intermediate is hydrolyzed to release the carboxylic acid component of the substrate and regenerate the free enzyme. Thus, one molecule of *p*-nitrophenolate is produced rapidly from each enzyme molecule as the acyl-enzyme intermediate is formed. However, it takes longer for the enzyme to be "reset" by the hydrolysis of the acyl-enzyme intermediate, and both phases are required for enzyme turnover.

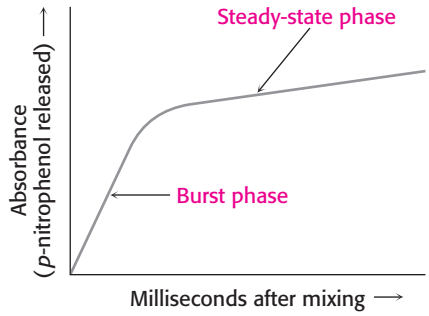


Figure 9.4 Kinetics of chymotrypsin catalysis. Two phases are evident in the cleaving of *N*-acetyl-L-phenylalanine *p*-nitrophenyl ester by chymotrypsin: a rapid burst phase (pre-steady-state) and a steady-state phase.

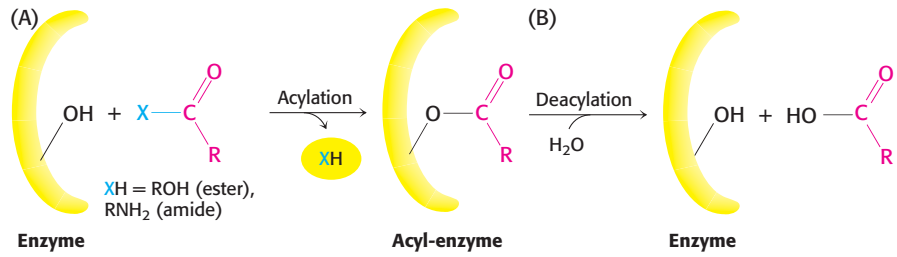


Figure 9.5 Covalent catalysis. Hydrolysis by chymotrypsin takes place in two phases: (A) acylation to form the acyl-enzyme intermediate followed by (B) deacylation to regenerate the free enzyme.

Serine is part of a catalytic triad that also includes histidine and aspartate

The three-dimensional structure of chymotrypsin was solved by David Blow in 1967. Overall, chymotrypsin is roughly spherical and comprises three polypeptide chains, linked by disulfide bonds. It is synthesized as a single polypeptide, termed *chymotrypsinogen*, which is activated by the proteolytic cleavage of the polypeptide to yield the three chains (Section 10.4). The active site of chymotrypsin, marked by serine 195, lies in a cleft on the surface of the enzyme (Figure 9.6). The structure of the active

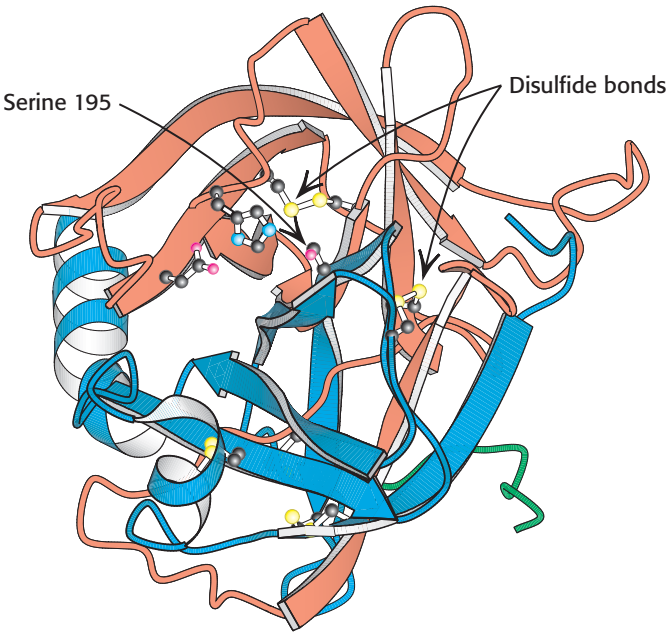


Figure 9.6 Location of the active site in chymotrypsin. Chymotrypsin consists of three chains, shown in ribbon form in orange, blue, and green. The side chains of the catalytic triad residues are shown as ball-and-stick representations. *Notice* these side chains, including serine 195, lining the active site in the upper half of the structure. *Also notice* two intrastrand and two interstrand disulfide bonds in various locations throughout the molecule. [Drawn from 1GCT.pdb.]

site explained the special reactivity of serine 195 (Figure 9.7). The side chain of serine 195 is hydrogen bonded to the imidazole ring of histidine 57. The —NH group of this imidazole ring is, in turn, hydrogen bonded to the carboxylate group of aspartate 102. This constellation of residues is referred to as the *catalytic triad.* How does this arrangement of residues lead to the high reactivity of serine 195? The histidine residue serves to position the serine side chain and to polarize its hydroxyl group so that it is poised for deprotonation. In the presence of the substrate, the histidine residue accepts the proton from the serine 195 hydroxyl group. In doing so, the residue acts as a general base catalyst. The withdrawal of the proton from the hydroxyl group generates an alkoxide ion, which is a much more powerful nucleophile than is an alcohol. The aspartate residue helps orient the histidine residue and make it a better proton acceptor through hydrogen bonding and electrostatic effects.

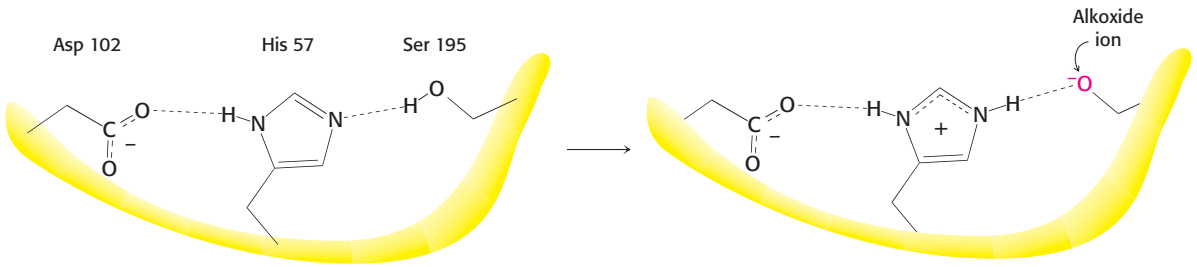


Figure 9.7 The catalytic triad. The catalytic triad, shown on the left, converts serine 195 into a potent nucleophile, as illustrated on the right.

These observations suggest a mechanism for peptide hydrolysis (Figure 9.8). After substrate binding (step 1), the reaction begins with the oxygen atom of the side chain of serine 195 making a nucleophilic attack on the carbonyl carbon atom of the target peptide bond (step 2). There are now four atoms bonded to the carbonyl carbon, arranged as a tetrahedron, instead of three atoms in a planar arrangement. This inherently unstable *tetrahedral intermediate* bears a formal negative charge on the oxygen atom derived from the carbonyl group. This charge is stabilized by interactions

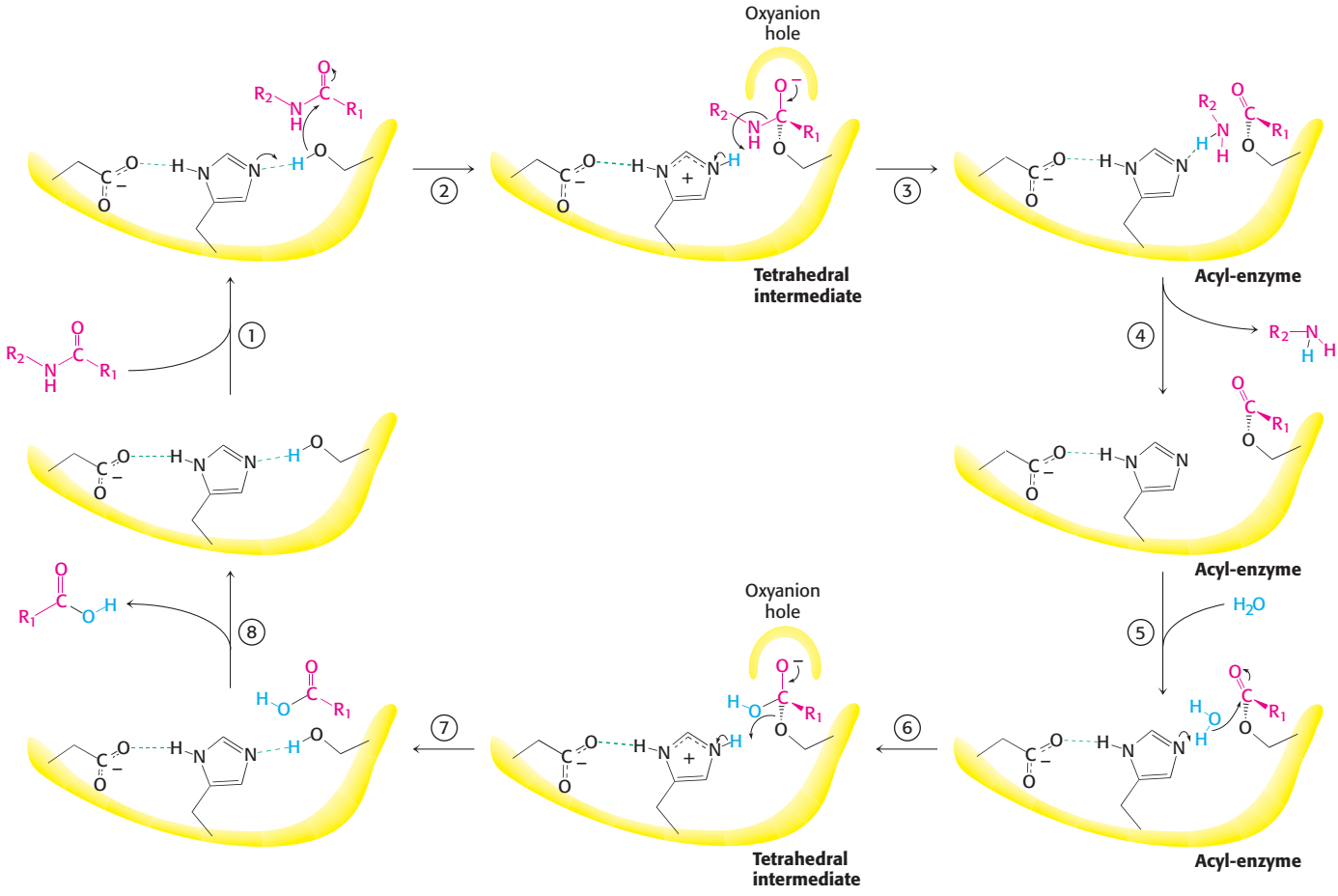


Figure 9.8 Peptide hydrolysis by chymotrypsin. The mechanism of peptide hydrolysis illustrates the principles of covalent and acid–base catalysis. The reaction proceeds in eight steps: (1) substrate binding, (2) nucleophilic attack of serine on the peptide carbonyl group, (3) collapse of the tetrahedral intermediate, (4) release of the amine component, (5) water binding, (6) nucleophilic attack of water on the acyl-enzyme intermediate, (7) collapse of the tetrahedral intermediate; and (8) release of the carboxylic acid component. The dashed green lines represent hydrogen bonds.

with NH groups from the protein in a site termed the *oxyanion hole* (Figure 9.9). These interactions also help stabilize the transition state that precedes the formation of the tetrahedral intermediate. This tetrahedral intermediate collapses to generate the acyl-enzyme (step 3). This step is facilitated by the transfer of the proton being held by the positively charged histidine residue to the amino group formed by cleavage of the peptide bond. The amine component is now free to depart from the enzyme (step 4), completing the first stage of the hydrolytic reaction—acylation of the enzyme.

The next stage—deacylation—begins when a water molecule takes the place occupied earlier by the amine component of the substrate (step 5). The ester group of the acyl-enzyme is now hydrolyzed by a process that essentially repeats steps 2 through 4. Now acting as a general acid catalyst, histidine 57 draws a proton away from the water molecule. The resulting OH^- ion attacks the carbonyl carbon atom of the acyl group, forming a tetrahedral intermediate (step 6). This structure breaks down to form the carboxylic acid product (step 7). Finally, the release of the carboxylic acid product (step 8) readies the enzyme for another round of catalysis.

This mechanism accounts for all characteristics of chymotrypsin action except the observed preference for cleaving the peptide bonds just past

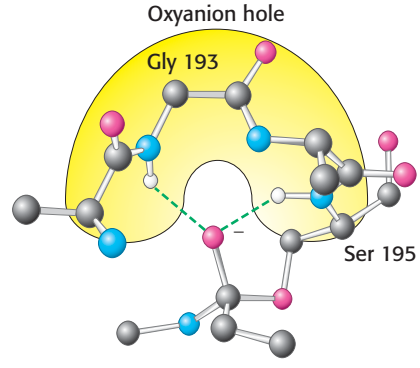


Figure 9.9 The oxyanion hole. The structure stabilizes the tetrahedral intermediate of the chymotrypsin reaction. *Notice* that hydrogen bonds (shown in green) link peptide NH groups and the negatively charged oxygen atom of the intermediate.

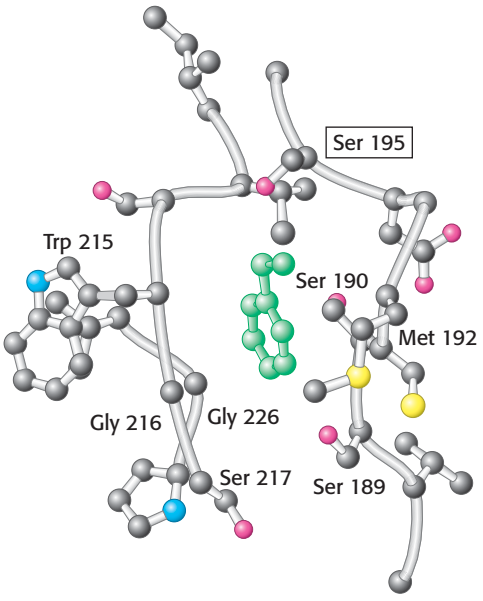


Figure 9.10 Specificity pocket of chymotrypsin. *Notice* that this pocket is lined with hydrophobic residues and is deep, favoring the binding of residues with long hydrophobic side chains such as phenylalanine (shown in green). *Also notice* that the active-site serine residue (serine 195) is positioned to cleave the peptide backbone between the residue bound in the pocket and the next residue in the sequence. The key amino acids that constitute the binding site are identified.

residues with large, hydrophobic side chains. Examination of the three-dimensional structure of chymotrypsin with substrate analogs and enzyme inhibitors revealed the presence of a deep hydrophobic pocket, called the S_1 pocket, into which the long, uncharged side chains of residues such as phenylalanine and tryptophan can fit. *The binding of an appropriate side chain into this pocket positions the adjacent peptide bond into the active site for cleavage* (Figure 9.10). The specificity of chymotrypsin depends almost entirely on which amino acid is directly on the amino-terminal side of the peptide bond to be cleaved. Other proteases have more-complex specificity patterns. Such enzymes have additional pockets on their surfaces for the recognition of other residues in the substrate. Residues on the amino-terminal side of the scissile bond (the bond to be cleaved) are labeled P_1, P_2, P_3, and so forth, heading away from the scissile bond (Figure 9.11). Likewise, residues on the carboxyl side of the scissile bond are labeled P_1', P_2', P_3', and so forth. The corresponding sites on the enzyme are referred to as S_1, S_2 or S_1', S_2', and so forth.

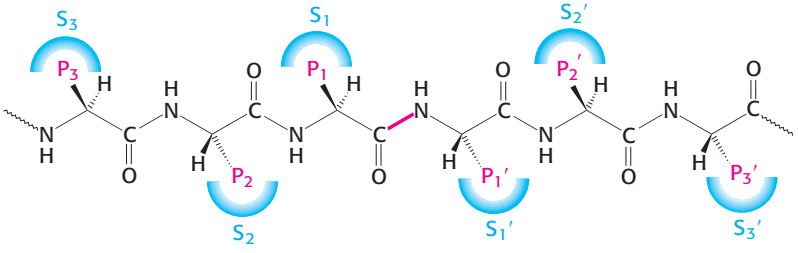


Figure 9.11 Specificity nomenclature for protease–substrate interactions. The potential sites of interaction of the substrate with the enzyme are designated P (shown in red), and corresponding binding sites on the enzyme are designated S. The scissile bond (also shown in red) is the reference point.

Catalytic triads are found in other hydrolytic enzymes

Many other peptide-cleaving proteins have subsequently been found to contain catalytic triads similar to that discovered in chymotrypsin. Some, such as trypsin and elastase, are obvious homologs of chymotrypsin. The sequences of these proteins are approximately 40% identical with that of chymotrypsin, and their overall structures are quite similar (Figure 9.12). These proteins operate by mechanisms identical with that of chymotrypsin.

However, the three enzymes differ markedly in substrate specificity. Chymotrypsin cleaves at the peptide bond after residues with an aromatic or long nonpolar side chain. Trypsin cleaves at the peptide bond after residues with long, positively charged side chains—namely, arginine and lysine. Elastase cleaves at the peptide bond after amino acids with small side chains—such as alanine and serine. Comparison of the S_1 pockets of these enzymes reveals that *these different specificities are due to small structural differences*. In trypsin, an aspartate residue (Asp 189) is present at the bottom of the S_1 pocket in place of a serine residue in chymotrypsin. The aspartate residue attracts and stabilizes a positively charged arginine or lysine residue in the substrate. In elastase, two residues at the top of the pocket in chymotrypsin and trypsin are replaced by much bulkier valine residues (Val 190 and Val 216). These residues close off the mouth of the pocket so that only small side chains can enter (Figure 9.13).

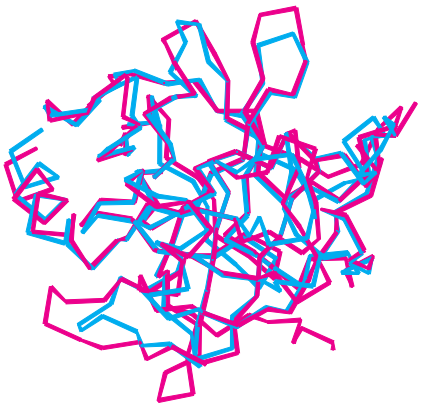


Figure 9.12 Structural similarity of trypsin and chymotrypsin. An overlay of the structure of chymotrypsin (red) on that of trypsin (blue) is shown. *Notice* the high degree of similarity. Only α-carbon-atom positions are shown. The mean deviation in position between corresponding α-carbon atoms is 1.7 Å. [Drawn from 5PTP.pdb and 1GCT.pdb.]

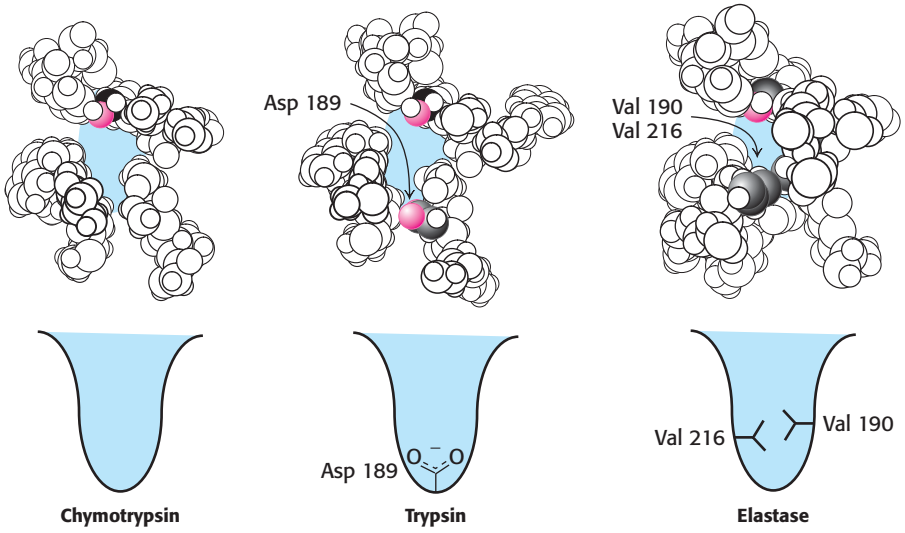


Figure 9.13 The S_1 pockets of chymotrypsin, trypsin, and elastase. Certain residues play key roles in determining the specificity of these enzymes. The side chains of these residues, as well as those of the active-site serine residues, are shown in color.

Other members of the chymotrypsin family include a collection of proteins that take part in blood clotting, to be discussed in Chapter 10, as well as the tumor marker protein prostate-specific antigen (PSA). In addition, a wide range of proteases found in bacteria, viruses, and plants belong to this clan.

Other enzymes that are not homologs of chymotrypsin have been found to contain very similar active sites. As noted in Chapter 6, the presence of very similar active sites in these different protein families is a consequence of convergent evolution. Subtilisin, a protease in bacteria such as *Bacillus amyloliquefaciens,* is a particularly well characterized example. The active site of this enzyme includes both the catalytic triad and the oxyanion hole. However, one of the NH groups that forms the oxyanion hole comes from the side chain of an asparagine residue rather than from the peptide backbone (Figure 9.14). Subtilisin is the founding member of another large family of proteases that includes representatives from Archaea, Bacteria, and Eukarya.

Finally, other proteases have been discovered that contain an active-site serine or threonine residue that is activated not by a histidine–aspartate pair

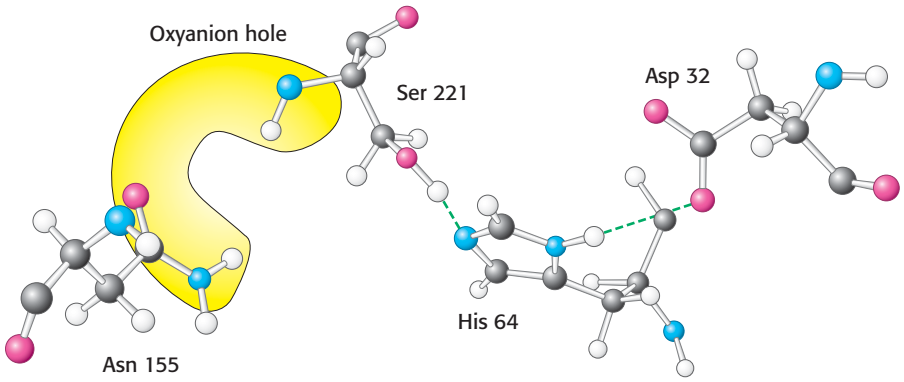


Figure 9.14 The catalytic triad and oxyanion hole of subtilisin. *Notice* the two enzyme NH groups (both in the backbone and in the side chain of Asn 155) located in the oxyanion hole. The NH groups will stabilize a negative charge that develops on the peptide bond attacked by nucleophilic serine 221 of the catalytic triad.

but by a primary amino group from the side chain of lysine or by the N-terminal amino group of the polypeptide chain.

Thus, the catalytic triad in proteases has emerged at least three times in the course of evolution. We can conclude that this catalytic strategy must be an especially effective approach to the hydrolysis of peptides and related bonds.

The catalytic triad has been dissected by site-directed mutagenesis

How can we be sure that the mechanism proposed for the catalytic triad is correct? One way is to test the contribution of individual amino acid residues to the catalytic power of a protease by using site-directed mutagenesis (Section 5.2). Subtilisin has been extensively studied by this method. Each of the residues within the catalytic triad, consisting of aspartic acid 32, histidine 64, and serine 221, has been individually converted into alanine, and the ability of each mutant enzyme to cleave a model substrate has been examined (Figure 9.15).

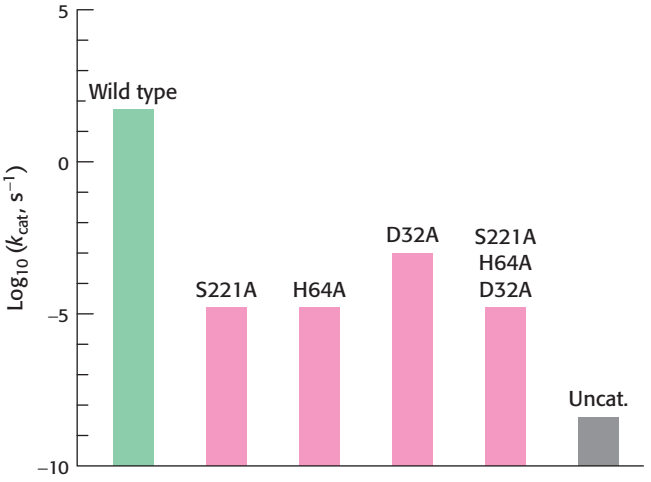


Figure 9.15 Site-directed mutagenesis of subtilisin. Residues of the catalytic triad were mutated to alanine, and the activity of the mutated enzyme was measured. Mutations in any component of the catalytic triad cause a dramatic loss of enzyme activity. Note that the activity is displayed on a logarithmic scale. The mutations are identified as follows: the first letter is the one-letter abbreviation for the amino acid being altered; the number identifies the position of the residue in the primary structure; and the second letter is the one-letter abbreviation for the amino acid replacing the original one. Uncat. refers to the estimated rate for the uncatalyzed reaction.

As expected, the conversion of active-site serine 221 into alanine dramatically reduced catalytic power; the value of k_{cat} fell to less than *one-millionth* of its value for the wild-type enzyme. The value of K_M was essentially unchanged; its increase by no more than a factor of two indicated that substrate continued to bind normally. The mutation of histidine 64 to alanine reduced catalytic power to a similar degree. The conversion of aspartate 32 into alanine reduced catalytic power by less, although the value of k_{cat} still fell to less than 0.005% of its wild-type value. The simultaneous conversion of all three residues into alanine was no more deleterious than the conversion of serine or histidine alone. These observations support the notion that the catalytic triad and, particularly, the serine–histidine pair act together to generate a nucleophile of sufficient power to attack the carbonyl carbon atom of a peptide bond. Despite the reduction in their catalytic power, the mutated enzymes still hydrolyze peptides a thousand times as fast as buffer at pH 8.6.

Site-directed mutagenesis also offered a way to probe the importance of the oxyanion hole for catalysis. The mutation of asparagine 155 to glycine eliminated the side-chain NH group from the oxyanion hole of subtilisin. The elimination of the NH group reduced the value of k_{cat} to 0.2% of its wild-type value but increased the value of K_M by only a factor of two. These observations

demonstrate that the NH group of the asparagine residue plays a significant role in stabilizing the tetrahedral intermediate and the transition state leading to it.

Cysteine, aspartyl, and metalloproteases are other major classes of peptide-cleaving enzymes

Not all proteases utilize strategies based on activated serine residues. Classes of proteins have been discovered that employ three alternative approaches to peptide-bond hydrolysis (Figure 9.16). These classes are the (1) cysteine proteases, (2) aspartyl proteases, and (3) metalloproteases. In each case, the strategy is to generate a nucleophile that attacks the peptide carbonyl group (Figure 9.17).

The strategy used by the *cysteine proteases* is most similar to that used by the chymotrypsin family. In these enzymes, a cysteine residue, activated by a histidine residue, plays the role of the nucleophile that attacks the peptide bond (see Figure 9.17) in a manner quite analogous to that of the serine residue in serine proteases. Because the sulfur atom in cysteine is inherently a better nucleophile than is the oxygen atom in serine, cysteine proteases appear to require only this histidine residue in addition to cysteine and not the full catalytic triad. A well-studied example of these proteins is papain,

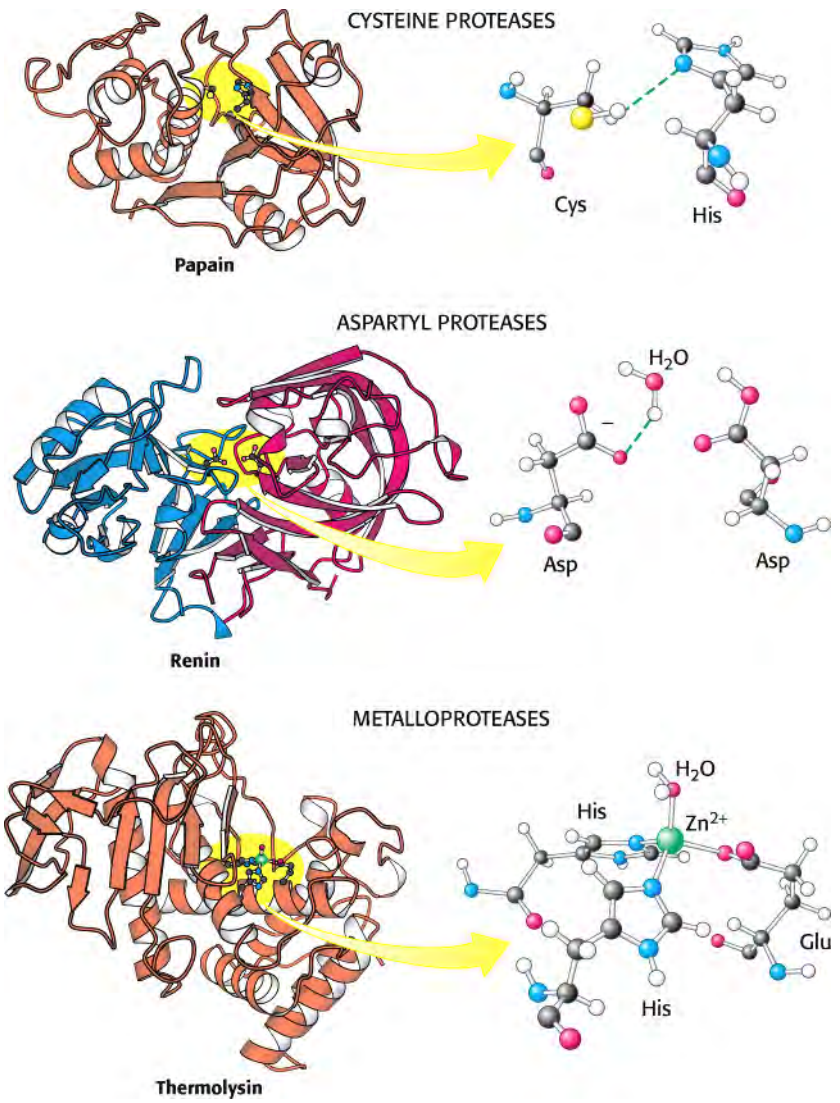


Figure 9.16 Three classes of proteases and their active sites. These examples of a cysteine protease, an aspartyl protease, and a metalloprotease use a histidine-activated cysteine residue, an aspartate-activated water molecule, and a metal-activated water molecule, respectively, as the nucleophile. The two halves of renin are in blue and red to highlight the approximate twofold symmetry of aspartyl proteases. *Notice* how different these active sites are despite the similarity in the reactions they catalyze. [Drawn from 1PPN.pdb.; 1HRN.pdb; 1LND.pdb.]

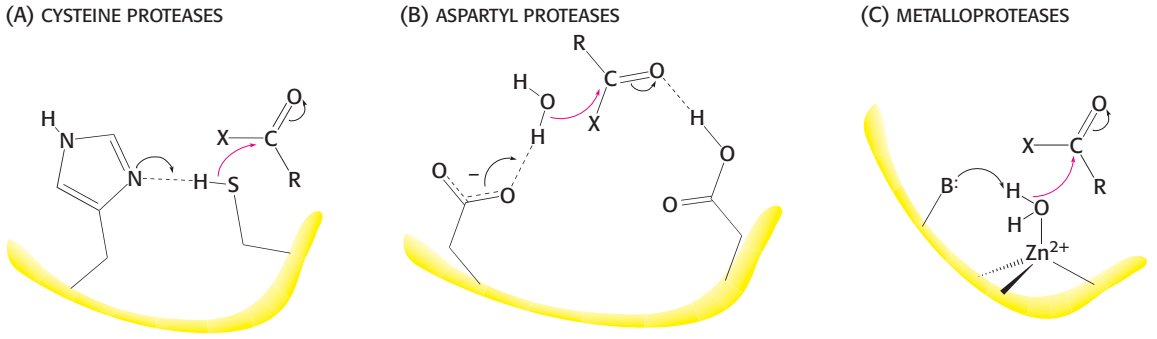


Figure 9.17 The activation strategies for three classes of proteases. The peptide carbonyl group is attacked by (A) a histidine-activated cysteine in the cysteine proteases, (B) an aspartate-activated water molecule in the aspartyl proteases, and (C) a metal-activated water molecule in the metalloproteases. For the metalloproteases, the letter B represents a base (often glutamate) that helps deprotonate the metal-bound water.

an enzyme purified from the fruit of the papaya. Mammalian proteases homologous to papain have been discovered, most notably the cathepsins, proteins having a role in the immune system and other systems. The cysteine-based active site arose independently at least twice in the course of evolution; the caspases, enzymes that play a major role in apoptosis, have active sites similar to that of papain, but their overall structures are unrelated.

The second class comprises the *aspartyl proteases*. The central feature of the active sites is a pair of aspartic acid residues that act together to allow a water molecule to attack the peptide bond. One aspartic acid residue (in its deprotonated form) activates the attacking water molecule by poising it for deprotonation. The other aspartic acid residue (in its protonated form) polarizes the peptide carbonyl group so that it is more susceptible to attack (see Figure 9.17). Members of this class include renin, an enzyme having a role in the regulation of blood pressure, and the digestive enzyme pepsin. These proteins possess approximate twofold symmetry. A likely scenario is that two copies of a gene for the ancestral enzyme fused to form a single gene that encoded a single-chain enzyme. Each copy of the gene would have contributed an aspartate residue to the active site. The individual chains are now joined to make a single chain in the aspartyl proteases present in human immunodeficiency virus (HIV) and other retroviruses (Figure 9.18). This observation is consistent with the idea that the enzyme may have originally existed as separate subunits.

The *metalloproteases* constitute the final major class of peptide-cleaving enzymes. The active site of such a protein contains a bound metal ion, almost always zinc, that activates a water molecule to act as a nucleophile to attack the peptide carbonyl group. The bacterial enzyme thermolysin and the digestive enzyme carboxypeptidase A are classic examples of the zinc proteases. Thermolysin, but not carboxypeptidase A, is a member of a large and diverse family of homologous zinc proteases that includes the matrix metalloproteases, enzymes that catalyze the reactions in tissue remodeling and degradation.

In each of these three classes of enzymes, the active site includes features that act to (1) activate a water molecule or another nucleophile, (2) polarize the peptide carbonyl group, and (3) stabilize a tetrahedral intermediate (see Figure 9.17).

Protease inhibitors are important drugs

Several important drugs are protease inhibitors. For example, captopril, used to regulate blood pressure, is an inhibitor of the angiotensin-converting enzyme (ACE), a metalloprotease. Indinavir (Crixivan), retrovir, and more than 20 other compounds used in the treatment of AIDS are inhibitors of HIV protease, which is an aspartyl protease. HIV protease

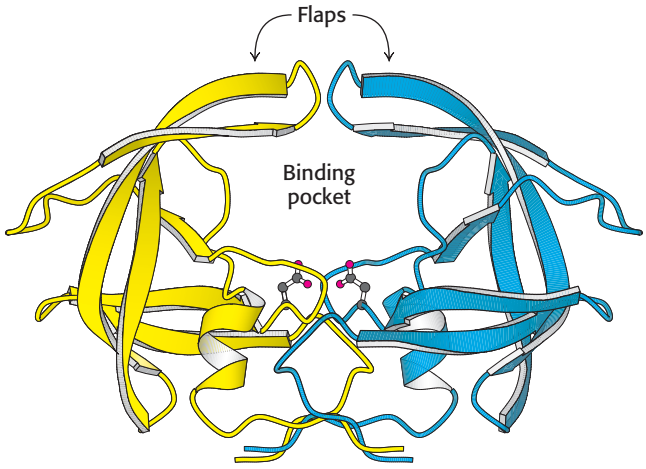


Figure 9.18 HIV protease, a dimeric aspartyl protease. The protease is a dimer of identical subunits, shown in blue and yellow, consisting of 99 amino acids each. *Notice* the placement of active-site aspartic acid residues, one from each chain, which are shown as ball-and-stick structures. The flaps will close down on the binding pocket after substrate has been bound. [Drawn from 3PHV.pdb.]

cleaves multidomain viral proteins into their active forms; blocking this process completely prevents the virus from being infectious (see Figure 9.18). HIV protease inhibitors, in combination with inhibitors of other key HIV enzymes, dramatically reduced deaths due to AIDS in circumstances where these drugs can be used (see Figure 36.21).

Indinavir resembles the peptide substrate of the HIV protease. Indinavir is constructed around an alcohol that mimics the tetrahedral intermediate; other groups are present to bind into the S_2, S_1, S_1', and S_2' recognition sites on the enzyme (Figure 9.19). X-ray crystallographic studies revealed that, in the active site, indinavir adopts a conformation that approximates the twofold symmetry of the enzyme (Figure 9.20). The active site of HIV protease is covered by two flexible flaps that fold down on top of the bound inhibitor. The OH group of the central alcohol interacts with the two aspartate residues of the active site. In addition, two carbonyl groups of the inhibitor are hydrogen bonded to a water molecule (not shown in Figure 9.20), which, in turn, is hydrogen bonded to a peptide NH group in each of the flaps. This interaction of the inhibitor with water and the enzyme is not possible within cellular aspartyl proteases such as renin. Thus the interaction may contribute to the specificity of indinavir for HIV protease.

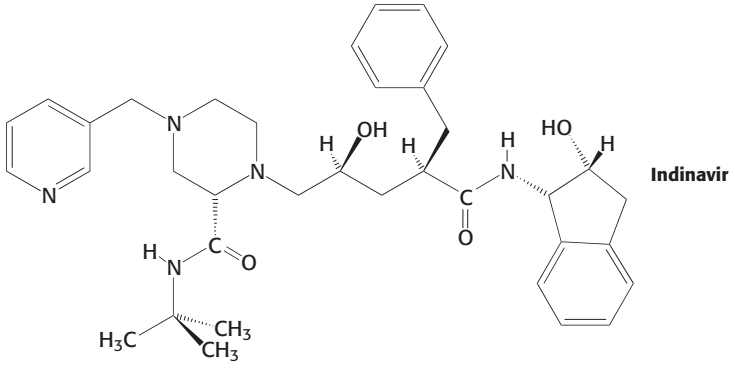


Figure 9.19 Indinavir, an HIV protease inhibitor. The structure of indinavir (Crixivan) is shown in comparison with that of a peptide substrate of HIV protease. The scissile bond in the substrate is highlighted in red.

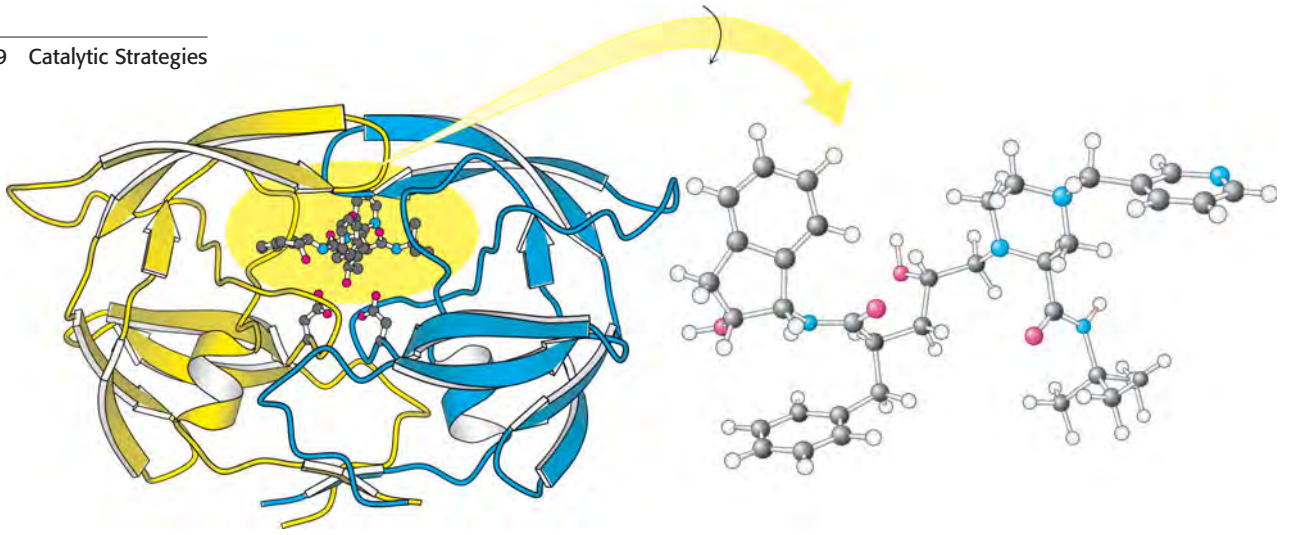
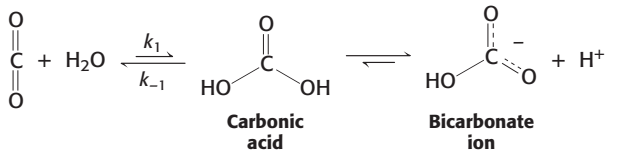


Figure 9.20 HIV protease–indinavir complex. (Left) The HIV protease is shown with the inhibitor indinavir bound at the active site. *Notice* the twofold symmetry of the enzyme structure. (Right) The drug has been rotated to reveal its approximately twofold symmetric conformation. [Drawn from 1HSH.pdb.]

Protease inhibitors used as drugs must be specific for one enzyme without inhibiting other proteins within the body to prevent side effects.

9.2 Carbonic Anhydrases Make a Fast Reaction Faster

Carbon dioxide is a major end product of aerobic metabolism. In mammals, this carbon dioxide is released into the blood and transported to the lungs for exhalation. While in the red blood cells, carbon dioxide reacts with water (Section 7.3). The product of this reaction is a moderately strong acid, carbonic acid ($pK_a = 3.5$), which is converted into bicarbonate ion (HCO_3^-) on the loss of a proton.



Even in the absence of a catalyst, this hydration reaction proceeds at a moderately fast pace. At 37°C near neutral pH, the second-order rate constant k_1 is 0.0027 M^{-1} s^{-1}. This value corresponds to an effective first-order rate constant of 0.15 s^{-1} in water ($[H_2O] = 55.5$ M). The reverse reaction, the dehydration of HCO_3^-, is even more rapid, with a rate constant of $k_{-1} = 50$ s^{-1}. These rate constants correspond to an equilibrium constant of $K_1 = 5.4 \times 10^{-5}$ and a ratio of $[CO_2]$ to $[H_2CO_3]$ of 340 : 1 at equilibrium.

Carbon dioxide hydration and HCO_3^- dehydration are often coupled to rapid processes, particularly transport processes. Thus, almost all organisms contain enzymes, referred to as *carbonic anhydrases,* that increase the rate of reaction beyond the already reasonable spontaneous rate. For example, carbonic anhydrases dehydrate HCO_3^- in the blood to form CO_2 for exhalation as the blood passes through the lungs. Conversely, they convert CO_2 into HCO_3^- to generate the aqueous humor of the eye and other secretions. Furthermore, both CO_2 and HCO_3^- are substrates and products

for a variety of enzymes, and the rapid interconversion of these species may be necessary to ensure appropriate substrate levels. So important are these enzymes in human beings that mutations in some carbonic anhydrases have been found to be associated with osteopetrosis (excessive formation of dense bones accompanied by anemia) and mental retardation.

Carbonic anhydrases accelerate CO_2 hydration dramatically. The most-active enzymes hydrate CO_2 at rates as high as $k_{cat} = 10^6\ s^{-1}$, or a million times a second per enzyme molecule. Fundamental physical processes such as diffusion and proton transfer ordinarily limit the rate of hydration, and so the enzymes employ special strategies to attain such prodigious rates.

Carbonic anhydrase contains a bound zinc ion essential for catalytic activity

Less than 10 years after the discovery of carbonic anhydrase in 1932, this enzyme was found to contain a bound zinc ion. Moreover, the zinc ion appeared to be necessary for catalytic activity. This discovery, remarkable at the time, made carbonic anhydrase the first known zinc-containing enzyme. At present, hundreds of enzymes are known to contain zinc. In fact, more than one-third of all enzymes either contain bound metal ions or require the addition of such ions for activity. Metal ions have several properties that increase chemical reactivity: their positive charges, their ability to form strong yet kinetically labile bonds, and, in some cases, their capacity to be stable in more than one oxidation state. The chemical reactivity of metal ions explains why catalytic strategies that employ metal ions have been adopted throughout evolution.

X-ray crystallographic studies have supplied the most-detailed and direct information about the zinc site in carbonic anhydrase. At least seven carbonic anhydrases, each with its own gene, are present in human beings. They are all clearly homologous, as revealed by substantial sequence identity. Carbonic anhydrase II, a major protein component of red blood cells, has been the most extensively studied (Figure 9.21). It is also one of the most active carbonic anhydrases.

Zinc is found only in the +2 state in biological systems. A zinc atom is essentially always bound to four or more ligands; in carbonic anhydrase, three coordination sites are occupied by the imidazole rings of three histidine residues and an additional coordination site is occupied by a water

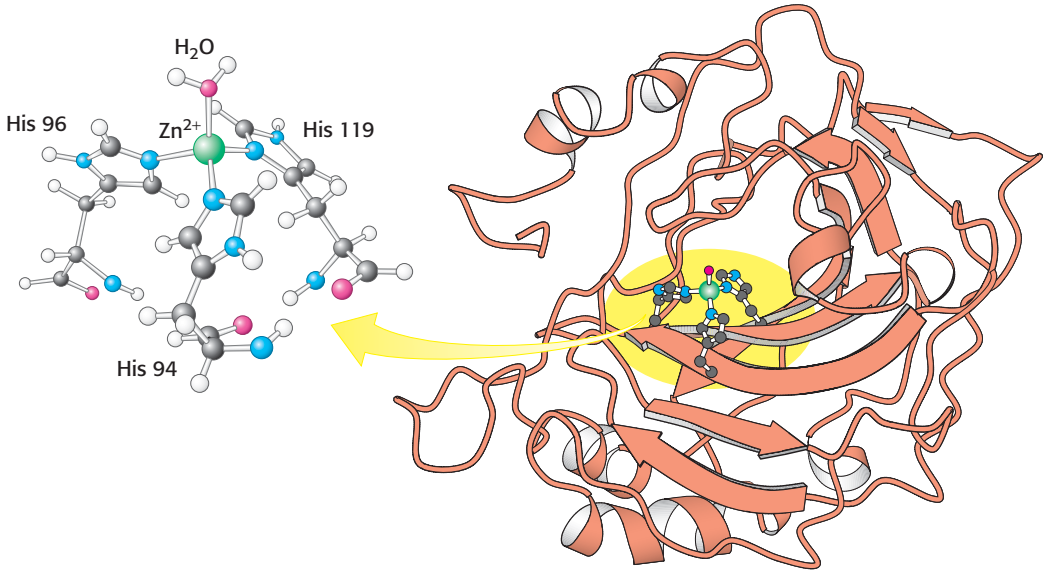


Figure 9.21 The structure of human carbonic anhydrase II and its zinc site. (Left) *Notice* that the zinc ion is bound to the imidazole rings of three histidine residues as well as to a water molecule. (Right) *Notice* the location of the zinc site in a cleft near the center of the enzyme. [Drawn from 1CA2.pdb.]

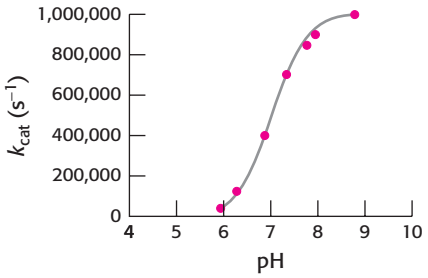


Figure 9.22 Effect of pH on carbonic anhydrase activity. Changes in pH alter the rate of carbon dioxide hydration catalyzed by carbonic anhydrase II. The enzyme is maximally active at high pH.

molecule (or hydroxide ion, depending on pH). Because the molecules occupying the coordination sites are neutral, the overall charge on the $Zn(His)_3$ unit remains +2.

Catalysis entails zinc activation of a water molecule

How does this zinc complex facilitate carbon dioxide hydration? A major clue comes from the pH profile of enzymatically catalyzed carbon dioxide hydration (Figure 9.22).

At pH 8, the reaction proceeds near its maximal rate. As the pH decreases, the rate of the reaction drops. The midpoint of this transition is near pH 7, suggesting that a group that loses a proton at pH 7 ($pK_a = 7$) plays an important role in the activity of carbonic anhydrase. Moreover, the curve suggests that the deprotonated (high pH) form of this group participates more effectively in catalysis. Although some amino acids, notably histidine, have pK_a values near 7, *a variety of evidence suggests that the group responsible for this transition is not an amino acid but is the zinc-bound water molecule.*

The binding of a water molecule to the positively charged zinc center reduces the pK_a of the water molecule from 15.7 to 7 (Figure 9.23).

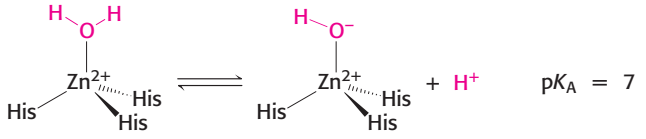


Figure 9.23 The pK_a of zinc-bound water. Binding to zinc lowers the pK_a of water from 15.7 to 7.

With the pK_a lowered, many water molecules lose a proton at neutral pH, generating a substantial concentration of hydroxide ion (bound to the zinc atom). A zinc-bound hydroxide ion (OH^-) is a potent nucleophile able to attack carbon dioxide much more readily than water does. Adjacent to the zinc site, carbonic anhydrase also possesses a hydrophobic patch that serves as a binding site for carbon dioxide (Figure 9.24). Based on these observations, a simple mechanism for carbon dioxide hydration can be proposed (Figure 9.25):

1. The zinc ion facilitates the release of a proton from a water molecule, which generates a hydroxide ion.

2. The carbon dioxide substrate binds to the enzyme's active site and is positioned to react with the hydroxide ion.

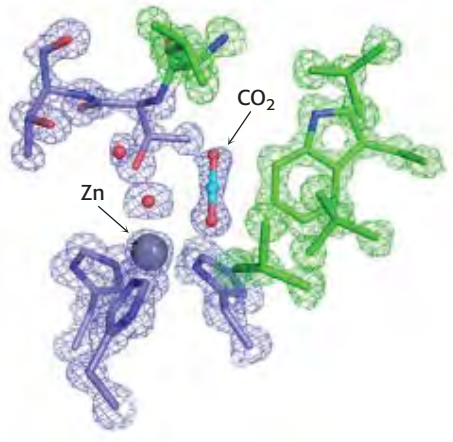


Figure 9.24 Carbon dioxide binding site. Crystals of carbonic anhydrase were exposed to carbon dioxide gas at high pressure and low temperature and x-ray diffraction data were collected. The electron density for carbon dioxide, clearly visible adjacent to the zinc and its bound water, reveals the carbon dioxide binding site. [After J. F. Domsic, B. S. Avvaru, C. U. Kim, S. M. Gruner, M. Agbandje-McKenna, D. N. Silverman, and R. McKenna. *J. Biol. Chem.* 283:30766–30771, 2008.]

3. The hydroxide ion attacks the carbon dioxide, converting it into bicarbonate ion, HCO_3^-.

4. The catalytic site is regenerated with the release of HCO_3^- and the binding of another molecule of water.

Thus, the binding of a water molecule to the zinc ion favors the formation of the transition state by facilitating proton release and by positioning the water molecule to be in close proximity to the other reactant.

Studies of a *synthetic analog model system* provide evidence for the mechanism's plausibility. A simple synthetic ligand binds zinc through four nitrogen atoms (compared with three histidine nitrogen atoms in the enzyme), as shown in Figure 9.26. One water molecule remains bound to the zinc ion in the complex. Direct measurements reveal that this water molecule has a pK_a value of 8.7, not as low as the value for the water molecule in carbonic anhydrase but substantially lower than the value for free water. At pH 9.2, this complex accelerates the hydration of carbon dioxide more than 100-fold. Although its rate of catalysis is much less efficient than catalysis by carbonic anhydrase, the model system strongly suggests that the zinc-bound hydroxide mechanism is likely to be correct. Carbonic anhydrases have evolved to employ the reactivity intrinsic to a zinc-bound hydroxide ion as a potent catalyst.

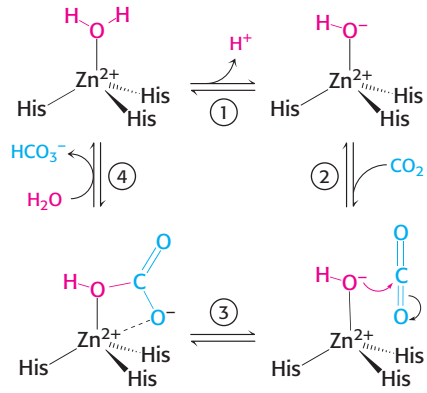


Figure 9.25 Mechanism of carbonic anhydrase. The zinc-bound hydroxide mechanism for the hydration of carbon dioxide reveals one aspect of metal ion catalysis. The reaction proceeds in four steps: (1) water deprotonation; (2) carbon dioxide binding; (3) nucleophilic attack by hydroxide on carbon dioxide; and (4) displacement of bicarbonate ion by water.

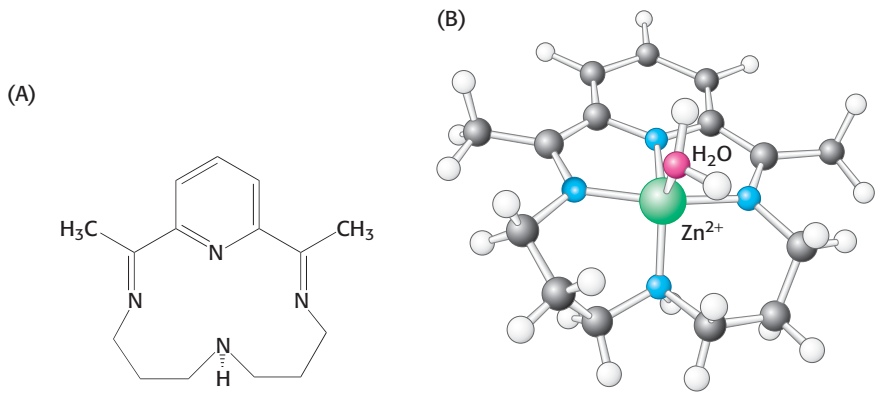


Figure 9.26 A synthetic analog model system for carbonic anhydrase. (A) An organic compound, capable of binding zinc, was synthesized as a model for carbonic anhydrase. The zinc complex of this ligand accelerates the hydration of carbon dioxide more than 100-fold under appropriate conditions. (B) The structure of the presumed active complex showing zinc bound to the ligand and to one water molecule.

A proton shuttle facilitates rapid regeneration of the active form of the enzyme

As noted earlier, some carbonic anhydrases can hydrate carbon dioxide at rates as high as a million times a second (10^6 s^{-1}). The magnitude of this rate can be understood from the following observations. In the first step of a carbon dioxide hydration reaction, the zinc-bound water molecule must lose a proton to regenerate the active form of the enzyme (Figure 9.27). The rate of the reverse reaction, the protonation of the zinc-bound hydroxide ion, is limited by the rate of proton diffusion. Protons diffuse very rapidly with second-order rate constants near 10^{-11} M^{-1} s^{-1}. Thus, the backward

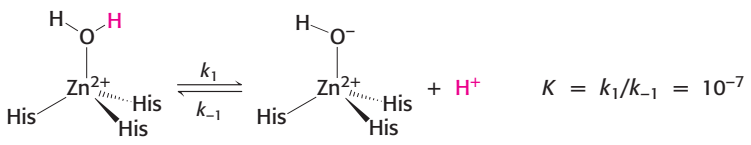
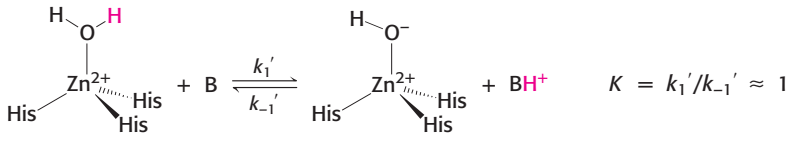


Figure 9.27 Kinetics of water deprotonation. The kinetics of deprotonation and protonation of the zinc-bound water molecule in carbonic anhydrase.

Figure 9.28 The effect of buffer on deprotonation. The deprotonation of the zinc-bound water molecule in carbonic anhydrase is aided by buffer component B.



rate constant k^{-1} must be less than 10^{11} M^{-1} s^{-1}. Because the equilibrium constant K is equal to k_1/k_{-1}, the forward rate constant is given by $k_1 = K \cdot k_{-1}$. Thus, if $k_{-1} \leq 10^{11}$ M^{-1} s^{-1} and $K = 10^{-7}$ M (because $pK_a = 7$), then k_1 must be less than or equal to 10^4 s^{-1}. In other words, the rate of proton diffusion limits the rate of proton release to less than 10^4 s^{-1} for a group with $pK_a = 7$. However, if carbon dioxide is hydrated at a rate of 10^6 s^{-1}, then every step in the mechanism (see Figure 9.25) must take place at least this fast. How is this apparent paradox resolved?

The answer became clear with the realization that *the highest rates of carbon dioxide hydration require the presence of buffer, suggesting that the buffer components participate in the reaction.* The buffer can bind or release protons. The advantage is that, whereas the concentrations of protons and hydroxide ions are limited to 10^{-7} M at neutral pH, the concentration of buffer components can be much higher, of the order of several millimolar. If the buffer component BH^+ has a pK_a of 7 (matching that for the zinc-bound water molecule), then the equilibrium constant for the reaction in Figure 9.28 is 1. The rate of proton abstraction is given by $k_1' \cdot [B]$. The second-order rate constants k_1' and k_{-1}' will be limited by buffer diffusion to values less than approximately 10^9 M^{-1} s^{-1}. Thus, buffer concentrations greater than $[B] = 10^{-3}$ M (1 mM) may be high enough to support carbon dioxide hydration rates of 10^6 M^{-1} s^{-1} because $k_1' \cdot [B] = (10^9\ M^{-1}\ s^{-1}) \cdot (10^{-3}\ M) = 10^6\ s^{-1}$. This prediction is confirmed experimentally (Figure 9.29).

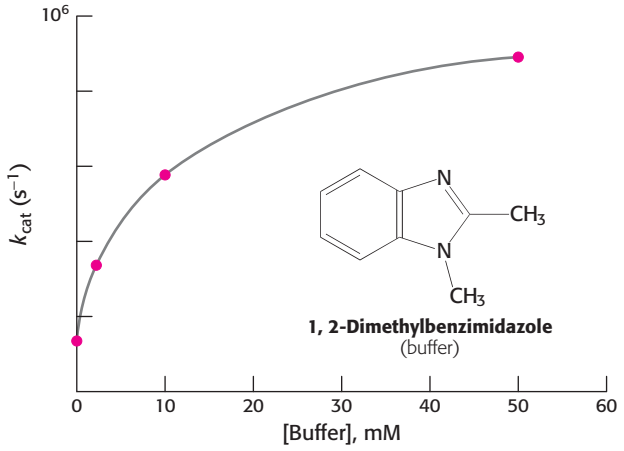
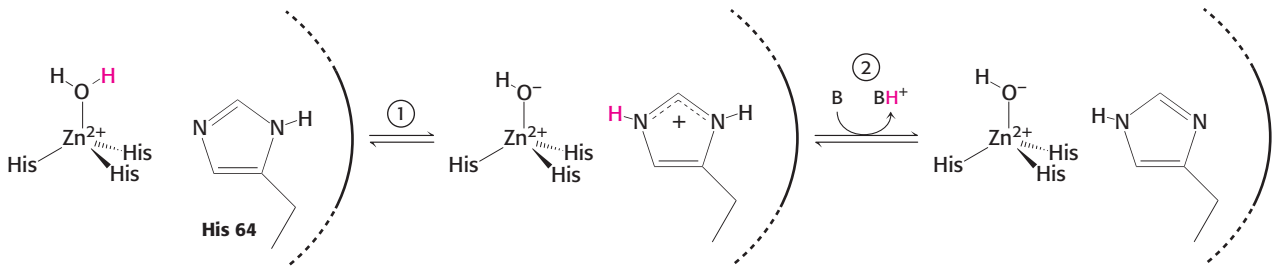


Figure 9.29 The effect of buffer concentration on the rate of carbon dioxide hydration. The rate of carbon dioxide hydration increases with the concentration of the buffer 1,2-dimethylbenzimidazole. The buffer enables the enzyme to achieve its high catalytic rates.

The molecular components of many buffers are too large to reach the active site of carbonic anhydrase. Carbonic anhydrase II has evolved a *proton shuttle* to allow buffer components to participate in the reaction from solution. The primary component of this shuttle is histidine 64. This residue transfers protons from the zinc-bound water molecule to the protein surface and then to the buffer (Figure 9.30). Thus, catalytic function has been enhanced through the evolution of an apparatus for controlling proton transfer from and to the active site. Because protons participate in many biochemical reactions, the manipulation of the proton inventory within active sites is crucial to the function of many enzymes and explains the prominence of acid–base catalysis.

Figure 9.30 Histidine proton shuttle. (1) Histidine 64 abstracts a proton from the zinc-bound water molecule, generating a nucleophilic hydroxide ion and a protonated histidine. (2) The buffer (B) removes a proton from the histidine, regenerating the unprotonated form.



Convergent evolution has generated zinc-based active sites in different carbonic anhydrases

Carbonic anhydrases homologous to the human enzymes, referred to as *α-carbonic anhydrases,* are common in animals and in some bacteria and algae. In addition, two other families of carbonic anhydrases have been discovered. Proteins in these families contain the zinc ion required for catalytic activity but are not significantly similar in sequence to the α-carbonic anhydrases. The *β-carbonic anhydrases* are found in higher plants and in many bacterial species, including *E. coli.* Spectroscopic and structural studies reveal that the zinc ion is bound by one histidine residue and two cysteine residues. Moreover, the overall enzyme structures are unrelated to those of the α-carbonic anhydrases. In plants, these enzymes facilitate the accumulation of carbon dioxide, crucial for the Calvin cycle in photosynthesis. A third family, the *γ-carbonic anhydrases,* was initially identified in the archaeon *Methanosarcina thermophila.* The crystal structure of this enzyme reveals three zinc sites extremely similar to the zinc site in the α-carbonic anhydrases. In this case, however, the three zinc sites lie at the interfaces between the three subunits of a trimeric enzyme (Figure 9.31). The very striking left-handed β-helical structure (a β strand twisted into a left-handed helix) present in this enzyme is, again, different from any structure present in the α- and β-carbonic anhydrases. Thus, convergent evolution has generated carbonic anhydrases that rely on coordinated zinc ions at least three times.

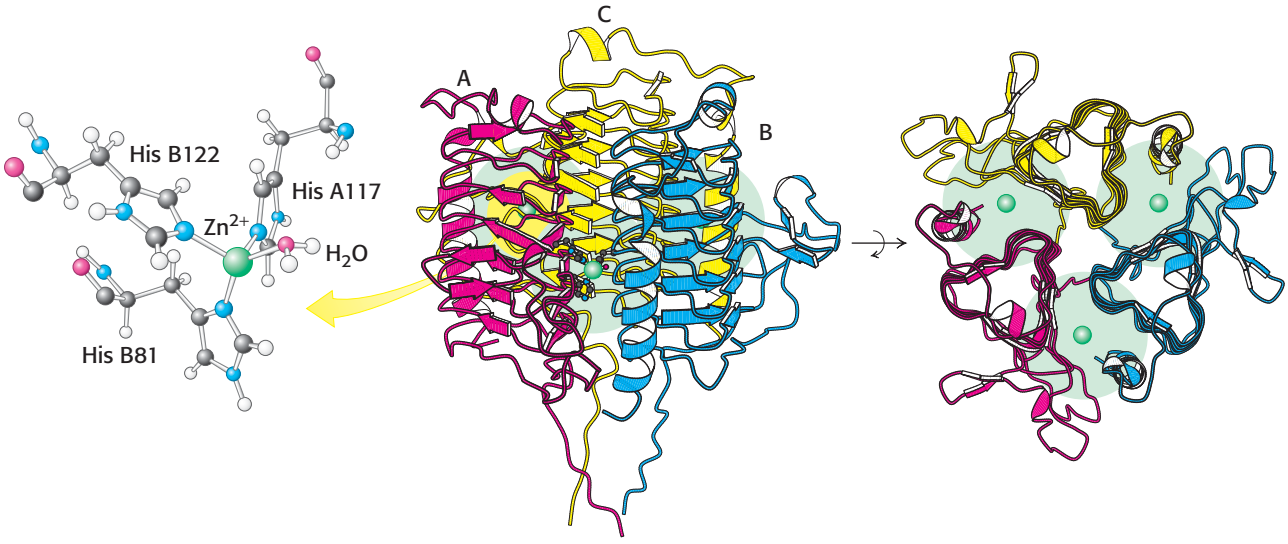


Figure 9.31 γ-Carbonic anhydrase. (Left) The zinc site of γ-carbonic anhydrase. *Notice* that the water-binding zinc ion is bound to three histidine residues. (Middle) The trimeric structure of the protein (individual chains are labeled A, B, and C). Each chain consists primarily of a left-handed β helix. (Right) The protein is rotated to show a top-down view that highlights its threefold symmetry. *Notice* the position of the zinc sites (green) at the interfaces between subunits. [Drawn from 1THJ.pdb.]

9.3 Restriction Enzymes Catalyze Highly Specific DNA-Cleavage Reactions

We next consider a hydrolytic reaction that results in the cleavage of DNA. Bacteria and archaea have evolved mechanisms to protect themselves from viral infections. Many viruses inject their DNA genomes into cells; once inside, the viral DNA hijacks the cell's machinery to drive the production of viral proteins and, eventually, of progeny virus. Often, a viral infection results in the death of the host cell. A major protective strategy for the host is to use *restriction endonucleases* (restriction enzymes) to degrade the viral DNA on its introduction into a cell. These enzymes recognize particular

base sequences, called *recognition sequences* or *recognition sites,* in their target DNA and cleave that DNA at defined positions. We have already considered the utility of these important enzymes for dissecting genes and genomes (Section 5.2). The most well studied class of restriction enzymes comprises the type II restriction enzymes, which cleave DNA *within* their recognition sequences. Other types of restriction enzymes cleave DNA at positions somewhat distant from their recognition sites.

Restriction endonucleases must show tremendous specificity at two levels. First, they must not degrade host DNA containing the recognition sequences. Second, they must cleave only DNA molecules that contain recognition sites (hereafter referred to as *cognate DNA*) without cleaving DNA molecules that lack these sites. How do these enzymes manage to degrade viral DNA while sparing their own? In *E. coli,* the restriction endonuclease *Eco*RV cleaves double-stranded viral DNA molecules that contain the sequence 5′-GATATC-3′ but leaves intact host DNA containing hundreds of such sequences. We shall return to the strategy by which host cells protect their own DNA at the end of this section.

Restriction enzymes must cleave DNA only at recognition sites, without cleaving at other sites. Suppose that a recognition sequence is six base pairs long. Because there are 4^6, or 4096, sequences having six base pairs, the concentration of sites that must not be cleaved will be approximately 4000-fold higher than the concentration of sites that should be cleaved. Thus, to keep from damaging host-cell DNA, restriction enzymes must cleave cognate DNA molecules much more than 4000 times as efficiently as they cleave nonspecific sites. We shall return to the mechanism used to achieve the necessary high specificity after considering the chemistry of the cleavage process.

Cleavage is by in-line displacement of 3′-oxygen from phosphorus by magnesium-activated water

A restriction endonuclease catalyzes the hydrolysis of the phosphodiester backbone of DNA. Specifically, the bond between the 3′-oxygen atom and the phosphorus atom is broken. The products of this reaction are DNA strands with a free 3′-hydroxyl group and a 5′-phosphoryl group at the cleavage site (Figure 9.32). This reaction proceeds by nucleophilic attack at the phosphorus atom. We will consider two alternative mechanisms, suggested by analogy with the proteases. The restriction endonuclease might cleave DNA by mechanism 1 through a covalent intermediate, employing a potent nucleophile (Nu), or by mechanism 2 through direct hydrolysis:

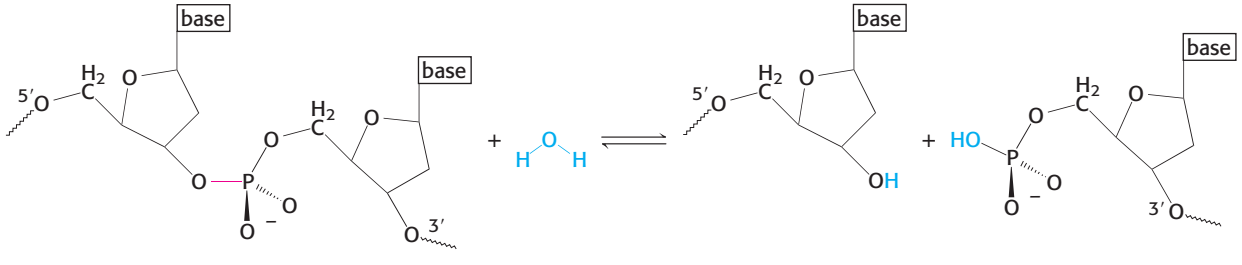
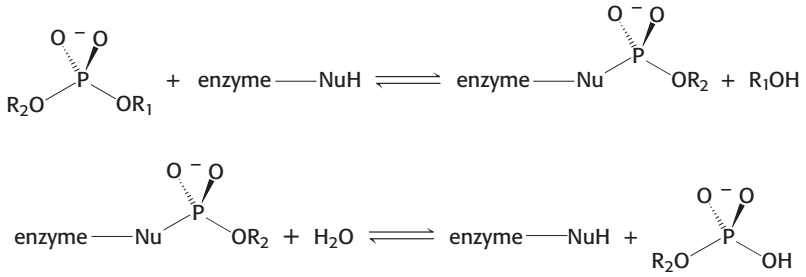
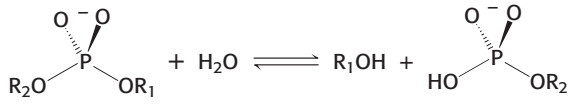


Figure 9.32 Hydrolysis of a phosphodiester bond. All restriction enzymes catalyze the hydrolysis of DNA phosphodiester bonds, leaving a phosphoryl group attached to the 5′ end. The bond that is cleaved is shown in red.

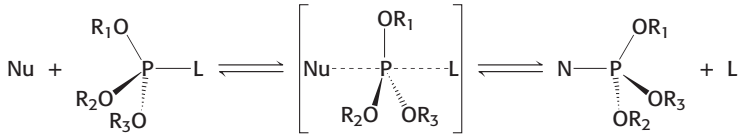
Mechanism 1 (covalent intermediate)



Mechanism 2 (direct hydrolysis)



Each mechanism postulates a different nucleophile to attack the phosphorus atom. In either case, each reaction takes place by *in-line displacement:*



The incoming nucleophile attacks the phosphorus atom, and a pentacoordinate transition state is formed. This species has a trigonal bipyramidal geometry centered at the phosphorus atom, with the incoming nucleophile at one apex of the two pyramids and the group that is displaced (the leaving group, L) at the other apex. Note that the displacement inverts the stereochemical conformation at the tetrahedral phosphorous atom, analogous to the interconversion of the R and S configurations around a tetrahedral carbon center (Section 2.1).

The two mechanisms differ in the number of times that the displacement takes place in the course of the reaction. In the first type of mechanism, a nucleophile in the enzyme (analogous to serine 195 in chymotrypsin) attacks the phosphate group to form a covalent intermediate. In a second step, this intermediate is hydrolyzed to produce the final products. In this case, two displacement reactions take place at the phosphorus atom. Consequently, the stereochemical configuration at the phosphorus atom would be inverted and then inverted again, and the overall configuration would be *retained*. In the second type of mechanism, analogous to that used by the aspartyl- and metalloproteases, an activated water molecule attacks the phosphorus atom directly. In this mechanism, a single displacement reaction takes place at the phosphorus atom. Hence, the stereochemical configuration at the phosphorus atom is *inverted* after cleavage. To determine which mechanism is correct, we examine the stereochemistry at the phosphorus atom after cleavage.

A difficulty is that the stereochemistry is not easily observed, because two of the groups bound to the phosphorus atom are simple oxygen atoms, identical with each other. This difficulty can be circumvented by replacing one oxygen atom with sulfur (producing a species called a phosphorothioate). Let us consider *Eco*RV endonuclease. This enzyme cleaves the phosphodiester bond between the T and the A at the center of the recognition

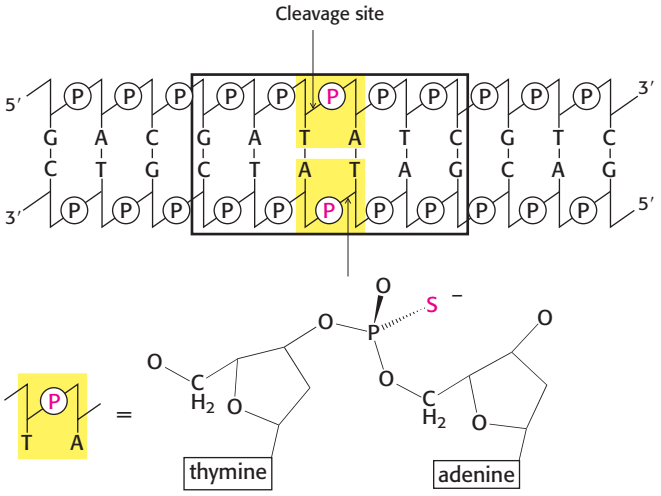


Figure 9.33 Labeling with phosphorothioates. Phosphorothioate groups, in which one of the nonbridging oxygen atoms is replaced by a sulfur atom, can be used to label specific sites in the DNA backbone to determine the overall stereochemical course of a displacement reaction. Here, a phosphorothioate is placed at sites that can be cleaved by *Eco*RV endonuclease.

sequence 5′-GATATC-3′. The first step is to synthesize an appropriate substrate for *Eco*RV containing phosphorothioates at the sites of cleavage (Figure 9.33). The reaction is then performed in water that has been greatly enriched in ^{18}O to allow the incoming oxygen atom to be marked. The location of the ^{18}O label with respect to the sulfur atom indicates whether the reaction proceeds with inversion or retention of stereochemistry. *The analysis revealed that the stereochemical configuration at the phosphorus atom was inverted only once with cleavage.* This result is consistent with a direct attack by water at the phosphorus atom and rules out the formation of any covalently bound intermediate (Figure 9.34).

Restriction enzymes require magnesium for catalytic activity

Many enzymes that act on phosphate-containing substrates require Mg^{2+} or some other similar divalent cation for activity. One or more Mg^{2+} (or similar) cations are essential to the function of restriction endonucleases. What are the functions of these metal ions?

Direct visualization of the complex between *Eco*RV endonuclease and cognate DNA molecules in the presence of Mg^{2+} by crystallization has not been possible, because the enzyme cleaves the substrate under these circumstances. Nonetheless, metal ion complexes can be visualized through several approaches. In one approach, crystals of *Eco*RV endonuclease are prepared bound to oligonucleotides that contain the enzyme's recognition sequence. These crystals are grown in the absence of magnesium to prevent cleavage; after their preparation, the crystals are soaked in solutions containing the metal. Alternatively, crystals have been grown with the use of a mutated form of the enzyme that is less active. Finally, Mg^{2+} can be replaced by metal ions such as Ca^{2+} that bind but do not result in much catalytic activity. In all cases, no cleavage takes place, and so the locations of the metal ion-binding sites are readily determined.

As many as three metal ions have been found to be present per active site. The roles of these multiple metal ions is still under investigation. One ion-binding site is occupied in essentially all structures. This metal ion is coordinated to the protein through two aspartate residues and to one of the phosphate-group oxygen atoms near the site of cleavage. This metal ion binds the water molecule that attacks the phosphorus atom, helping to position and activate it in a manner similar to that for the Zn^{2+} ion of carbonic anhydrase (Figure 9.35).

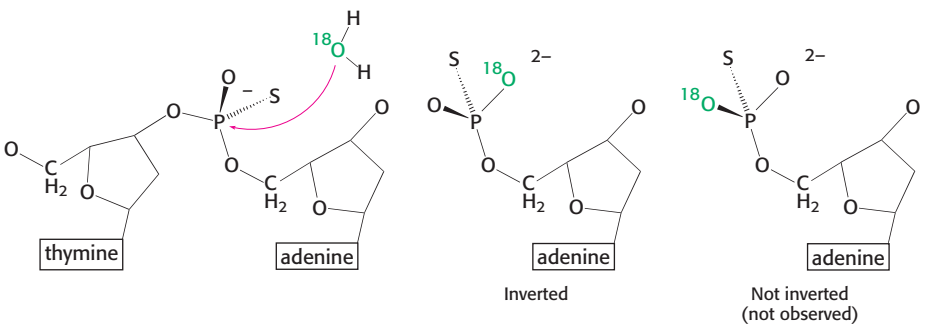


Figure 9.34 Stereochemistry of cleaved DNA. Cleavage of DNA by *Eco*RV endonuclease results in overall inversion of the stereochemical configuration at the phosphorus atom, as indicated by the stereochemistry of the phosphorus atom bound to one bridging oxygen atom, one ^{16}O, one ^{18}O, and one sulfur atom. This configuration strongly suggests that the hydrolysis takes place by water's direct attack at the phosphorus atom.

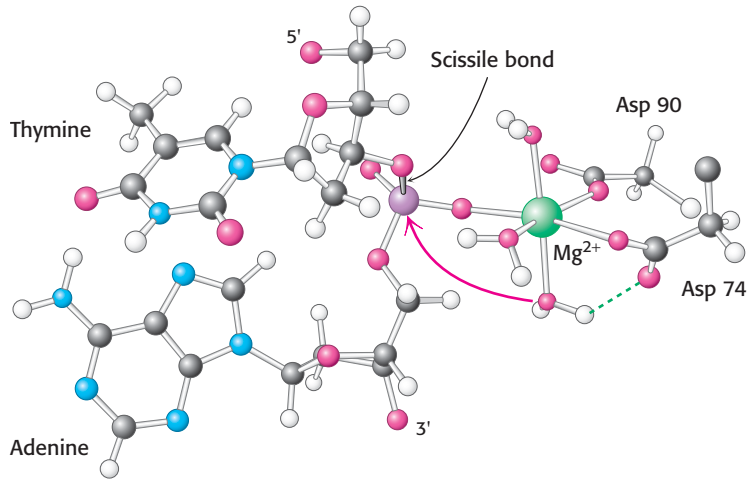


Figure 9.35 A magnesium ion-binding site in *Eco*RV endonuclease. The magnesium ion helps to activate a water molecule and positions it so that it can attack the phosphorus atom.

The complete catalytic apparatus is assembled only within complexes of cognate DNA molecules, ensuring specificity

We now return to the question of specificity, the defining feature of restriction enzymes. The recognition sequences for most restriction endonucleases are *inverted repeats*. This arrangement gives the three-dimensional structure of the recognition site a *twofold rotational symmetry* (Figure 9.36).

The restriction enzymes display a corresponding symmetry: they are dimers whose two subunits are related by twofold rotational symmetry. The matching symmetry of the recognition sequence and the enzyme facilitates the recognition of cognate DNA by the enzyme. This similarity in structure has been confirmed by the determination of the structure of the complex between *Eco*RV endonuclease and DNA fragments containing its recognition sequence (Figure 9.37). The enzyme surrounds the DNA in a tight embrace.

An enzyme's binding affinity for substrates often determines specificity. Surprisingly, however, binding studies performed in the absence of magnesium have demonstrated that the *Eco*RV endonuclease binds to all sequences, both cognate and noncognate, with approximately equal affinity. Why, then, does the enzyme cleave only cognate sequences? The answer lies in a unique set of interactions between the enzyme and a cognate DNA sequence.

Within the 5′-GATATC-3′ sequence, the G and A bases at the 5′ end of each strand and their Watson–Crick partners directly contact the enzyme by hydrogen bonding with residues that are located in two loops, one

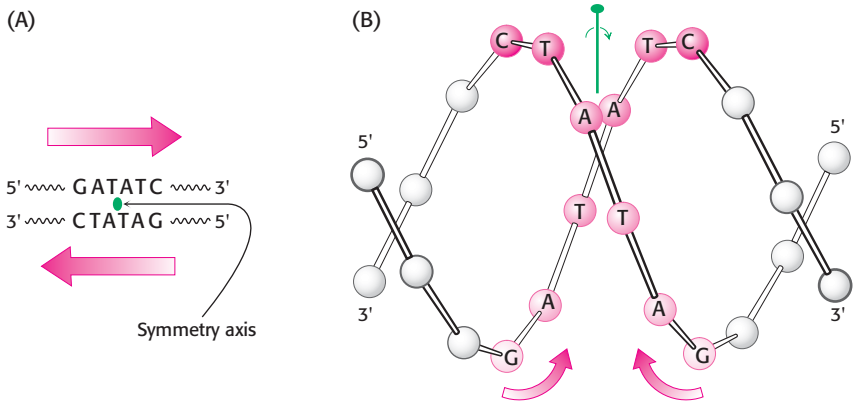


Figure 9.36 Structure of the recognition site of *Eco*RV endonuclease. (A) The sequence of the recognition site, which is symmetric around the axis of rotation designated in green. (B) The inverted repeat within the recognition sequence of *Eco*RV (and most other restriction endonucleases) endows the DNA site with twofold rotational symmetry.

The structures of complexes formed with noncognate DNA fragments are strikingly different from those formed with cognate DNA: the noncognate DNA conformation is not substantially distorted (Figure 9.39). *This lack of distortion has important consequences with regard to catalysis. No phosphate is positioned sufficiently close to the active-site aspartate residues to complete a magnesium ion-binding site* (see Figure 9.35). Hence, the nonspecific complexes do not bind the magnesium ions and the complete catalytic apparatus is never assembled. The distortion of the substrate and the subsequent binding of the magnesium ion account for the catalytic specificity of more than 1,000,000-fold that is observed for *Eco*RV endonuclease. *Thus, enzyme specificity may be determined by the specificity of enzyme action rather than the specificity of substrate binding.*

We can now see the role of binding energy in this strategy for attaining catalytic specificity. The distorted DNA makes additional contacts with the enzyme, increasing the binding energy. However, the increase in binding energy is canceled by the energetic cost of distorting the DNA from its relaxed conformation (Figure 9.40). Thus, for *Eco*RV endonuclease, there is little difference in binding affinity for cognate and nonspecific DNA fragments. However, the distortion in the cognate complex dramatically affects catalysis by completing the magnesium ion-binding site. This example illustrates how enzymes can utilize available binding energy to deform substrates and poise them for chemical transformation. Interactions that take place within the distorted substrate complex stabilize the transition state leading to DNA hydrolysis.

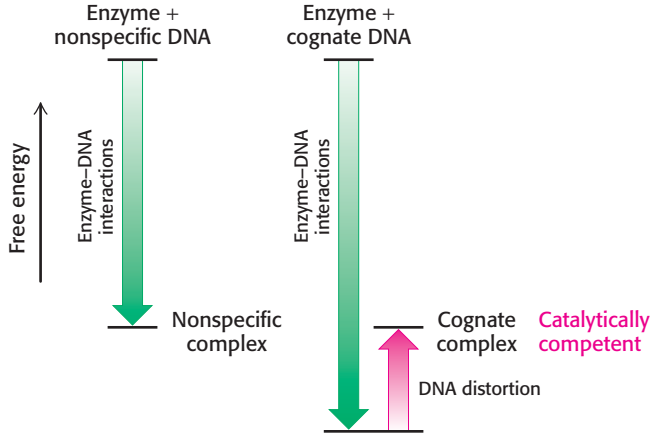


Figure 9.40 Greater binding energy of *Eco*RV endonuclease bound to cognate versus noncognate DNA. The additional interactions between *Eco*RV endonuclease and cognate DNA increase the binding energy, which can be used to drive DNA distortions necessary for forming a catalytically competent complex.

Host-cell DNA is protected by the addition of methyl groups to specific bases

How does a host cell harboring a restriction enzyme protect its own DNA? The host DNA is methylated on specific adenine bases within host recognition sequences by other enzymes called *methylases* (Figure 9.41). An endonuclease will not cleave DNA if its recognition sequence is methylated. For each restriction endonuclease, the host cell produces a corresponding methylase that marks the host DNA at the appropriate methylation site. These pairs of enzymes are referred to as *restriction-modification systems.*

The distortion in the DNA explains how methylation blocks catalysis and protects host-cell DNA. The host *E. coli* adds a methyl group to the amino group of the adenine nucleotide at the 5′ end of the recognition sequence. The presence of the methyl group blocks the formation of a

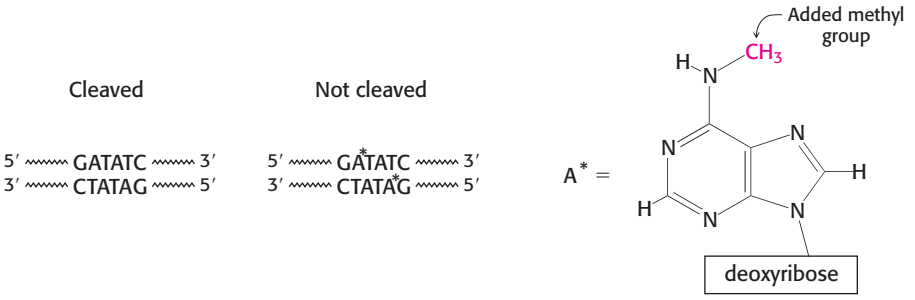


Figure 9.41 Protection by methylation. The recognition sequence for *Eco*RV endonuclease (left) and the sites of methylation (right) in DNA protected from the catalytic action of the enzyme.

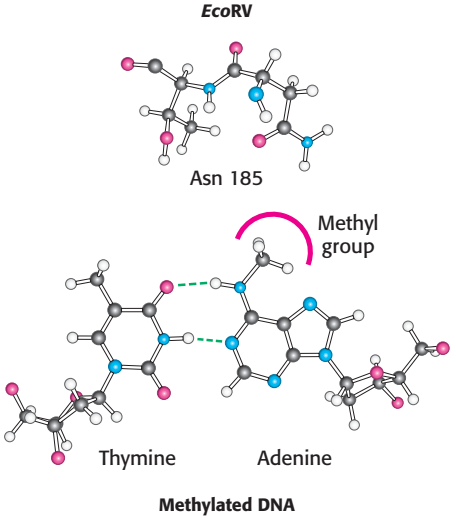


Figure 9.42 Methylation of adenine. The methylation of adenine blocks the formation of hydrogen bonds between *Eco*RV endonuclease and cognate DNA molecules and prevents their hydrolysis.

hydrogen bond between the amino group and the side-chain carbonyl group of asparagine 185 (Figure 9.42). This asparagine residue is closely linked to the other amino acids that form specific contacts with the DNA. The absence of the hydrogen bond disrupts other interactions between the enzyme and the DNA substrate, and the distortion necessary for cleavage will not take place.

Type II restriction enzymes have a catalytic core in common and are probably related by horizontal gene transfer

Type II restriction enzymes are prevalent in Archaea and Bacteria. What can we tell of the evolutionary history of these enzymes? Comparison of the amino acid sequences of a variety of type II restriction endonucleases did not reveal significant sequence similarity between most pairs of enzymes. However, a careful examination of three-dimensional structures, taking into account the location of the active sites, revealed the presence of a core structure conserved in the different enzymes. This structure includes β strands that contain the aspartate (or, in some cases, glutamate) residues forming the magnesium ion-binding sites (Figure 9.43).

These observations indicate that many type II restriction enzymes are indeed evolutionarily related. Analyses of the sequences in greater detail suggest that bacteria may have obtained genes encoding these enzymes from other species by *horizontal gene transfer*, the passing between species of pieces of DNA (such as plasmids) that provide a selective advantage in a particular environment. For example, *Eco*RI (from *E. coli*) and *Rsr*I (from *Rhodobacter sphaeroides*) are 50% identical in sequence over 266 amino acids, clearly indicative of a close evolutionary relationship. However, these species of bacteria are not closely related. Thus, *these species appear to have obtained the gene for these restriction endonucleases from a common source more recently than the time of their evolutionary divergence.* Moreover, the codons used by the gene encoding *Eco*RI endonuclease to specify given

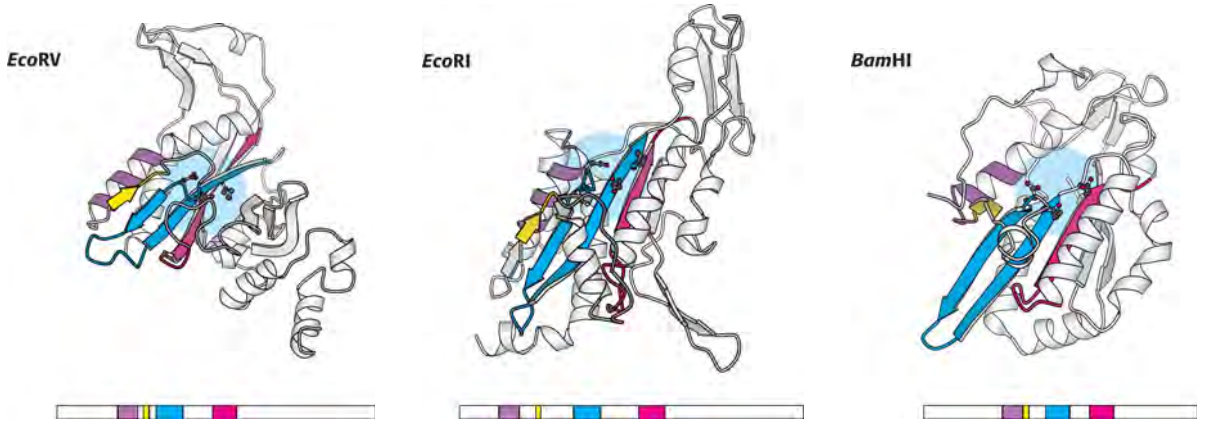


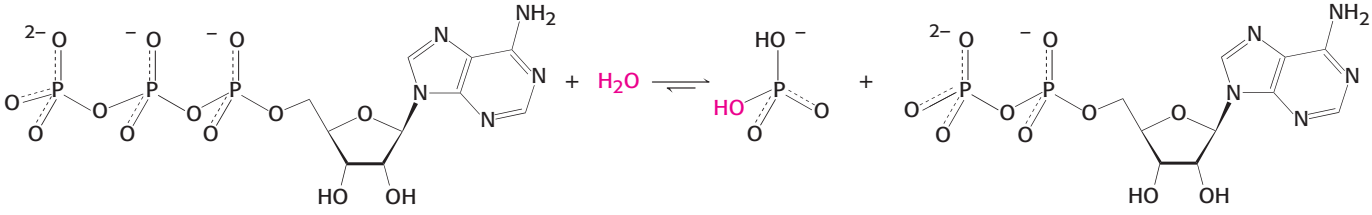
Figure 9.43 A conserved structural core in type II restriction enzymes. Four conserved structural elements, including the active-site region (in blue), are highlighted in color in these models of a single monomer from each dimeric enzyme. *Notice* that these elements adapt similar structures in each enzyme. The positions of the amino acid sequences that form these elements within each overall sequence are represented schematically below each structure. [Drawn from 1RVB.pdb; 1ERI.pdb; 1BHM.pdb.]

amino acids are strikingly different from the codons used by most *E. coli* genes, which suggests that the gene did not originate in *E. coli*.

Horizontal gene transfer may be a common event. For example, genes that inactivate antibiotics are often transferred, leading to the transmission of antibiotic resistance from one species to another. For restriction-modification systems, protection against viral infections may have favored horizontal gene transfer.

9.4 Myosins Harness Changes in Enzyme Conformation to Couple ATP Hydrolysis to Mechanical Work

The final enzymes that we will consider are the myosins. These enzymes catalyze the hydrolysis of adenosine triphosphate (ATP) to form adenosine diphosphate (ADP) and inorganic phosphate (P_i) and use the energy associated with this thermodynamically favorable reaction to drive the motion of molecules within cells.



For example, when we lift a book, the energy required comes from ATP hydrolysis catalyzed by myosin in our muscles. Myosins are found in all eukaryotes and the human genome encodes more than 40 different myosins. Myosins generally have elongated structures with globular domains that actually carry out ATP hydrolysis (Figure 9.44). In this chapter, we will focus on the globular ATPase domains, particularly the strategies that allow myosins to hydrolyze ATP in a controlled manner and to use the free energy associated with this reaction to promote substantial conformational changes within the myosin molecule. These conformational changes are amplified by other structures in the elongated myosin molecules to transport proteins or other cargo substantial distances within cells. In Chapter 35, we will examine the action of myosins and other molecular-motor proteins in much more detail.

As will be discussed in Chapter 15, ATP is used as the major currency of energy inside cells. Many enzymes use ATP hydrolysis to drive other reactions and processes. In almost all cases, an enzyme that hydrolyzed ATP without any such coupled processes would simply drain the energy reserves of a cell without benefit.

ATP hydrolysis proceeds by the attack of water on the gamma phosphoryl group

In our examination of the mechanism of restriction enzymes, we learned that an activated water molecule performs a nucleophilic attack on phosphorus to cleave the phosphodiester backbone of DNA. The cleavage of ATP by myosins follows an analogous mechanism. To understand the myosin mechanism in more detail, we must first examine the structure of the myosin *ATPase* domain.

The structures of the ATPase domains of several different myosins have been examined. One such domain, that from the soil-living amoeba

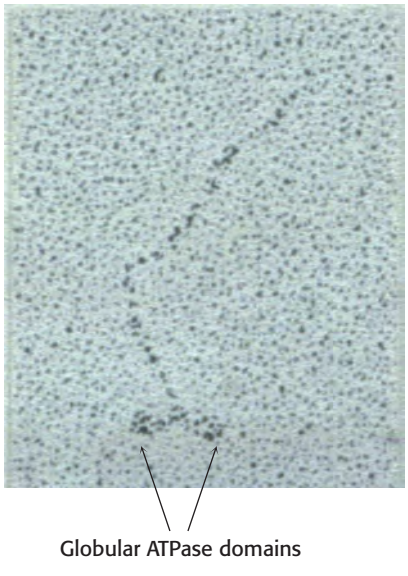


Figure 9.44 Elongated structure of muscle myosin. An electron micrograph showing myosin from mammalian muscle. This dimeric protein has an elongated structure with two globular ATPase domains per dimer. [Courtesy of Dr. Paula Flicker, Dr. Theo Walliman, and Dr. Peter Vibert.]

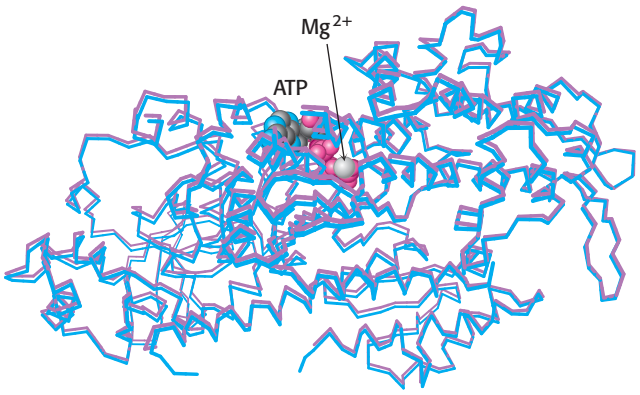


Figure 9.45 Myosin–ATP complex structure. An overlay of the structures of the ATPase domain from *Dictyostelium discoideum* myosin with no ligands bound (blue) and the complex of this protein with ATP and magnesium bound (red). *Notice* that the two structures are extremely similar to one another. [Drawn from 1FMV.pdb and 1FMW.pdb].

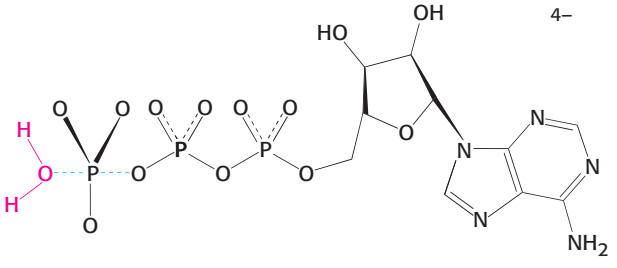
Dictyostelium discoideum, an organism that has been extremely useful for studying cell movement and molecular-motor proteins, has been studied in great detail. The crystal structure of this protein fragment in the absence of nucleotides revealed a single globular domain comprising approximately 750 amino acids. A water-filled pocket is present toward the center of the structure, suggesting a possible nucleotide-binding site. Crystals of this protein were soaked in a solution containing ATP and the structure was examined again. Remarkably, this structure revealed intact ATP bound in the active site with very little change in the overall structure and without evidence of significant hydrolysis (Figure 9.45). The ATP is also bound to a Mg^{2+} ion.

Kinetic studies of myosins, as well as many other enzymes having ATP or other nucleoside triphosphates as a substrate, reveal that these enzymes are essentially inactive in the absence of divalent metal ions such as magnesium (Mg^{2+}) or manganese (Mn^{2+}) but acquire activity on the addition of these ions. In contrast with the enzymes discussed so far, the metal is not a component of the active site. Rather, nucleotides such as ATP bind these ions, and it is the metal ion–nucleotide complex that is the true substrate for the enzymes. The dissociation constant for the ATP–Mg^{2+} complex is approximately 0.1 mM, and thus, given that intracellular Mg^{2+} concentrations are typically in the millimolar range, essentially all nucleoside triphosphates are present as NTP–Mg^{2+} complexes. *Magnesium or manganese complexes of nucleoside triphosphates are the true substrates for essentially all NTP-dependent enzymes.*

The nucleophilic attack by a water molecule on the γ-phosphoryl group requires some mechanism such as a basic residue or a bound metal ion to activate the water. Examination of the myosin–ATP complex structure shows no basic residue in an appropriate position and reveals that the bound Mg^{2+} ion is too far away from the phosphoryl group to play this role. These observations suggest why this ATP complex is relatively stable; the enzyme is not in a conformation that is competent to catalyze the reaction. This observation suggests that the domain must undergo a conformational change to catalyze the ATP-hydrolysis reaction.

Formation of the transition state for ATP hydrolysis is associated with a substantial conformational change

The catalytically competent conformation of the myosin ATPase domain must bind and stabilize the transition state of the reaction. In analogy with restriction enzymes, we expect that ATP hydrolysis includes a pentacoordinate transition state.



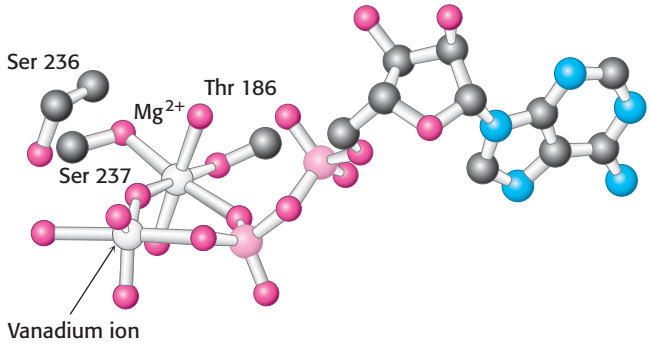
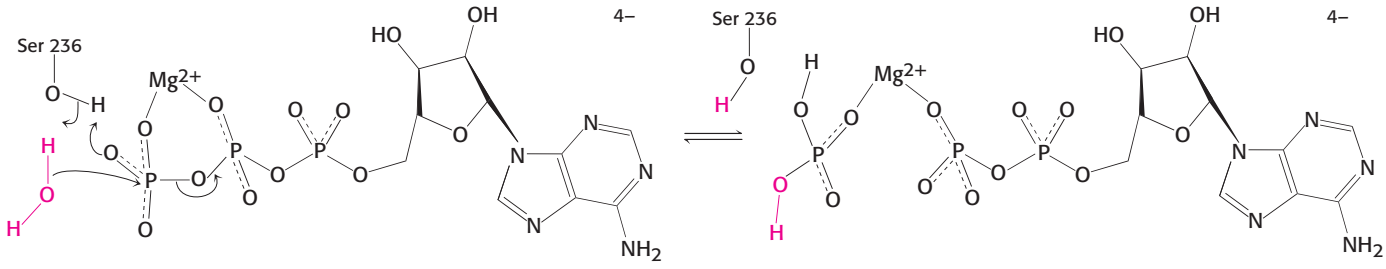


Figure 9.46 Myosin ATPase Transition-State Analog. The structure of the transition-state analog formed by treating the myosin ATPase domain with ADP and vanadate (VO_4^{3-}) in the presence of magnesium is shown. *Notice* that the vanadium ion is coordinated to five oxygen atoms including one from ADP. The positions of two residues that bind magnesium as well as Ser 236, a residue that appears to play a direct role in catalysis, are shown. [Drawn from 1VOM.pdb]

Such pentacoordinate structures based on phosphorus are too unstable to be readily observed. However, transition-state analogs in which other atoms replace phosphorus are more stable. The transition metal vanadium, in particular, forms similar structures. The myosin ATPase domain was crystallized in the presence of ADP and vanadate, VO_4^{3-}. The result was the formation of a complex that closely matches the expected transition-state structure (Figure 9.46). As expected, the vanadium atom is coordinated to five oxygen atoms, including one oxygen atom from ADP diametrically opposite an oxygen atom that is analogous to the attacking water molecule in the transition state. The Mg^{2+} ion is coordinated to one oxygen atom from the vanadate, one oxygen atom from the ADP, two hydroxyl groups from the enzyme, and two water molecules. In this position, this ion does not appear to play any direct role in activating the attacking water. However, an additional residue from the enzyme, Ser 236, is well positioned to play a role in catalysis (see Figure 9.46). In the proposed mechanism of ATP hydrolysis based on this structure, the water molecule attacks the γ-phosphoryl group, with the hydroxyl group of Ser 236 facilitating the transfer of a proton from the attacking water to the hydroxyl group of Ser 236, which, in turn, is deprotonated by one of the oxygen atoms of the γ-phosphoryl group (Figure 9.47). *Thus, in effect, the ATP serves as a base to promote its own hydrolysis.*

Comparison of the overall structures of the myosin ATPase domain complexed with ATP and with the ADP–vanadate reveals some remarkable differences. Around the active site, some residues move somewhat. In particular, a stretch of amino acids moves closer to the nucleotide by approximately 2 Å and interact with the oxygen atom that corresponds to the attacking water molecule. These changes help facilitate the hydrolysis

Figure 9.47 Facilitating Water Attack. The water molecule attacking the γ-phosphoryl group of ATP is deprotonated by the hydroxyl group of Ser 236, which, in turn, is deprotonated by one of the oxygen atoms of the γ-phosphoryl group forming the $H_2PO_4^-$ product.



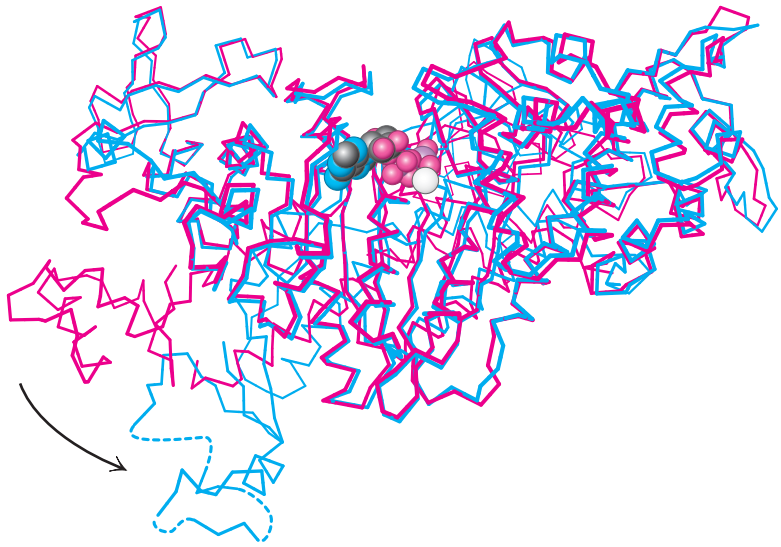


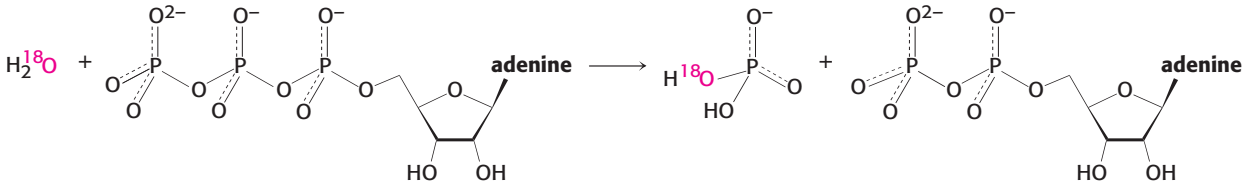
Figure 9.48 Myosin conformational changes. A comparison of the overall structures of the myosin ATPase domain with ATP bound (shown in red) and that with the transition-state analog ADP–vanadate (shown in blue). *Notice* the large conformational change of a region at the carboxyl-terminus of the domain, some parts of which move as much as 25 Å. [Drawn from 1FMW.pdb and 1VOM.pdb].

reaction by stabilizing the transition state. However, examination of the overall structure shows even more striking changes.

A region comprising approximately 60 amino acids at the carboxyl-terminus of the domain adopts a different configuration in the ADP–vanadate complex, displaced by as much as 25 Å from its position in the ATP complex (Figure 9.48). This displacement tremendously amplifies the relatively subtle changes that take place in the active site. The effect of this motion is amplified even more as this carboxyl-terminal domain is connected to other structures within the elongated structures typical of myosin molecules (see Figure 9.44). Thus, the conformation that is capable of promoting the ATP hydrolysis reaction is itself substantially different from other conformational changes that take place in the course of the catalytic cycle.

The altered conformation of myosin persists for a substantial period of time

Myosins are slow enzymes, typically turning over approximately once per second. What steps limit the rate of turnover? In an experiment that was particularly revealing, the hydrolysis of ATP was catalyzed by the myosin ATPase domain from mammalian muscle. The reaction took place in water labeled with ^{18}O to track the incorporation of solvent oxygen into the reaction products. The fraction of oxygen in the phosphate product was analyzed. In the simplest case, the phosphate would be expected to contain one oxygen atom derived from water and three initially present in the terminal phosphoryl group of ATP.



Instead, between two and three of the oxygen atoms in the phosphate were found, on average, to be derived from water. These observations indicate

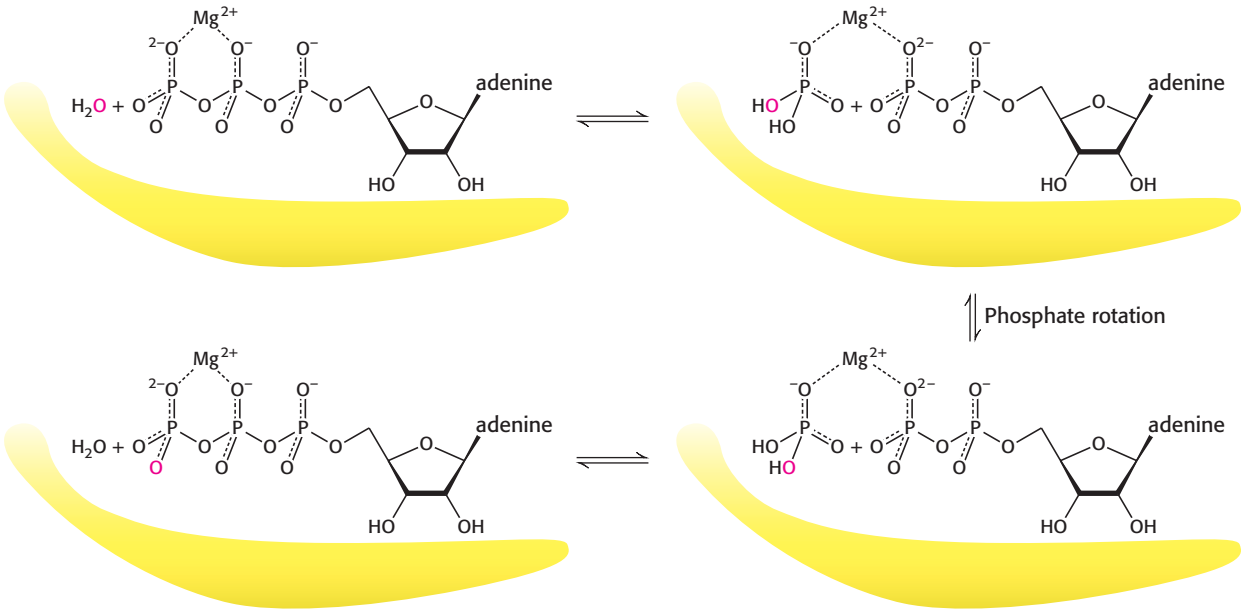


Figure 9.49 Reversible hydrolysis of ATP within the myosin active site. For myosin, more than one atom of oxygen from water is incorporated in inorganic phosphate. The oxygen atoms are incorporated in cycles of hydrolysis of ATP to ADP and inorganic phosphate, phosphate rotation within the active site, and reformation of ATP now containing oxygen from water.

that the ATP hydrolysis reaction within the enzyme active site is reversible. Each molecule of ATP is cleaved to ADP and P_i and then re-formed from these products several times before the products are released from the enzyme (Figure 9.49). At first glance, this observation is startling because ATP hydrolysis is a very favorable reaction with an equilibrium constant of approximately 140,000. However, this equilibrium constant applies to the molecules free in solution, not within the active site of an enzyme. Indeed, more-extensive analysis suggests that this equilibrium constant on the enzyme is approximately 10, indicative of a general strategy used by enzymes. Enzymes catalyze reactions by stabilizing the transition state. The structure of this transition state is intermediate between the enzyme-bound reactants and the enzyme-bound products. Many of the interactions that stabilize the transition state will help equalize the stabilities of the reactants and the products. Thus, *the equilibrium constant between enzyme-bound reactants and products is often close to 1, regardless of the equilibrium constant for the reactants and products free in solution.*

These observations reveal that the hydrolysis of ATP to ADP and P_i is not the rate-limiting step for the reaction catalyzed by myosin. Instead, the release of the products, particularly P_i, from the enzyme is rate limiting. The fact that a conformation of myosin with ATP hydrolyzed but still bound to the enzyme persists for a significant period of time is critical for coupling conformational changes that take place in the course of the reaction to other processes.

Myosins are a family of enzymes containing P-loop structures

X-ray crystallography has yielded the three-dimensional structures of a number of different enzymes that share key structural characteristics and, almost certainly, an evolutionary history with myosin. In particular, a conserved NTP-binding core domain is present. This domain consists of a central β sheet, surrounded on both sides by α helices (Figure 9.50). A characteristic feature of this domain is a loop between the first β strand and the first helix. This loop typically has several glycine residues that are often conserved between more closely related members of this large and diverse family. The loop is often referred to as the *P-loop* because

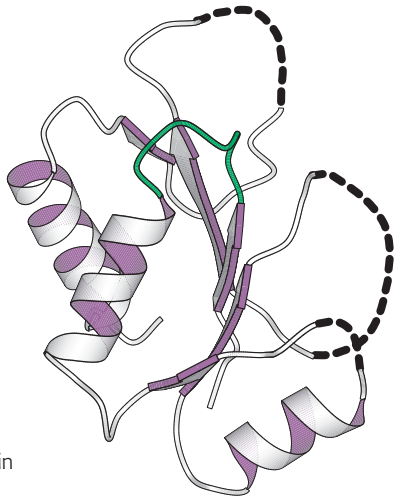
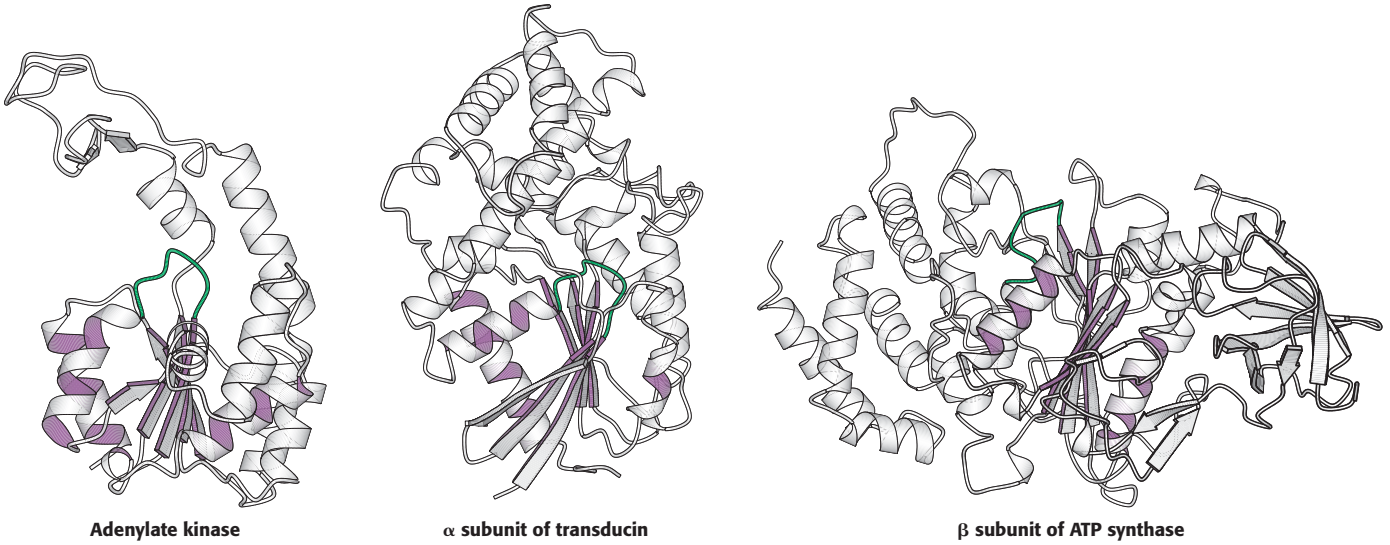


Figure 9.50 The core domain of NMP kinases. *Notice* the P-loop shown in green. The dashed lines represent the remainder of the protein structure. [Drawn from 1GKY.pdb.]

it interacts with phosphoryl groups on the bound nucleotide. P-loop NTPase domains are present in a remarkably wide array of proteins, many of which participate in essential biochemical processes. Examples include ATP synthase, the key enzyme responsible for ATP generation; signal-transduction proteins such as G proteins; proteins essential for translating mRNA into proteins, such as elongation factor Tu; and DNA and RNA unwinding helicases. The wide utility of P-loop NTPase domains is perhaps best explained by their ability to undergo substantial conformational changes on nucleoside triphosphate binding and hydrolysis. We shall encounter these domains throughout the book and shall observe how they function as springs, motors, and clocks. To allow easy recognition of these domains in the book, they will be depicted with the inner surfaces of the ribbons in a ribbon diagram shown in purple and the P-loop shown in green (Figure 9.51).



Adenylate kinase **α subunit of transducin** **β subunit of ATP synthase**

Figure 9.51 Three proteins containing P-loop NTPase domains. *Notice* the conserved domains shown with the inner surfaces of the ribbons in purple and the P-loops in green. [Drawn from 4AKE.pdb; 1TND.pdb; 1BMF.pdb.]

Summary

Enzymes adopt conformations that are structurally and chemically complementary to the transition states of the reactions that they catalyze. Sets of interacting amino acid residues make up sites with the special structural and chemical properties necessary to stabilize the transition state. Enzymes use five basic strategies to form and stabilize the transition state: (1) the use of binding energy, (2) covalent catalysis, (3) general acid–base catalysis, (4) metal ion catalysis, and (5) catalysis by approximation. The four classes of enzymes examined in this chapter catalyze the addition of water to their substrates but have different requirements for catalytic speed, specificity, and coupling to other processes.

9.1 Proteases Facilitate a Fundamentally Difficult Reaction

The cleavage of peptide bonds by chymotrypsin is initiated by the attack by a serine residue on the peptide carbonyl group. The attacking hydroxyl group is activated by interaction with the imidazole group of a histidine residue, which is, in turn, linked to an aspartate residue. This Ser-His-Asp catalytic triad generates a powerful nucleophile. The product of this initial reaction is a covalent intermediate formed by the enzyme and an acyl group derived from the bound substrate. The hydrolysis of this acyl-enzyme intermediate completes the cleavage process. The tetrahedral intermediates for these reactions have a negative charge on the peptide carbonyl oxygen atom. This negative charge is stabilized by interactions with peptide NH groups in a region on the enzyme termed the oxyanion hole.

Other proteases employ the same catalytic strategy. Some of these proteases, such as trypsin and elastase, are homologs of chymotrypsin. Other proteases, such as subtilisin, contain a very similar catalytic triad that has arisen by convergent evolution. Active-site structures that differ from the catalytic triad are present in a number of other classes of proteases. These classes employ a range of catalytic strategies but, in each case, a nucleophile is generated that is sufficiently powerful to attack the peptide carbonyl group. In some enzymes, the nucleophile is derived from a side chain whereas, in others, an activated water molecule attacks the peptide carbonyl directly.

9.2 Carbonic Anhydrases Make a Fast Reaction Faster

Carbonic anhydrases catalyze the reaction of water with carbon dioxide to generate carbonic acid. The catalysis can be extremely fast: some carbonic anhydrases hydrate carbon dioxide at rates as high as 1 million times per second. A tightly bound zinc ion is a crucial component of the active sites of these enzymes. Each zinc ion binds a water molecule and promotes its deprotonation to generate a hydroxide ion at neutral pH. This hydroxide ion attacks carbon dioxide to form bicarbonate ion, HCO_3^-. Because of the physiological roles of carbon dioxide and bicarbonate ions, speed is of the essence for this enzyme. To overcome limitations imposed by the rate of proton transfer from the zinc-bound water molecule, the most-active carbonic anhydrases have evolved a proton shuttle to transfer protons to a buffer.

9.3 Restriction Enzymes Catalyze Highly Specific DNA-Cleavage Reactions

A high level of substrate specificity is often the key to biological function. Restriction endonucleases that cleave DNA at specific recognition sequences discriminate between molecules that contain these recognition

sequences and those that do not. Within the enzyme–substrate complex, the DNA substrate is distorted in a manner that generates a magnesium ion-binding site between the enzyme and DNA. The magnesium ion binds and activates a water molecule, which attacks the phosphodiester backbone.

Some enzymes discriminate between potential substrates by binding them with different affinities. Others may bind many potential substrates but promote chemical reactions efficiently only on specific molecules. Restriction endonucleases such as *Eco*RV endonuclease employ the latter mechanism. Only molecules containing the proper recognition sequence are distorted in a manner that allows magnesium ion binding and, hence, catalysis. Restriction enzymes are prevented from acting on the DNA of a host cell by the methylation of key sites within its recognition sequences. The added methyl groups block specific interactions between the enzymes and the DNA such that the distortion necessary for cleavage does not take place.

9.4 Myosins Harness Changes in Enzyme Conformation to Couple ATP Hydrolysis to Mechanical Work

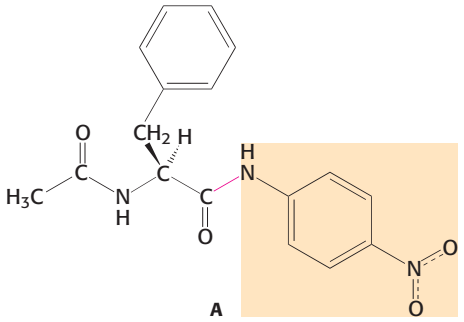
Finally, myosins catalyze the hydrolysis of adenosine triphosphate (ATP) to form adenosine diphosphate (ADP) and inorganic phosphate (P_i). The conformations of myosin ATPase domains free of bound nucleotides and with bound ATP are quite similar. Through the use of ADP and vanadate (VO_4^{3-}), an excellent mimic of the transition state for ATP hydrolysis bound to the myosin ATPase domain can be produced. The structure of this complex reveals that dramatic conformational changes take place on formation of this species from the ATP complex. These conformational changes are used to drive substantial motions in molecular motors. The rate of ATP hydrolysis by myosin is relatively low and is limited by the rate of product release from the enzyme. The hydrolysis of ATP to ADP and P_i within the enzyme is reversible with an equilibrium constant of approximately 10, compared with an equilibrium constant of 140,000 for these species free in solution. Myosins are examples of P-loop NTPase enzymes, a large collection of protein families that play key roles in a range of biological processes by virtue of the conformational changes that they undergo with various nucleotides bound.

Key Terms

binding energy (p. 254)
induced fit (p. 254)
covalent catalysis (p. 254)
general acid–base catalysis (p. 254)
catalysis by approximation (p. 254)
metal ion catalysis (p. 254)
chemical modification reaction (p. 256)
catalytic triad (p. 258)
oxyanion hole (p. 259)
protease inhibitor (p. 264)
proton shuttle (p. 270)
recognition sequence (p. 272)
in-line displacement (p. 273)
methylases (p. 277)
restriction-modification system (p. 277)
horizontal gene transfer (p. 278)
ATPase (p. 279)
P-loop (p. 283)

Problems

1. *No burst.* Examination of the cleavage of the amide substrate, A, by chymotrypsin with the use of stopped-flow kinetic methods reveals no burst. The reaction is monitored by noting the color produced by the release of the amino part of the substrate (highlighted in orange). Why is no burst observed?



2. *Contributing to your own demise.* Consider the subtilisin substrates A and B.

Phe-Ala-Gln-Phe-X	Phe-Ala-His-Phe-X
A	**B**

These substrates are cleaved (between Phe and X) by native subtilisin at essentially the same rate. However, the His 64-to-Ala mutant of subtilisin cleaves substrate B more than 1000-fold as rapidly as it cleaves substrate A. Propose an explanation.

3. *1 + 1 ≠ 2.* Consider the following argument. In subtilisin, mutation of Ser 221 to Ala results in a 10^6-fold decrease in activity. Mutation of His 64 to Ala results in a similar 10^6-fold decrease. Therefore, simultaneous mutation of Ser 221 to Ala and His 64 to Ala should result in a $10^6 \times 10^6 = 10^{12}$-fold reduction in activity. Is this reduction correct? Why or why not?

4. *Adding a charge.* In chymotrypsin, a mutant was constructed with Ser 189, which is in the bottom of the substrate-specificity pocket, changed to Asp. What effect would you predict for this Ser 189→Asp 189 mutation?

5. *Conditional results.* In carbonic anhydrase II, mutation of the proton-shuttle residue His 64 to Ala was expected to result in a decrease in the maximal catalytic rate. However, in buffers such as imidazole with relatively small molecular components, no rate reduction was observed. In buffers with larger molecular components, significant rate reductions were observed. Propose an explanation.

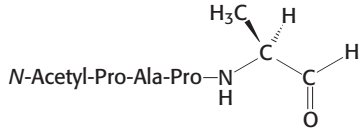
6. *How many sites?* A researcher has isolated a restriction endonuclease that cleaves at only one particular 10-base-pair site. Would this enzyme be useful in protecting cells from viral infections, given that a typical viral genome is 50,000 base pairs long? Explain.

7. *Is faster better?* Restriction endonucleases are, in general, quite slow enzymes with typical turnover numbers of 1 s^{-1}. Suppose that endonucleases were faster with turnover numbers similar to those for carbonic anhydrase (10^6 s^{-1}). Would this increased rate be beneficial to host cells, assuming that the fast enzymes have similar levels of specificity?

8. *Adopting a new gene.* Suppose that one species of bacteria obtained one gene encoding a restriction endonuclease by horizontal gene transfer. Would you expect this acquisition to be beneficial?

9. *Chelation therapy.* Treatment of carbonic anhydrase with high concentrations of the metal chelator EDTA (ethylenediaminetetraacetic acid) results in the loss of enzyme activity. Propose an explanation.

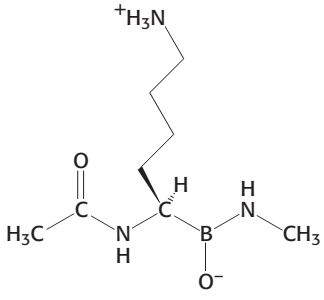
10. *An aldehyde inhibitor.* Elastase is specifically inhibited by an aldehyde derivative of one of its substrates:



(a) Which residue in the active site of elastase is most likely to form a covalent bond with this aldehyde?

(b) What type of covalent link would be formed?

11. *Identify the enzyme.* Consider the structure of molecule A. Which enzyme discussed in the chapter do you think molecule A will most effectively inhibit?



Molecule A

12. *Acid test.* At pH 7.0, carbonic anhydrase exhibits a k_{cat} of 600,000 s^{-1}. Estimate the value expected for k_{cat} at pH 6.0.

13. *Restriction.* To terminate a reaction in which a restriction enzyme cleaves DNA, researchers often add high concentrations of the metal chelator EDTA (ethylenediaminetetraacetic acid). Why does the addition of EDTA terminate the reaction?

14. *Labeling strategy.* ATP is added to the myosin ATPase domain in water labeled with ^{18}O. After 50% of the ATP has been hydrolyzed, the remaining ATP is isolated and found to contain ^{18}O. Explain.

15. *Viva le resistance.* Many patients become resistant to HIV protease inhibitors with the passage of time owing to mutations in the HIV gene that encodes the protease.

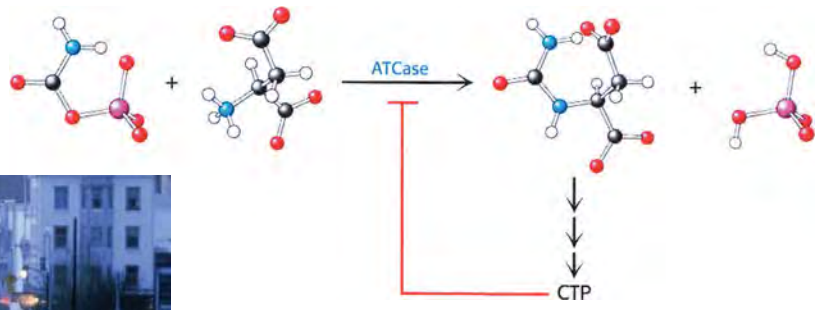
Mutations are not found in the aspartate residue that interacts with the drugs. Why not?

16. *More than one way to skin* k_{cat}. Serine 236 in *Dictyostelium discoideum* myosin has been mutated to alanine. The mutated protein showed modestly reduced ATPase activity. Analysis of the crystal structure of the mutated protein revealed that a water molecule occupied the position of the hydroxyl group of the serine residue in the wild-type protein. Propose a mechanism for the ATPase activity of the mutated enzyme.

Mechanism Problem

17. *Complete the mechanism.* On the basis of the information provided in Figure 9.17, complete the mechanisms for peptide-bond cleavage by (a) a cysteine protease, (b) an aspartyl protease, and (c) a metalloprotease.

Regulatory Strategies



Like motor traffic, metabolic pathways flow more efficiently when regulated by signals. Cytidine triphosphate (CTP), the final product of a multistep pathway, controls flux through the pathway by inhibiting the committed step catalyzed by aspartate transcarbamoylase (ATCase). [(Left) Michael Winokur Photography/ Getty Images.]

The activity of enzymes must often be regulated so that they function at the proper time and place. This regulation is essential for coordination of the vast array of biochemical processes taking place at any instant in an organism. Enzymatic activity is regulated in five principal ways:

1. *Allosteric Control.* Allosteric proteins contain distinct regulatory sites and multiple functional sites. The binding of small signal molecules at regulatory sites is a significant means of controlling the activity of these proteins. Moreover, allosteric proteins show the property of *cooperativity*: activity at one functional site affects the activity at others. Proteins displaying allosteric control are thus information transducers: their activity can be modified in response to signal molecules or to information shared among active sites. This chapter examines one of the best-understood allosteric proteins: the enzyme *aspartate transcarbamoylase* (ATCase). Catalysis by aspartate transcarbamoylase of the first step in pyrimidine biosynthesis is inhibited by cytidine triphosphate, the final product of that biosynthesis, in an example of *feedback inhibition*. We have already examined an allosteric protein—hemoglobin, the oxygen transport protein in the blood (Chapter 7).

2. *Multiple Forms of Enzymes.* Isozymes, or isoenzymes, provide an avenue for varying regulation of the same reaction at distinct locations or times to meet the specific physiological needs in the particular tissue at a particular time. Isozymes are homologous enzymes within a single organism that

OUTLINE

catalyze the same reaction but differ slightly in structure and more obviously in K_M and V_{max} values as well as in regulatory properties. Often, isozymes are expressed in a distinct tissue or organelle or at a distinct stage of development.

3. *Reversible Covalent Modification.* The catalytic properties of many enzymes are markedly altered by the covalent attachment of a modifying group, most commonly a phosphoryl group. ATP serves as the phosphoryl donor in these reactions, which are catalyzed by *protein kinases.* The removal of phosphoryl groups by hydrolysis is catalyzed by *protein phosphatases.* This chapter considers the structure, specificity, and control of *protein kinase A* (PKA), a ubiquitous eukaryotic enzyme that regulates diverse target proteins.

4. *Proteolytic Activation.* The enzymes controlled by some of these regulatory mechanisms cycle between active and inactive states. A different regulatory strategy is used to *irreversibly* convert an inactive enzyme into an active one. Many enzymes are activated by the hydrolysis of a few peptide bonds or even one such bond in inactive precursors called *zymogens* or *proenzymes.* This regulatory mechanism generates digestive enzymes such as chymotrypsin, trypsin, and pepsin. Blood clotting is due to a remarkable cascade of zymogen activations. Active digestive and clotting enzymes are switched off by the irreversible binding of specific inhibitory proteins that are irresistible lures to their molecular prey.

5. *Controlling the Amount of Enzyme Present.* Enzyme activity can also be regulated by adjusting the amount of enzyme present. This important form of regulation usually takes place at the level of transcription. We will consider the control of gene transcription in Chapter 31.

To begin here, we will consider the principles of allostery by examining the enzyme aspartate transcarbamoylase.

10.1 Aspartate Transcarbamoylase Is Allosterically Inhibited by the End Product of Its Pathway

Aspartate transcarbamoylase catalyzes the first step in the biosynthesis of pyrimidines: the condensation of aspartate and carbamoyl phosphate to form *N*-carbamoylaspartate and orthophosphate (Figure 10.1). This reaction

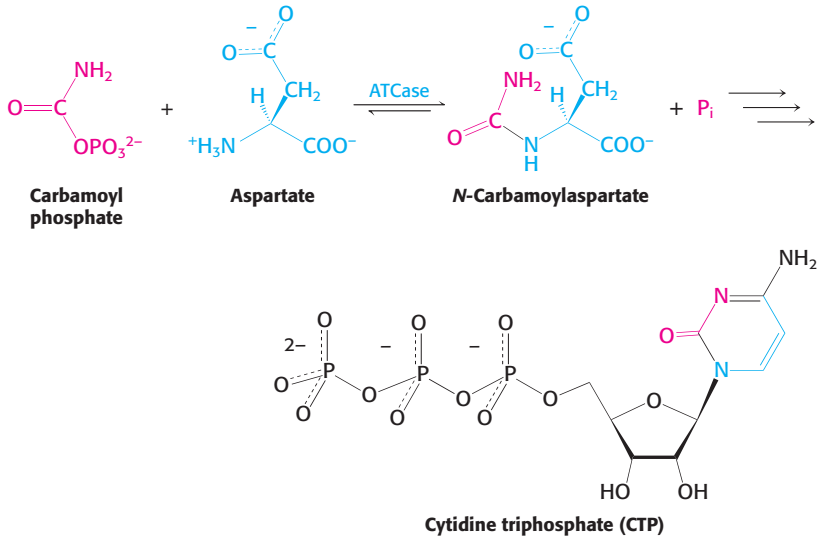


Figure 10.1 ATCase reaction. Aspartate transcarbamoylase catalyzes the committed step, the condensation of aspartate and carbamoyl phosphate to form *N*-carbamoylaspartate, in pyrimidine synthesis.

is the committed step in the pathway that will ultimately yield pyrimidine nucleotides such as cytidine triphosphate (CTP). How is this enzyme regulated to generate precisely the amount of CTP needed by the cell?

John Gerhart and Arthur Pardee found that ATCase is inhibited by CTP, the final product of the ATCase-initiated pathway. The rate of the reaction catalyzed by ATCase is fast at low concentrations of CTP but slows as CTP concentration increases (Figure 10.2). Thus, the pathway continues to make new pyrimidines until sufficient quantities of CTP have accumulated. The inhibition of ATCase by CTP is an example of *feedback inhibition*, the inhibition of an enzyme by the end product of the pathway. Feedback inhibition by CTP ensures that *N*-carbamoylaspartate and subsequent intermediates in the pathway are not needlessly formed when pyrimidines are abundant.

The inhibitory ability of CTP is remarkable because *CTP is structurally quite different from the substrates of the reaction* (see Figure 10.1). Thus CTP must bind to a site distinct from the active site at which substrate binds. Such sites are called *allosteric* or *regulatory sites*. CTP is an example of an *allosteric inhibitor*. In ATCase (but not all allosterically regulated enzymes), the catalytic sites and the regulatory sites are on separate polypeptide chains.

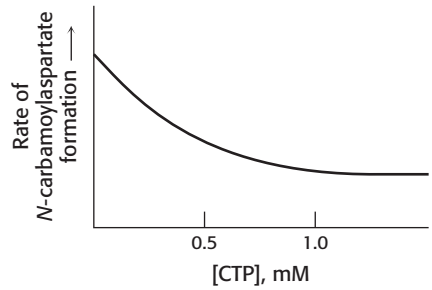


Figure 10.2 CTP inhibits ATCase. Cytidine triphosphate, an end product of the pyrimidine-synthesis pathway, inhibits aspartate transcarbamoylase despite having little structural similarity to reactants or products.

Allosterically regulated enzymes do not follow Michaelis–Menten kinetics

Allosteric enzymes are distinguished by their response to changes in substrate concentration in addition to their susceptibility to regulation by other molecules. Let us examine the rate of product formation as a function of substrate concentration for ATCase (Figure 10.3). The curve differs from that expected for an enzyme that follows Michaelis–Menten kinetics. The observed curve is referred to as sigmoidal because it resembles the letter "S." The vast majority of allosteric enzymes display sigmoidal kinetics. Recall from the discussion of hemoglobin that sigmoidal curves result from cooperation between subunits: the binding of substrate to one active site in a molecule increases the likelihood that substrate will bind to other active sites. To understand the basis of sigmoidal enzyme kinetics and inhibition by CTP, we need to examine the structure of ATCase.

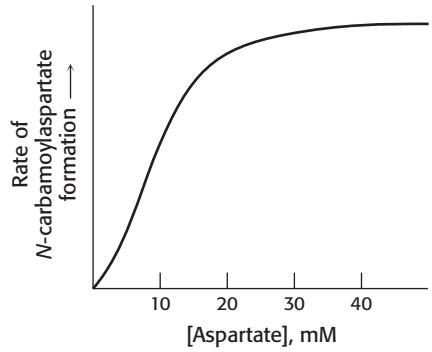


Figure 10.3 ATCase displays sigmoidal kinetics. A plot of product formation as a function of substrate concentration produces a sigmoidal curve because the binding of substrate to one active site increases the activity at the other active sites. Thus, the enzyme shows cooperativity.

ATCase consists of separable catalytic and regulatory subunits

What is the evidence that ATCase has distinct regulatory and catalytic sites? ATCase can be literally separated into regulatory (r) and catalytic (c) subunits by treatment with a mercurial compound such as *p*-hydroxymercuribenzoate, which reacts with sulfhydryl groups (Figure 10.4). Ultracentrifugation following treatment with mercurials revealed that ATCase is composed of two kinds of subunits (Figure 10.5). The subunits can be readily separated by ion-exchange chromatography because they differ markedly in charge or by centrifugation in a sucrose density gradient because they differ in size. These size differences are manifested in the sedimentation coefficients: that of the native enzyme is 11.6S, whereas those of the dissociated subunits are 2.8S and 5.8S. The attached *p*-mercuribenzoate groups can be removed from the separated subunits by adding an excess of mercaptoethanol, providing isolated subunits for study.

The larger subunit is the *catalytic subunit*. This subunit displays catalytic activity but is unresponsive to CTP and does not display sigmoidal kinetics. The isolated smaller subunit can bind CTP, but has no catalytic activity. Hence, that subunit is the *regulatory subunit*. The catalytic subunit (c_3) consists of three chains (34 kd each), and the regulatory subunit (r_2)

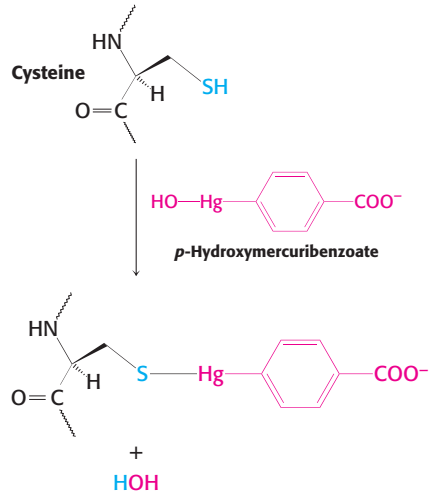


Figure 10.4 Modification of cysteine residues. *p*-Hydroxymercuribenzoate reacts with crucial cysteine residues in aspartate transcarbamoylase.

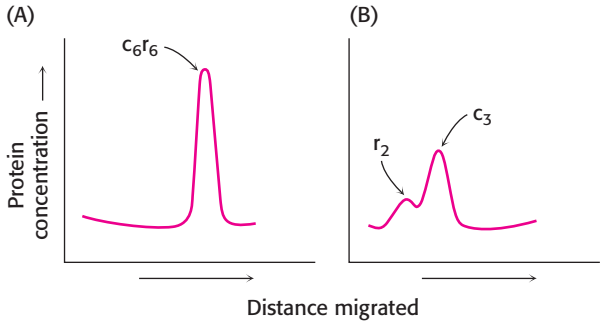


Figure 10.5 Ultracentrifugation studies of ATCase. Sedimentation velocity patterns of (A) native ATCase and (B) the enzyme after treatment with *p*-hydroxymercuribenzoate show that the enzyme can be dissociated into regulatory (r) and catalytic (c) subunits. [After J. C. Gerhart and H. K. Schachman. *Biochemistry* 4:1054–1062, 1965.]

consists of two chains (17 kd each). The catalytic and regulatory subunits combine rapidly when they are mixed. The resulting complex has the same structure, c_6r_6, as the native enzyme: two catalytic trimers and three regulatory dimers.

$$2\,c + 3\,r_2 \longrightarrow c_6r_6$$

Most strikingly, the reconstituted enzyme has the same allosteric and kinetic properties as those of the native enzyme. Thus, ATCase is composed of discrete catalytic and regulatory subunits, and *the interaction of the subunits in the native enzyme produces its regulatory and catalytic properties*.

Allosteric interactions in ATCase are mediated by large changes in quaternary structure

What are the subunit interactions that account for the properties of ATCase? Significant clues have been provided by the three-dimensional structure of ATCase in various forms. Two catalytic trimers are stacked one on top of the other, linked by three dimers of the regulatory chains (Figure 10.6). There are significant contacts between the catalytic and the

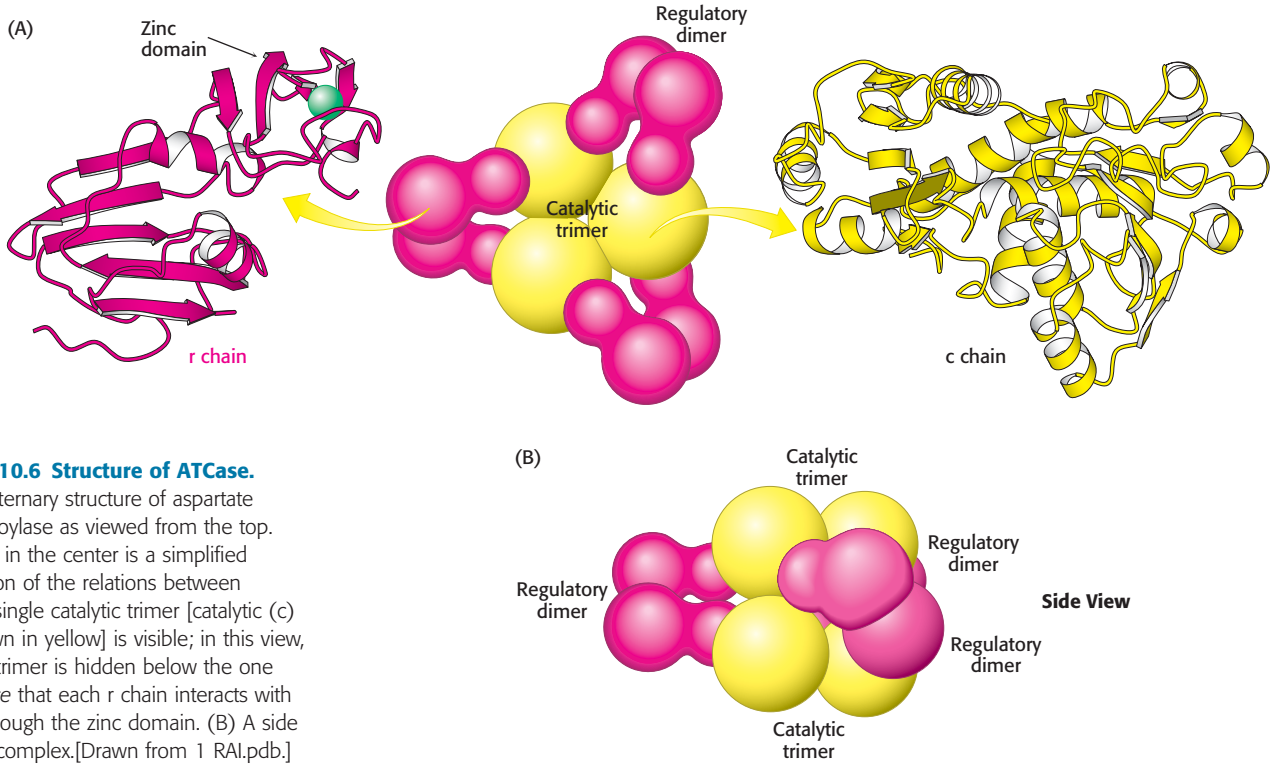


Figure 10.6 Structure of ATCase. (A) The quaternary structure of aspartate transcarbamoylase as viewed from the top. The drawing in the center is a simplified representation of the relations between subunits. A single catalytic trimer [catalytic (c) chains, shown in yellow] is visible; in this view, the second trimer is hidden below the one visible. *Notice* that each r chain interacts with a c chain through the zinc domain. (B) A side view of the complex.[Drawn from 1 RAI.pdb.]

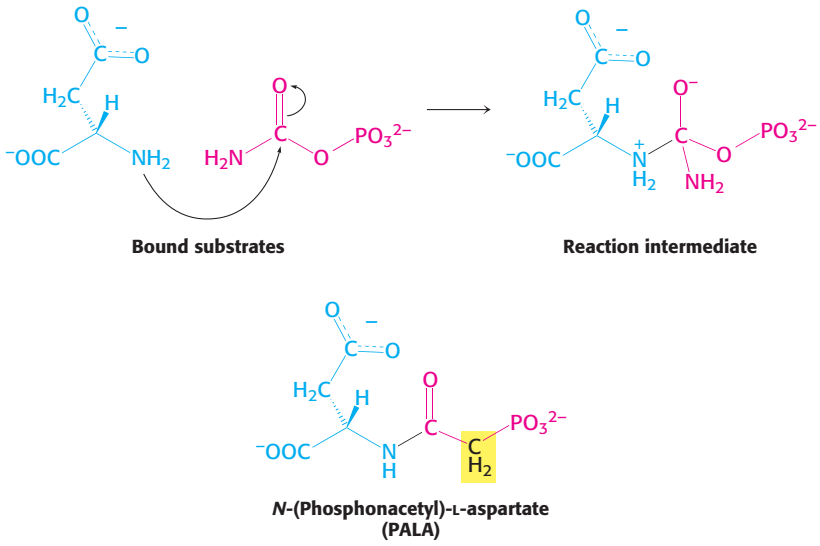


Figure 10.7 PALA, a bisubstrate analog. (Top) Nucleophilic attack by the amino group of aspartate on the carbonyl carbon atom of carbamoyl phosphate generates an intermediate on the pathway to the formation of *N*-carbamoylaspartate. (Bottom) *N*-(Phosphonacetyl)-L-aspartate (PALA) is an analog of the reaction intermediate and a potent competitive inhibitor of aspartate transcarbamoylase.

regulatory subunits: each r chain within a regulatory dimer interacts with a c chain within a catalytic trimer. The c chain makes contact with a structural domain in the r chain that is stabilized by a zinc ion bound to four cysteine residues. The mercurial compound *p*-hydroxymercuribenzoate is able to dissociate the catalytic and regulatory subunits because mercury binds strongly to the cysteine residues, displacing the zinc and destabilizing this r-subunit domain.

To locate the active sites, the enzyme was crystallized in the presence of *N*-(phosphonacetyl)-L-aspartate (PALA), a bisubstrate analog (an analog of the two substrates) that resembles an intermediate along the pathway of catalysis (Figure 10.7). PALA is a potent competitive inhibitor of ATCase; it binds to the active sites and blocks them. The structure of the ATCase–PALA complex reveals that PALA binds at sites lying at the boundaries between pairs of c chains within a catalytic trimer (Figure 10.8). Each catalytic trimer contributes three active sites to the complete enzyme. Further examination of the ATCase–PALA complex reveals a remarkable change in

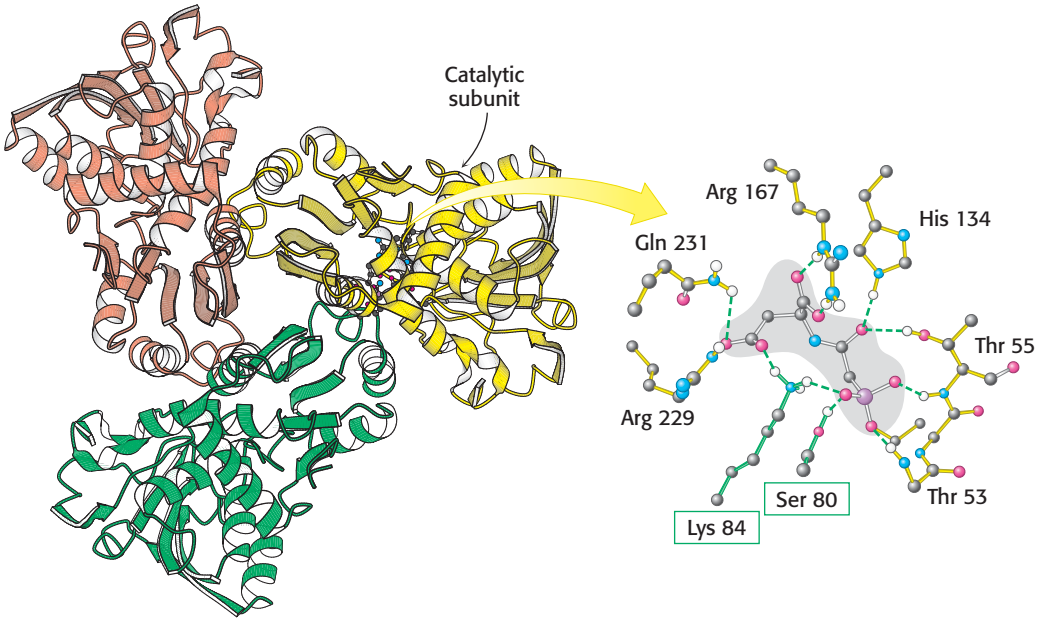


Figure 10.8 The active site of ATCase. Some of the crucial active-site residues are shown binding to the inhibitor PALA (shaded gray). *Notice* that the active site is composed mainly of residues from one c chain, but an adjacent c chain also contributes important residues (boxed in green). [Drawn from 8ATC.pdb.]

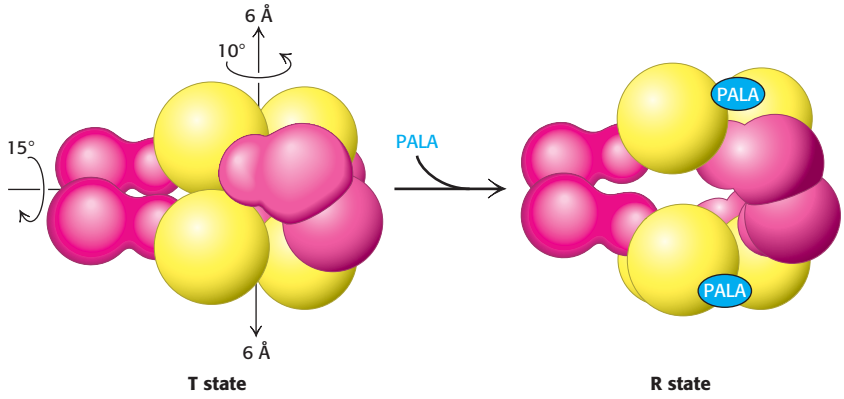


Figure 10.9 The T-to-R state transition in ATCase. Aspartate transcarbamoylase exists in two conformations: a compact, relatively inactive form called the tense (T) state and an expanded form called the relaxed (R) state. *Notice* that the structure of ATCase changes dramatically in the transition from the T state to the R State. PALA binding stabilizes the R state.

quaternary structure on binding of PALA. The two catalytic trimers move 12 Å farther apart and rotate approximately 10 degrees about their common threefold axis of symmetry. Moreover, the regulatory dimers rotate approximately 15 degrees to accommodate this motion (Figure 10.9). The enzyme literally expands on PALA binding. In essence, ATCase has two distinct quaternary forms: one that predominates in the absence of substrate or substrate analogs and another that predominates when substrates or analogs are bound. We call these forms the T (for tense) state and the R (for relaxed) state, respectively, as we did for the two quaternary states of hemoglobin.

How can we explain the enzyme's sigmoidal kinetics in light of the structural observations? Like hemoglobin, the enzyme exists in an equilibrium between the T state and the R state. In the absence of substrate, almost all the enzyme molecules are in the T state. The T state has a low affinity for substrate and hence shows a low catalytic activity. The occasional binding of a substrate molecule to one active site in an enzyme increases the likelihood that the entire enzyme shifts to the R state with its higher binding affinity. The addition of more substrate has two effects. First, it increases the probability that each enzyme molecule will bind at least one substrate molecule. Second, it increases the average number of substrate molecules bound to each enzyme. The presence of additional substrate will increase the fraction of enzyme molecules in the more active R state because *the position of the equilibrium depends on the number of active sites that are occupied by substrate.* We considered this property, called *cooperativity* because the subunits cooperate with one another, when we discussed the sigmoidal oxygen-binding curve of hemoglobin. The effects of substrates on allosteric enzymes are referred to as *homotropic effects* (from the Greek *homós,* "same").

This mechanism for allosteric regulation is referred to as the *concerted mechanism* because the change in the enzyme is "all or none"; the entire enzyme is converted from T into R, affecting all of the catalytic sites equally. In contrast, the *sequential model* assumes that the binding of ligand to one site on the complex can affect neighboring sites without causing all subunits to undergo the T-to-R transition. Although the concerted mechanism explains the behavior of ATCase well, most other allosteric enzymes have features of both models.

The sigmoidal curve for ATCase can be pictured as a composite of two Michaelis–Menten curves, one corresponding to the T state and the other to the R state. An increase in substrate concentration favors a transition from the T-state curve to the R-state curve (Figure 10.10). Note that such sigmoidal behavior has an additional consequence: in the concentration range at which the T-to-R transition is taking place, the curve depends quite steeply on the substrate concentration. The enzyme is switched from a less

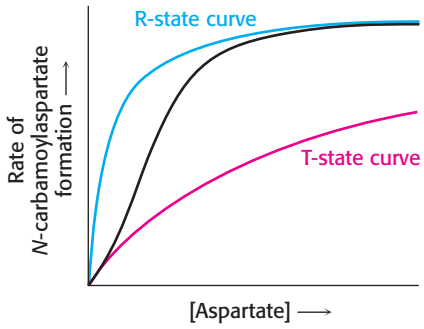


Figure 10.10 Basis for the sigmoidal curve. The generation of the sigmoidal curve by the property of cooperativity can be understood by imagining an allosteric enzyme as a mixture of two Michaelis–Menten enzymes, one with a high value of K_M that corresponds to the T state and another with a low value of K_M that corresponds to the R state. As the concentration of substrate is increased, the equilibrium shifts from the T state to the R state, which results in a steep rise in activity with respect to substrate concentration.

active state to a more active state within a narrow range of substrate concentration. This behavior is beneficial when a response to small changes in substrate concentration is physiologically important.

In studies of the isolated catalytic trimer, the catalytic subunit shows the hyperbolic curve characteristic of Michaelis–Menten kinetics, which is indistinguishable from the curve deduced for the R state (see Figure 10.10). Thus, the term *tense* is apt: in the T state, the regulatory dimers hold the two catalytic trimers sufficiently close to each other that key loops on their surfaces collide and interfere with conformational adjustments necessary for high-affinity substrate binding and catalysis.

Allosteric regulators modulate the T-to-R equilibrium

We now turn our attention to the effects of CTP. As noted earlier, CTP inhibits the action of ATCase. X-ray studies of ATCase in the presence of CTP revealed (1) that the enzyme is in the T state when bound to CTP and (2) that a binding site for this nucleotide exists in each regulatory chain in a domain that does not interact with the catalytic subunit (Figure 10.11). Each active site is more than 50 Å from the nearest CTP-binding site. The question naturally arises, How can CTP inhibit the catalytic activity of the enzyme when it does not interact with the catalytic chain?

The quaternary structural changes observed on substrate-analog binding suggest a mechanism for inhibition by CTP (Figure 10.12). *The binding of the inhibitor CTP shifts the equilibrium toward the T state, decreasing net enzyme activity*. The binding of CTP makes it more difficult for substrate binding to convert the enzyme into the R state. Consequently, CTP increases the initial phase of the sigmoidal curve (Figure 10.13). More substrate is required to attain a given reaction rate.

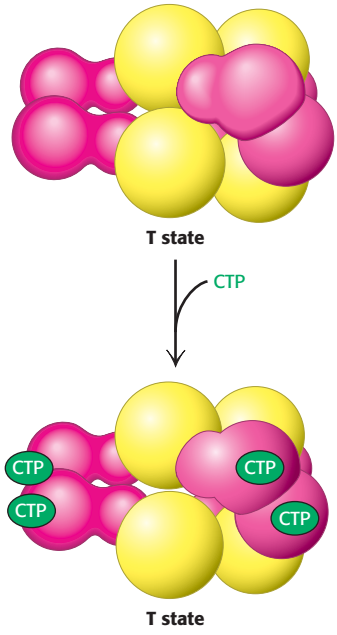


Figure 10.11 CTP stabilizes the T state. The binding of CTP to the regulatory subunit of aspartate transcarbamoylase stabilizes the T state.

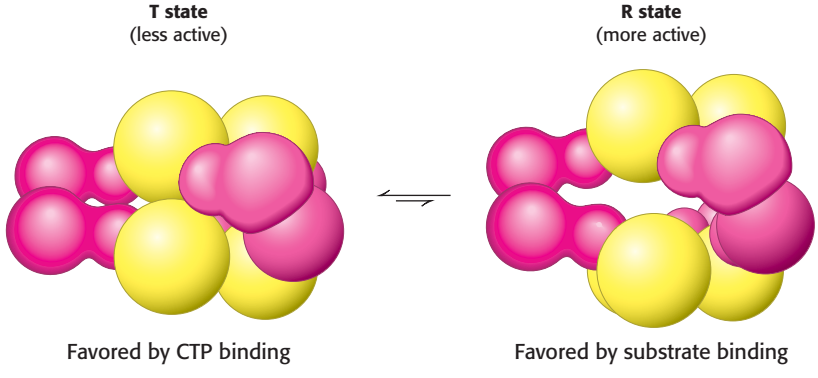


Figure 10.12 The R state and the T state are in equilibrium. Even in the absence of any substrate or regulators, aspartate transcarbamoylase exists in equilibrium between the R and the T states. Under these conditions, the T state is favored by a factor of approximately 200.

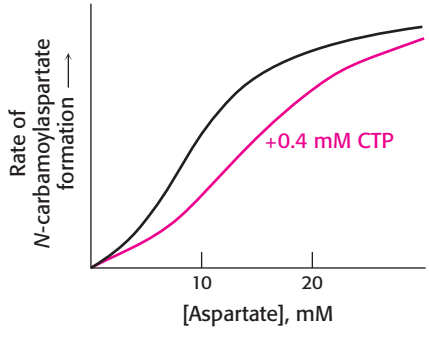


Figure 10.13 Effect of CTP on ATCase kinetics. Cytidine triphosphate (CTP) stabilizes the T state of aspartate transcarbamoylase, making it more difficult for substrate binding to convert the enzyme into the R state. As a result, the curve is shifted to the right, as shown in red.

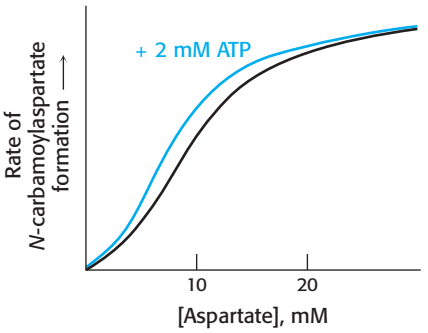


Figure 10.14 Effect of ATP on ATCase kinetics. ATP is an allosteric activator of aspartate transcarbamoylase because it stabilizes the R state, making it easier for substrate to bind. As a result, the curve is shifted to the left, as shown in blue.

Interestingly, ATP, too, is an allosteric effector of ATCase. However, the effect of ATP is to *increase* the reaction rate at a given aspartate concentration (Figure 10.14). At high concentrations of ATP, the kinetic profile shows a less-pronounced sigmoidal behavior. ATP competes with CTP for binding to regulatory sites. Consequently, high levels of ATP prevent CTP from inhibiting the enzyme. The effects of nonsubstrate molecules on allosteric enzymes (such as those of CTP and ATP on ATCase) are referred to as *heterotropic effects* (from the Greek *héteros,* "different"). Substrates generate the sigmoidal curve (homotropic effects), whereas regulators shift the K_M (heterotropic effects). Note, however, that both types of effect are generated by altering the T/R ratio.

The increase in ATCase activity in response to increased ATP concentration has two potential physiological explanations. First, high ATP concentration signals a high concentration of purine nucleotides in the cell; the increase in ATCase activity will tend to balance the purine and pyrimidine pools. Second, a high concentration of ATP indicates that energy is available for mRNA synthesis and DNA replication and leads to the synthesis of pyrimidines needed for these processes.

The Appendix to Chapter 7 includes a quantitative description of the concerted model. Although developed to describe a binding process, the model also applies to enzyme activity because the fraction of enzymatic active sites with substrate bound is proportional to enzymatic activity. A key aspect of this model is the equilibrium between the T and the R states. We defined L as the equilibrium constant between the R and the T forms.

$$\mathrm{R} \rightleftharpoons \mathrm{T} \qquad L = \frac{\mathrm{T}}{\mathrm{R}}$$

The effects of CTP and ATP can be modeled simply by changing the value of L. For the CTP-saturated form, the value of L increases from 200 to 1250. Thus, more substrate is required to shift the equilibrium appreciably to the R form. For the ATP-saturated form, the value of L decreases to 70 (Figure 10.15).

Thus, the concerted model provides us with a good description of the kinetic behavior of ATCase in the presence of its key regulators.

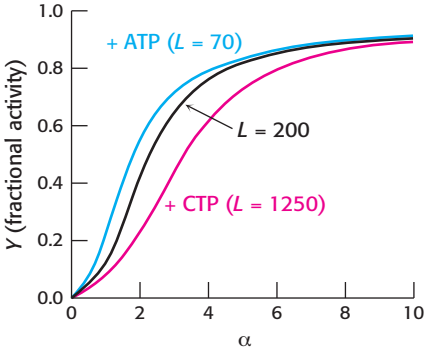


Figure 10.15 Quantitative description of the concerted model. In this description of the concerted model, fractional activity, Y, is the fraction of active sites bound to substrate and is directly proportional to reaction velocity; α is the ratio of [S] to the dissociation constant of S with the enzyme in the R state; and L is the ratio of the concentration of enzyme in the T state to that in the R state. The binding of the regulators ATP and CTP to ATCase changes the value of L and thus the response to substrate concentration. To construct these curves, the formula describing the concerted model in the appendix to Chapter 7 was used, with $c = 0.1$ and $n = 6$.

10.2 Isozymes Provide a Means of Regulation Specific to Distinct Tissues and Developmental Stages

Isozymes, or *isoenzymes,* are enzymes that differ in amino acid sequence yet catalyze the same reaction. Usually, these enzymes display different kinetic parameters, such as K_M, or respond to different regulatory molecules. They are encoded by different genes, which usually arise through gene duplication

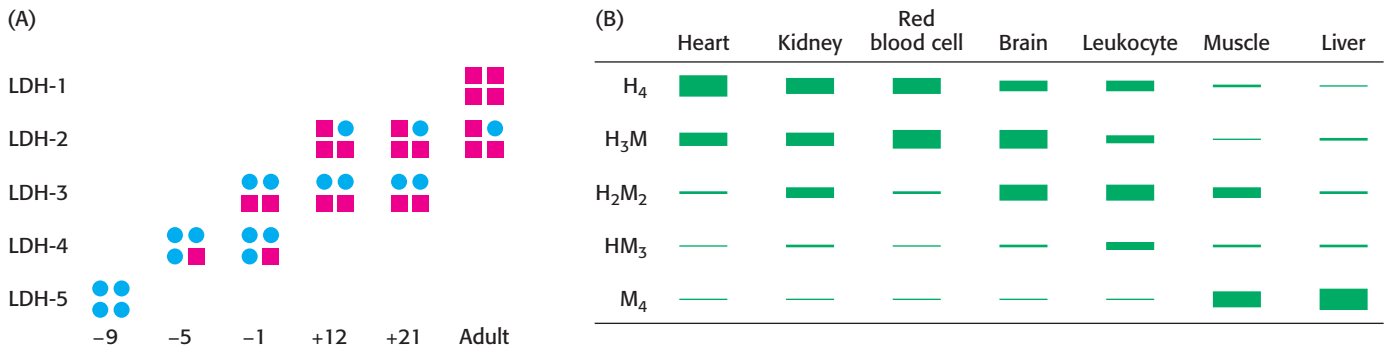


Figure 10.16 Isozymes of lactate dehydrogenase. (A) The rat heart lactate dehydrogenase (LDH) isozyme profile changes in the course of development. The H isozyme is represented by squares and the M isozyme by circles. The negative and positive numbers denote the days before and after birth, respectively. (B) LDH isozyme content varies by tissue. [(A) After W.-H. Li, *Molecular Evolution* (Sinauer, 1997), p. 283; (B) after K. Urich, *Comparative Animal Biochemistry* (Springer Verlag, 1990), p. 542.]

and divergence. Isozymes can often be distinguished from one another by biochemical properties such as electrophoretic mobility.

The existence of isozymes permits the fine-tuning of metabolism to meet the needs of a given tissue or developmental stage. Consider the example of lactate dehydrogenase (LDH), an enzyme that catalyzes a step in anaerobic glucose metabolism and glucose synthesis. Human beings have two isozymic polypeptide chains for this enzyme: the H isozyme is highly expressed in heart muscle and the M isozyme is expressed in skeletal muscle. The amino acid sequences are 75% identical. Each functional enzyme is tetrameric, and many different combinations of the two isozymic polypeptide chains are possible. The H_4 isozyme, found in the heart, has a higher affinity for substrates than does the M_4 isozyme. The two isozymes also differ in that high levels of pyruvate allosterically inhibit the H_4 but not the M_4 isozyme. The other combinations, such as H_3M, have intermediate properties. We will consider these isozymes in their biological context in Chapter 16.

The M_4 isozyme functions optimally in the anaerobic environment of hard-working skeletal muscle, whereas the H_4 isozyme does so in the aerobic environment of heart muscle. Indeed, the proportions of these isozymes change throughout the development of the rat heart as the tissue switches from an anaerobic environment to an aerobic one (Figure 10.16A). Figure 10.16B shows the tissue-specific forms of lactate dehydrogenase in adult rat tissues.

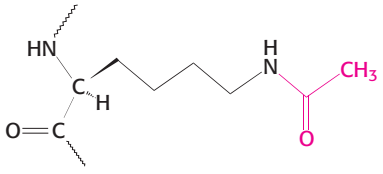
The appearance of some isozymes in the blood is a sign of tissue damage, useful for clinical diagnosis. For instance, an increase in serum levels of H_4 relative to H_3M is an indication that a myocardial infarction, or heart attack, has damaged heart-muscle cells, leading to the release of cellular material.

10.3 Covalent Modification Is a Means of Regulating Enzyme Activity

The covalent attachment of a molecule to an enzyme or protein can modify its activity. In these instances, a donor molecule provides the functional moiety being attached. Most modifications are reversible. Phosphorylation and dephosphorylation are the most common means of covalent modification. The attachment of acetyl groups and their removal are another common means. Histones—proteins that are packaged with DNA into chromosomes—are extensively acetylated and deacetylated in vivo on lysine residues (Section 31.3). More heavily acetylated histones are associated with genes that are being actively transcribed. The acetyltransferase

Table 10.1 Common covalent modifications of protein activity

Modification	Donor molecule	Example of modified protein	Protein function
Phosphorylation	ATP	Glycogen phosphorylase	Glucose homeostasis; energy transduction
Acetylation	Acetyl CoA	Histones	DNA packing; transcription
Myristoylation	Myristoyl CoA	Src	Signal transduction
ADP ribosylation	NAD^+	RNA polymerase	Transcription
Farnesylation	Farnesyl pyrophosphate	Ras	Signal transduction
γ-Carboxylation	HCO_3^-	Thrombin	Blood clotting
Sulfation	3′-Phosphoadenosine-5′-phosphosulfate	Fibrinogen	Blood-clot formation
Ubiquitination	Ubiquitin	Cyclin	Control of cell cycle



Acetylated lysine

and deacetylase enzymes are themselves regulated by phosphorylation, showing that the covalent modification of a protein can be controlled by the covalent modification of the modifying enzymes.

Modification is not readily reversible in some cases. The irreversible attachment of a lipid group causes some proteins in signal-transduction pathways, such as Ras (a GTPase) and Src (a protein tyrosine kinase), to become affixed to the cytoplasmic face of the plasma membrane. Fixed in this location, the proteins are better able to receive and transmit information that is being passed along their signaling pathways (Chapter 14). Mutations in both Ras and Src are seen in a wide array of cancers. The attachment of the small protein ubiquitin can signal that a protein is to be destroyed, the ultimate means of regulation (Chapter 23). The protein cyclin must be ubiquitinated and destroyed before a cell can enter anaphase and proceed through the cell cycle.

Virtually all the metabolic processes that we will examine are regulated in part by covalent modification. Indeed, the allosteric properties of many enzymes are modified by covalent modification. Table 10.1 lists some of the common covalent modifications.

Kinases and phosphatases control the extent of protein phosphorylation

We will see phosphorylation used as a regulatory mechanism in virtually every metabolic process in eukaryotic cells. Indeed, as much as 30% of eukaryotic proteins are phosphorylated. The enzymes catalyzing phosphorylation reactions are called *protein kinases*. These enzymes constitute one of the largest protein families known: there are more than 100 homologous protein kinases in yeast and more than 500 in human beings. This multiplicity of enzymes allows regulation to be fine-tuned according to a specific tissue, time, or substrate.

ATP is the most common donor of phosphoryl groups. The terminal (γ) phosphoryl group of ATP is transferred to a specific amino acid of the acceptor protein or enzyme. In eukaryotes, the acceptor residue is commonly one of the three containing a hydroxyl group in its side chain. Transfers to *serine* and *threonine* residues are handled by one class of protein kinases and to *tyrosine* residues by another. Tyrosine kinases, which are unique to multicellular organisms, play pivotal roles in growth regulation, and mutations in these enzymes are commonly observed in cancer cells.

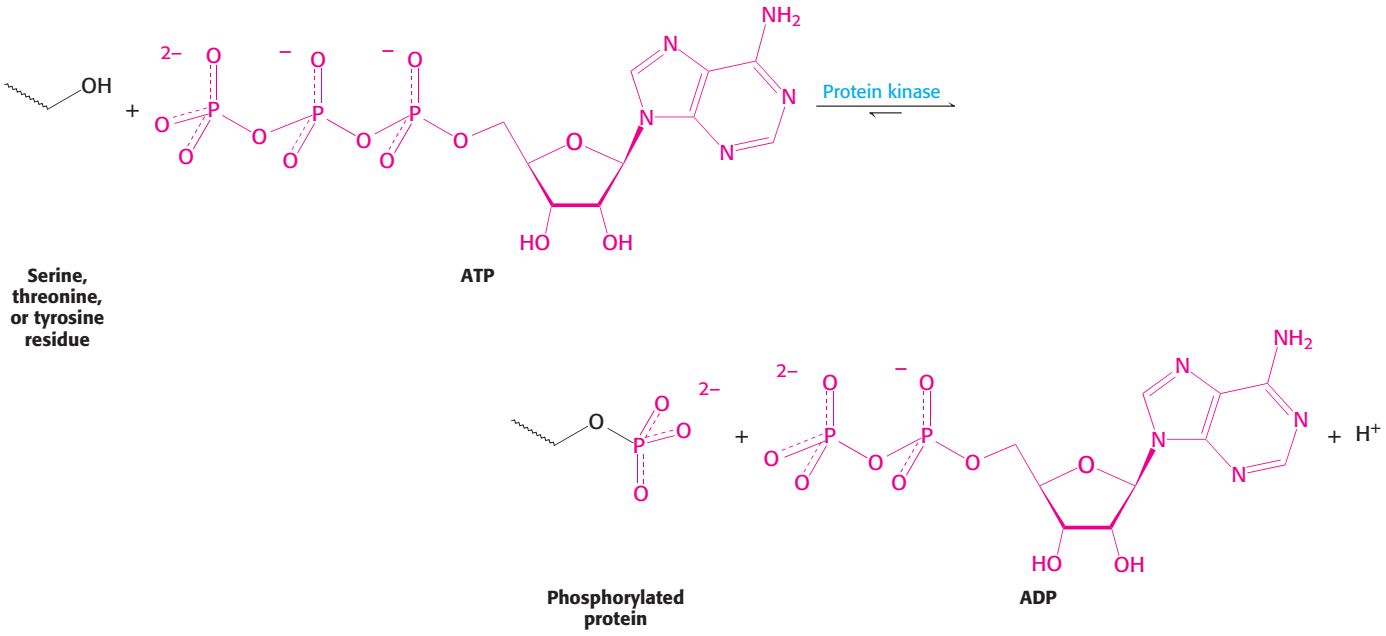


Table 10.2 lists a few of the known serine and threonine protein kinases. The acceptors in protein-phosphorylation reactions are located inside cells, where the phosphoryl-group donor ATP is abundant. Proteins that are entirely extracellular are not regulated by reversible phosphorylation.

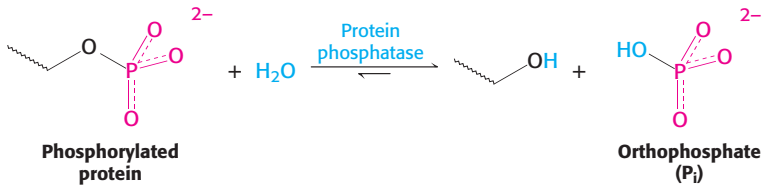
Protein kinases vary in their degree of specificity. *Dedicated protein kinases* phosphorylate a single protein or several closely related ones. *Multifunctional protein kinases* modify many different targets; they have a wide reach and can coordinate diverse processes. Comparisons of amino acid sequences of many phosphorylation sites show that a multifunctional kinase recognizes related sequences. For example, the *consensus sequence* recognized by protein kinase A is Arg-Arg-X-*Ser*-Z or Arg-Arg-X-*Thr*-Z, in which X is a small residue, Z is a large hydrophobic one, and *Ser* or *Thr* is the site of phosphorylation. However, this sequence is not absolutely required. Lysine, for example, can substitute for one of the arginine residues but with some loss of affinity. Short synthetic peptides containing a consensus motif are nearly always phosphorylated by serine–threonine protein kinases. Thus, *the primary determinant of specificity is the amino acid sequence surrounding the serine or threonine phosphorylation site.* However, distant residues can contribute to specificity. For instance, a change in protein conformation can open or close access to a possible phosphorylation site.

Table 10.2 Examples of serine and threonine kinases and their activating signals

Signal	Enzyme
Cyclic nucleotides	Cyclic AMP-dependent protein kinase Cyclic GMP-dependent protein kinase
Ca^{2+} and calmodulin	Ca^{2+}–calmodulin protein kinase Phosphorylase kinase or glycogen synthase kinase 2
AMP	AMP-activated kinase
Diacylglycerol	Protein kinase C
Metabolic intermediates and other "local" effectors	Many target-specific enzymes, such as pyruvate dehydrogenase kinase and branched-chain ketoacid dehydrogenase kinase

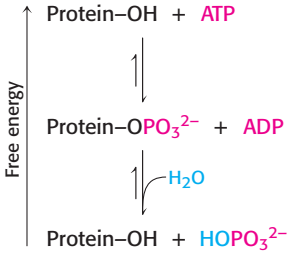
Source: After D. Fell, *Understanding the Control of Metabolism* (Portland Press, 1997), Table 7.2.

Protein phosphatases reverse the effects of kinases by catalyzing the removal of phosphoryl groups attached to proteins. The enzyme hydrolyzes the bond attaching the phosphoryl group.



The unmodified hydroxyl-containing side chain is regenerated and orthophosphate (P_i) is produced. These enzymes play a vital role in cells because they turn off the signaling pathways that are activated by kinases. One class of highly conserved phosphatase called PP2A suppresses the cancer-promoting activity of certain kinases.

Importantly, the phosphorylation and dephosphorylation reactions are not the reverse of one another; each is essentially irreversible under physiological conditions. Furthermore, both reactions take place at negligible rates in the absence of enzymes. Thus, phosphorylation of a protein substrate will take place only through the action of a specific protein kinase and at the expense of ATP cleavage, and dephosphorylation will take place only through the action of a phosphatase. The result is that target proteins cycle unidirectionally between unphosphorylated and phosphorylated forms. The rate of cycling between the phosphorylated and the dephosphorylated states depends on the relative activities of kinases and phosphatases.



Phosphorylation is a highly effective means of regulating the activities of target proteins

Phosphorylation is a common covalent modification of proteins in all forms of life, which leads to the question, What makes protein phosphorylation so valuable in regulating protein function that its use is ubiquitous? Phosphorylation is a highly effective means of controlling the activity of proteins for several reasons:

1. *The free energy of phosphorylation is large.* Of the -50 kJ mol^{-1} (-12 kcal mol^{-1}) provided by ATP, about half is consumed in making phosphorylation irreversible; the other half is conserved in the phosphorylated protein. A free-energy change of 5.69 kJ mol^{-1} (1.36 kcal mol^{-1}) corresponds to a factor of 10 in an equilibrium constant. Hence, phosphorylation can change the conformational equilibrium between different functional states by a large factor, of the order of 10^4. In essence, the energy expenditure allows for a stark shift from one state to another.

2. *A phosphoryl group adds two negative charges to a modified protein.* These new charges may disrupt electrostatic interactions in the unmodified protein and allow new electrostatic interactions to be formed. Such structural changes can markedly alter substrate binding and catalytic activity.

3. *A phosphoryl group can form three or more hydrogen bonds.* The tetrahedral geometry of a phosphoryl group makes these bonds highly directional, allowing for specific interactions with hydrogen-bond donors.

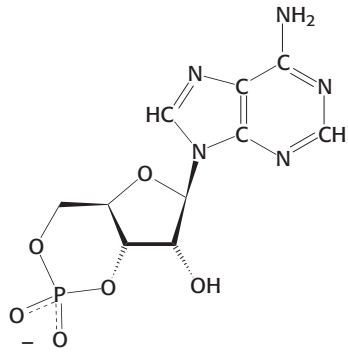
4. *Phosphorylation and dephosphorylation can take place in less than a second or over a span of hours.* The kinetics can be adjusted to meet the timing needs of a physiological process.

5. *Phosphorylation often evokes highly amplified effects.* A single activated kinase can phosphorylate hundreds of target proteins in a short interval. If the target protein is an enzyme, it can in turn transform a large number of substrate molecules.

6. *ATP is the cellular energy currency* (Chapter 15). The use of this compound as a phosphoryl-group donor links the energy status of the cell to the regulation of metabolism.

Cyclic AMP activates protein kinase A by altering the quaternary structure

Let us examine a specific protein kinase that helps animals cope with stressful situations. The "flight or fight" response is common to many animals presented with a dangerous or exciting situation. Muscle becomes primed for action. This priming is the result of the activity of a particular protein kinase. In this case, the hormone epinephrine (adrenaline) triggers the formation of cyclic AMP (cAMP), an intracellular messenger formed by the cyclization of ATP. Cyclic AMP subsequently activates a key enzyme: *protein kinase A* (PKA). The kinase alters the activities of target proteins by phosphorylating specific serine or threonine residues. The striking finding is that *most effects of cAMP in eukaryotic cells are achieved through the activation by cAMP of PKA.*



Cyclic adenosine monophosphate (cAMP)

PKA provides a clear example of the integration of allosteric regulation and phosphorylation. PKA is activated by cAMP concentrations near 10 nM. The activation mechanism is reminiscent of that of aspartate transcarbamoylase. Like that enzyme, PKA in muscle consists of two kinds of subunits: a 49-kd regulatory (R) subunit and a 38-kd catalytic (C) subunit. In the absence of cAMP, the regulatory and catalytic subunits form an R_2C_2 complex that is enzymatically inactive (Figure 10.17). The binding of two molecules of cAMP to each of the regulatory subunits leads to the dissociation of R_2C_2 into an R_2 subunit and two C subunits. These free catalytic subunits are then enzymatically active. Thus, *the binding of cAMP to the regulatory subunit relieves its inhibition of the catalytic subunit.* PKA and most other kinases exist in isozymic forms for fine-tuning regulation to meet the needs of a specific cell or developmental stage.

How does the binding of cAMP activate the kinase? Each R chain contains the sequence Arg-Arg-Gly-*Ala*-Ile, which matches the consensus sequence for phosphorylation except for the presence of alanine in place of serine. In the R_2C_2 complex, this *pseudosubstrate sequence* of R occupies the catalytic site of C, thereby preventing the entry of protein substrates (see Figure 10.17). The binding of cAMP to the R chains allosterically moves the pseudosubstrate sequences out of the catalytic sites. The released C chains are then free to bind and phosphorylate substrate proteins.

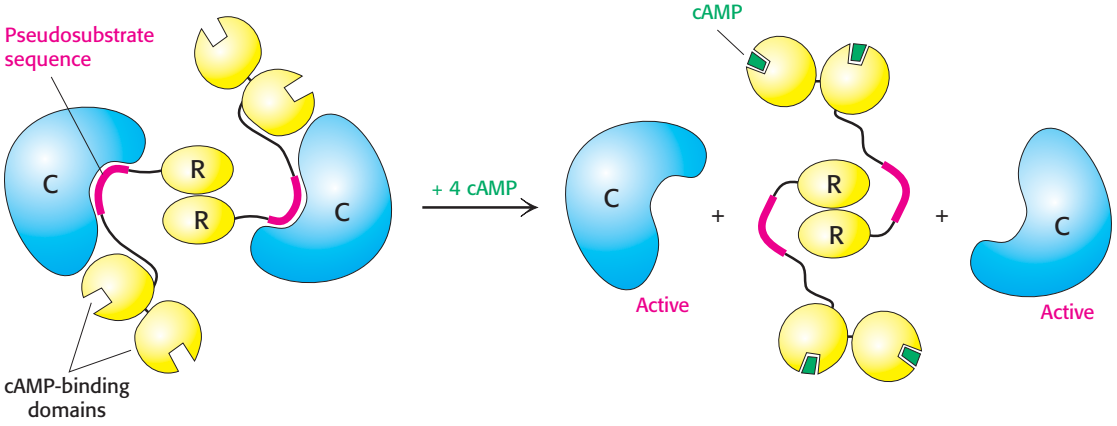


Figure 10.17 Regulation of protein kinase A. The binding of four molecules of cAMP activates protein kinase A by dissociating the inhibited holoenzyme (R_2C_2) into a regulatory subunit (R_2) and two catalytically active subunits (C). Each R chain includes cAMP-binding domains and a pseudosubstrate sequence.

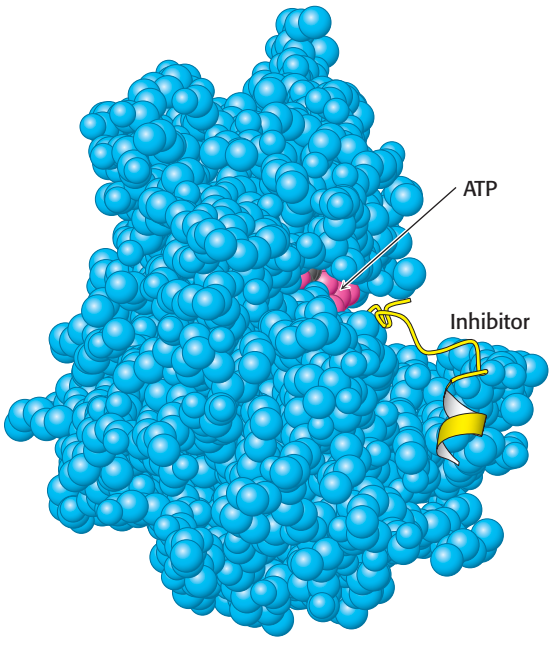


Figure 10.18 Protein kinase A bound to an inhibitor. This space-filling model shows a complex of the catalytic subunit of protein kinase A with an inhibitor bearing a pseudosubstrate sequence. *Notice* that the inhibitor (yellow) binds to the active site, a cleft between the domains of the enzyme. The bound ATP, shown in red, is in the active site adjacent to the site to which the inhibitor is bound. [Drawn from 1ATP.pdb.]

ATP and the target protein bind to a deep cleft in the catalytic subunit of protein kinase A

X-ray crystallography revealed the three-dimensional structure of the catalytic subunit of PKA bound to ATP and a 20-residue peptide inhibitor. The 350-residue catalytic subunit of PKA has two lobes (Figure 10.18). ATP and part of the inhibitor fill a deep cleft between the lobes. The smaller lobe makes many contacts with ATP–Mg^{2+}, whereas the larger lobe binds the peptide and contributes the key catalytic residues. As with other kinases, the two lobes move closer to one another on substrate binding; mechanisms that restrict this domain closure provide a means of regulating protein kinase activity. *The PKA structure has broad significance because residues 40 to 280 constitute a conserved catalytic core that is common to essentially all known protein kinases.* We see here an example of a successful biochemical solution to a problem (in this case, protein phosphorylation) being employed many times in the course of evolution.

The bound peptide in this crystal occupies the active site because it contains the pseudosubstrate sequence Arg-Arg-Asn-*Ala*-Ile (Figure 10.19). The structure of the complex reveals the interactions by which the enzyme recognizes the consensus sequence. The guanidinium group of the first arginine residue forms an ion pair with the carboxylate side chain of a glutamate residue (Glu 127) of the enzyme. The second arginine likewise interacts with two other carboxylate groups. The nonpolar side chain of isoleucine, which matches Z in the consensus sequence (p. 299), fits snugly in a hydrophobic groove formed by two leucine residues of the enzyme.

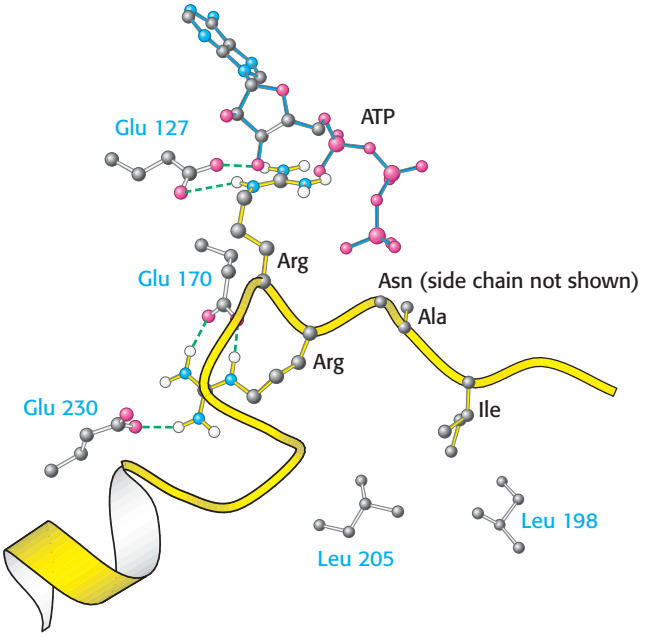


Figure 10.19 Binding of pseudosubstrate to protein kinase A. *Notice* that the inhibitor makes multiple contacts with the enzyme. The two arginine side chains of the pseudosubstrate form salt bridges with three glutamate carboxylate groups. Hydrophobic interactions also are important in the recognition of substrate. The isoleucine residue of the pseudosubstrate is in contact with a pair of leucine residues of the enzyme.

10.4 Many Enzymes Are Activated by Specific Proteolytic Cleavage

We turn now to a different mechanism of enzyme regulation. Many enzymes acquire full enzymatic activity as they spontaneously fold into their characteristic three-dimensional forms. In contrast, the folded forms

Table 10.3 Gastric and pancreatic zymogens

Site of synthesis	Zymogen	Active enzyme
Stomach	Pepsinogen	Pepsin
Pancreas	Chymotrypsinogen	Chymotrypsin
Pancreas	Trypsinogen	Trypsin
Pancreas	Procarboxypeptidase	Carboxypeptidase

of other enzymes are inactive until the cleavage of one or a few specific peptide bonds. The inactive precursor is called a *zymogen* or a *proenzyme*. An energy source such as ATP is not needed for cleavage. Therefore, in contrast with reversible regulation by phosphorylation, even proteins located outside cells can be activated by this means. Another noteworthy difference is that proteolytic activation, in contrast with allosteric control and reversible covalent modification, takes place just once in the life of an enzyme molecule.

Specific proteolysis is a common means of activating enzymes and other proteins in biological systems. For example:

1. The *digestive enzymes* that hydrolyze proteins are synthesized as zymogens in the stomach and pancreas (Table 10.3).

2. *Blood clotting* is mediated by a cascade of proteolytic activations that ensures a rapid and amplified response to trauma.

3. Some protein hormones are synthesized as inactive precursors. For example, *insulin* is derived from *proinsulin* by proteolytic removal of a peptide.

4. The fibrous protein *collagen,* the major constituent of skin and bone, is derived from *procollagen,* a soluble precursor.

5. Many *developmental processes* are controlled by the activation of zymogens. For example, in the metamorphosis of a tadpole into a frog, large amounts of collagen are resorbed from the tail in the course of a few days. Likewise, much collagen is broken down in a mammalian uterus after delivery. The conversion of *procollagenase* into *collagenase,* the active protease, is precisely timed in these remodeling processes.

6. *Programmed cell death,* or *apoptosis,* is mediated by proteolytic enzymes called *caspases,* which are synthesized in precursor form as *procaspases.* When activated by various signals, caspases function to cause cell death in most organisms, ranging from *C. elegans* to human beings. Apoptosis provides a means of sculpting the shapes of body parts in the course of development and a means of eliminating damaged or infected cells.

We next examine the activation and control of zymogens, using as examples several digestive enzymes as well as blood-clot formation.

Chymotrypsinogen is activated by specific cleavage of a single peptide bond

Chymotrypsin is a digestive enzyme that hydrolyzes proteins in the small intestine. Its mechanism of action was described in detail in Chapter 9. Its inactive precursor, *chymotrypsinogen,* is synthesized in the pancreas, as are several other zymogens and digestive enzymes. Indeed, the pancreas is one of the most active organs in synthesizing and secreting proteins. The enzymes and zymogens are synthesized in the acinar cells of the pancreas

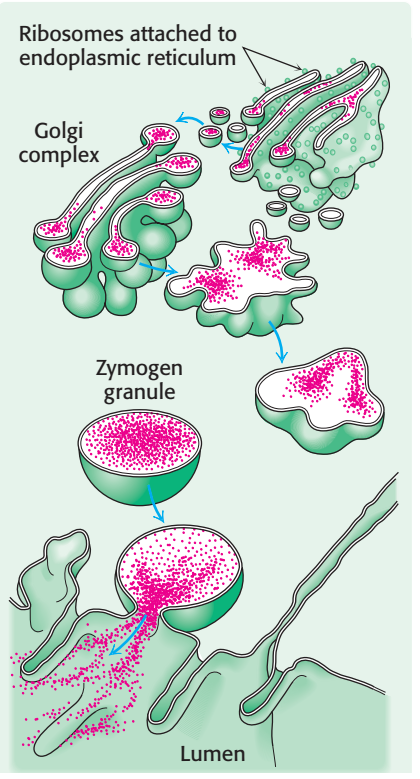


Figure 10.20 Secretion of zymogens by an acinar cell of the pancreas. Zymogens are synthesized on ribosomes attached to the endoplasmic reticulum. They are subsequently processed in the Golgi apparatus and packaged into zymogen or secretory granules. With the proper signal, the granules fuse with the plasma membrane, discharging their contents into the lumen of the pancreatic ducts. Cell cytoplasm is depicted as pale green. Membranes and lumen are shown as dark green.

and stored inside membrane-bounded granules (Figure 10.20). The zymogen granules accumulate at the apex of the acinar cell; when the cell is stimulated by a hormonal signal or a nerve impulse, the contents of the granules are released into a duct leading into the duodenum.

Chymotrypsinogen, a single polypeptide chain consisting of 245 amino acid residues, is virtually devoid of enzymatic activity. It is converted into a fully active enzyme when the peptide bond joining arginine 15 and isoleucine 16 is cleaved by trypsin (Figure 10.21). The resulting active enzyme, called π-chymotrypsin, then acts on other π-chymotrypsin molecules by removing two dipeptides to yield α-chymotrypsin, the stable form of the enzyme. The three resulting chains in α-chymotrypsin remain linked to one another by two interchain disulfide bonds. The striking feature of this activation process is that *cleavage of a single specific peptide bond transforms the protein from a catalytically inactive form into one that is fully active.*

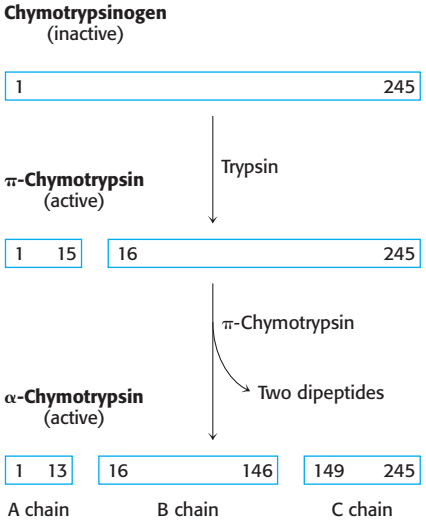


Figure 10.21 Proteolytic activation of chymotrypsinogen. The three chains of α-chymotrypsin are linked by two interchain disulfide bonds (A to B, and B to C).

Proteolytic activation of chymotrypsinogen leads to the formation of a substrate-binding site

How does cleavage of a single peptide bond activate the zymogen? The cleavage of the peptide bond between amino acids 15 and 16 triggers key conformational changes, which were revealed by the elucidation of the three-dimensional structure of chymotrypsinogen.

1. The newly formed *amino-terminal group of isoleucine 16 turns inward and forms an ionic bond with aspartate 194* in the interior of the chymotrypsin molecule (Figure 10.22).

2. This electrostatic interaction triggers a number of conformational changes. Methionine 192 moves from a deeply buried position in the zymogen to the surface of the active enzyme, and residues 187 and 193 move

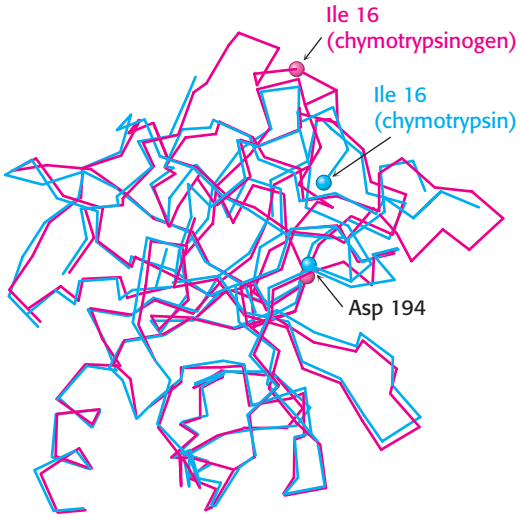


Figure 10.22 Conformations of chymotrypsinogen (red) and chymotrypsin (blue). *Notice* the alteration of the position of isoleucine 16 in chymotrypsin. The electrostatic interaction between the α-amino group of isoleucine 16 and the carboxylate of aspartate 194, essential for the structure of active chymotrypsin, is possible only in chymotrypsin. [Drawn from 1GCT.pdb and 2GCA.pdb.]

farther apart from each other. These changes result in the formation of the *substrate-specificity site* for aromatic and bulky nonpolar groups. One side of this site is made up of residues 189 through 192. *This cavity for binding part of the substrate is not fully formed in the zymogen.*

3. The tetrahedral transition state in catalysis by chymotrypsin is stabilized by hydrogen bonds between the negatively charged carbonyl oxygen atom of the substrate and two NH groups of the main chain of the enzyme (see Figure 9.8). One of these NH groups is not appropriately located in chymotrypsinogen, and so *the oxyanion hole is incomplete in the zymogen.*

4. The conformational changes elsewhere in the molecule are very small. Thus, *the switching on of enzymatic activity in a protein can be accomplished by discrete, highly localized conformational changes that are triggered by the hydrolysis of a single peptide bond.*

The generation of trypsin from trypsinogen leads to the activation of other zymogens

The structural changes accompanying the activation of *trypsinogen,* the precursor of the proteolytic enzyme *trypsin,* are somewhat different from those in the activation of chymotrypsinogen. X-ray analyses have shown that the conformation of four stretches of polypeptide, constituting about 15% of the molecule, changes markedly on activation. *These regions are very flexible in the zymogen, whereas they have a well-defined conformation in trypsin.* Furthermore, the oxyanion hole in trypsinogen is too far from histidine 57 to promote the formation of the tetrahedral transition state.

The digestion of proteins in the duodenum requires the concurrent action of several proteolytic enzymes, because each is specific for a limited number of side chains. Thus, the zymogens must be switched on at the same time. Coordinated control is achieved by the action of *trypsin as the common activator of all the pancreatic zymogens*—trypsinogen, chymotrypsinogen, proelastase, procarboxypeptidase, and prolipase, a lipid degrading enzyme. To produce active trypsin, the cells that line the duodenum secrete an enzyme, *enteropeptidase,* which hydrolyzes a unique lysine–isoleucine peptide bond in trypsinogen as the zymogen enters the duodenum from the pancreas. The small amount of trypsin produced in this way activates more trypsinogen and the other zymogens (Figure 10.23). Thus, *the formation of trypsin by enteropeptidase is the master activation step.*

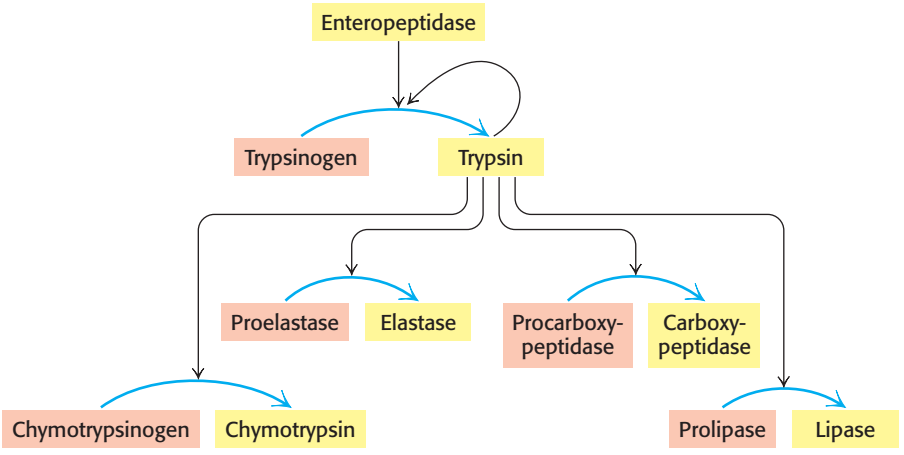


Figure 10.23 Zymogen activation by proteolytic cleavage. Enteropeptidase initiates the activation of the pancreatic zymogens by activating trypsin, which then activates other zymogens. Active enzymes are shown in yellow; zymogens are shown in orange.

Some proteolytic enzymes have specific inhibitors

The conversion of a zymogen into a protease by cleavage of a single peptide bond is a precise means of switching on enzymatic activity. However, this activation step is irreversible, and so a different mechanism is needed to stop proteolysis. Specific protease inhibitors accomplish this task. For example, *pancreatic trypsin inhibitor,* a 6-kd protein, inhibits trypsin by binding very tightly to its active site. The dissociation constant of the complex is 0.1 pM, which corresponds to a standard free energy of binding of about -75 kJ mol^{-1} (-18 kcal mol^{-1}). In contrast with nearly all known protein assemblies, this complex is not dissociated into its constituent chains by treatment with denaturing agents such as 8 M urea or 6 M guanidine hydrochloride.

The reason for the exceptional stability of the complex is that pancreatic trypsin inhibitor is a very effective substrate analog. X-ray analyses showed that the inhibitor lies in the active site of the enzyme, positioned such that the side chain of lysine 15 of this inhibitor interacts with the aspartate side chain in the specificity pocket of trypsin. In addition, there are many hydrogen bonds between the main chain of trypsin and that of its inhibitor. Furthermore, the carbonyl group of lysine 15 and the surrounding atoms of the inhibitor fit snugly in the active site of the enzyme. Comparison of the structure of the inhibitor bound to the enzyme with that of the free inhibitor reveals that *the structure is essentially unchanged on binding to the enzyme* (Figure 10.24). Thus, the inhibitor is preorganized into a structure that is highly complementary to the enzyme's active site. Indeed, the peptide bond between lysine 15 and alanine 16 in pancreatic trypsin inhibitor is cleaved but at a very slow rate: the half-life of the trypsin–inhibitor complex is several months. In essence, the inhibitor is a substrate, but its intrinsic structure is so nicely complementary to the enzyme's active site that it binds very tightly, rarely progressing to the transition state and is turned over slowly.

The amount of trypsin is much greater than the amount of inhibitor. Why does trypsin inhibitor exist? Recall that trypsin activates other zymogens. Consequently, the prevention of even small amounts of trypsin from initiating the inappropriately activated cascade prematurely is vital. Trypsin inhibitor binds to trypsin molecules in the pancreas or pancreatic ducts. This inhibition prevents severe damage to those tissues, which could lead to acute pancreatitis.

Pancreatic trypsin inhibitor is not the only important protease inhibitor. *α_1-Antitrypsin* (also called *α_1-antiproteinase*), a 53-kd plasma protein,

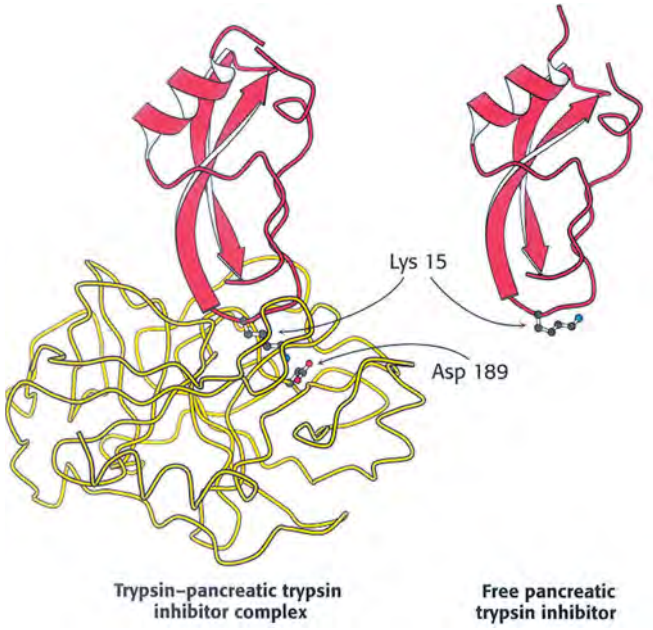


Figure 10.24 Interaction of trypsin with its inhibitor. Structure of a complex of trypsin (yellow) and pancreatic trypsin inhibitor (red). *Notice* that lysine 15 of the inhibitor penetrates into the active site of the enzyme. There it forms a salt bridge with aspartate 189 in the active site. *Also notice* that bound inhibitor and the free inhibitor are almost identical in structure. [Drawn from 1BPI.pdb.]

protects tissues from digestion by elastase, a secretory product of neutrophils (white blood cells that engulf bacteria). *Antielastase* would be a more accurate name for this inhibitor, because it blocks *elastase* much more effectively than it blocks trypsin. Like pancreatic trypsin inhibitor, α_1-antitrypsin blocks the action of target enzymes by binding nearly irreversibly to their active sites. Genetic disorders leading to a deficiency of α_1-antitrypsin show that this inhibitor is physiologically important. For example, the substitution of lysine for glutamate at residue 53 in the type Z mutant slows the secretion of this inhibitor from liver cells. Serum levels of the inhibitor are about 15% of normal in people homozygous for this defect. The consequence is that excess elastase destroys alveolar walls in the lungs by digesting elastic fibers and other connective-tissue proteins.

The resulting clinical condition is called *emphysema* (also known as *destructive lung disease*). People with emphysema must breathe much harder than normal people to exchange the same volume of air because their alveoli are much less resilient than normal. Cigarette smoking markedly increases the likelihood that even a type Z heterozygote will develop emphysema. The reason is that smoke oxidizes methionine 358 of the inhibitor (Figure 10.25), a residue essential for binding elastase. Indeed, this methionine side chain is the bait that selectively traps elastase. The *methionine sulfoxide* oxidation product, in contrast, does not lure elastase, a striking consequence of the insertion of just one oxygen atom into a protein and a striking example of the effect of behavior on biochemistry. We will consider another protease inhibitor, antithrombin III, when we examine the control of blood clotting.

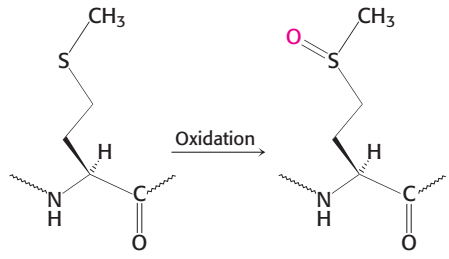


Figure 10.25 Oxidation of methionine to methionine sulfoxide.

Blood clotting is accomplished by a cascade of zymogen activations

Enzymatic cascades are often employed in biochemical systems to achieve a rapid response. In a cascade, an initial signal institutes a series of steps, each of which is catalyzed by an enzyme. At each step, the signal is amplified. For instance, if a signal molecule activates an enzyme that in turn activates 10 enzymes and each of the 10 enzymes in turn activates 10 additional enzymes, after four steps the original signal will have been amplified 10,000-fold. Blood clots are formed by a *cascade of zymogen activations:* the activated form of one clotting factor catalyzes the activation

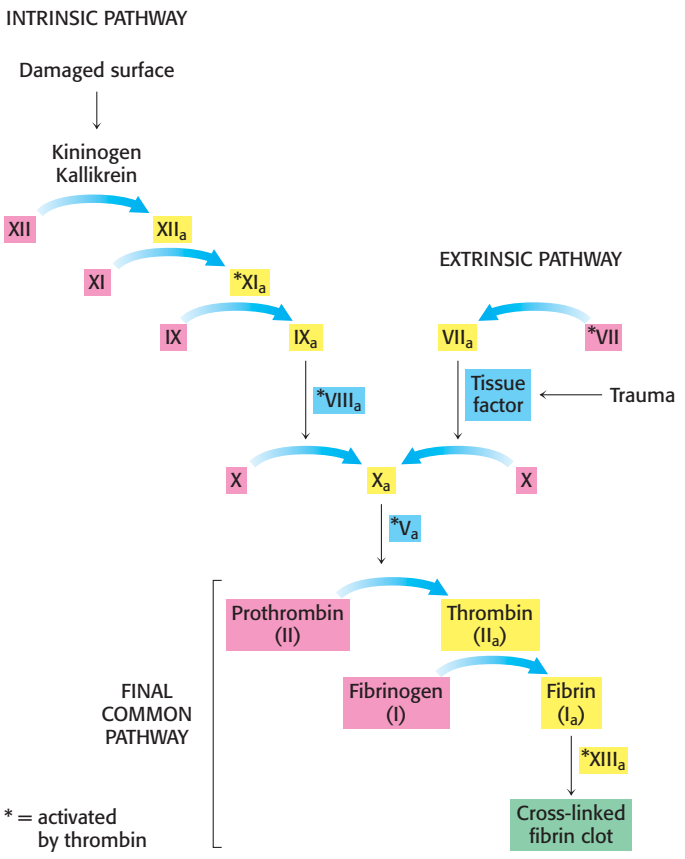


Figure 10.26 Blood-clotting cascade. A fibrin clot is formed by the interplay of the intrinsic, extrinsic, and final common pathways. The intrinsic pathway begins with the activation of factor XII (Hageman factor) by contact with abnormal surfaces produced by injury. The extrinsic pathway is triggered by trauma, which releases tissue factor (TF). TF forms a complex with VII, which initiates a cascade-activating thrombin. Inactive forms of clotting factors are shown in red; their activated counterparts (indicated by the subscript "a") are in yellow. Stimulatory proteins that are not themselves enzymes are shown in blue boxes. A striking feature of this process is that the activated form of one clotting factor catalyzes the activation of the next factor.

of the next (Figure 10.26). Thus, very small amounts of the initial factors suffice to trigger the cascade, ensuring a rapid response to trauma.

Two means of initiating blood clotting have been described, the *intrinsic pathway* and the *extrinsic pathway*. The intrinsic clotting pathway is activated by exposure of anionic surfaces on rupture of the endothelial lining of the blood vessels. The extrinsic pathway, which appears to be most crucial in blood clotting, is initiated when trauma exposes *tissue factor* (TF), an integral membrane glycoprotein. Shortly after the tissue factor is exposed, small amounts of *thrombin*, the key protease in clotting, are generated. Thrombin then amplifies the clotting process by activating enzymes and factors that lead to the generation of yet more thrombin, an example of positive feedback. The extrinsic and intrinsic pathways converge on a common sequence of final steps to form a clot composed of the protein fibrin (see Figure 10.26). Note that the active forms of the clotting factors are designated with a subscript "a," whereas factors that are activated by thrombin are designated with an asterisk.

Fibrinogen is converted by thrombin into a fibrin clot

The best-characterized part of the clotting process is the final step in the cascade: the conversion of *fibrinogen* into fibrin by thrombin, a proteolytic enzyme. Fibrinogen is made up of three globular units connected by two rods (Figure 10.27). This 340-kd protein consists of six chains: two each of Aα, Bβ, and γ. The rod regions are triple-stranded α-helical coiled coils, a recurring motif in proteins (Section 2.3). Thrombin cleaves four *arginine–glycine peptide bonds* in the central globular region of fibrinogen. On cleavage, an A peptide of 18 residues is released from each of the two Aα chains, as is a B peptide of 20 residues from each of the two Bβ chains. These A and B peptides are called *fibrinopeptides.* A fibrinogen molecule devoid of these fibrinopeptides is called a *fibrin monomer* and has the subunit structure $(\alpha\beta\gamma)_2$.

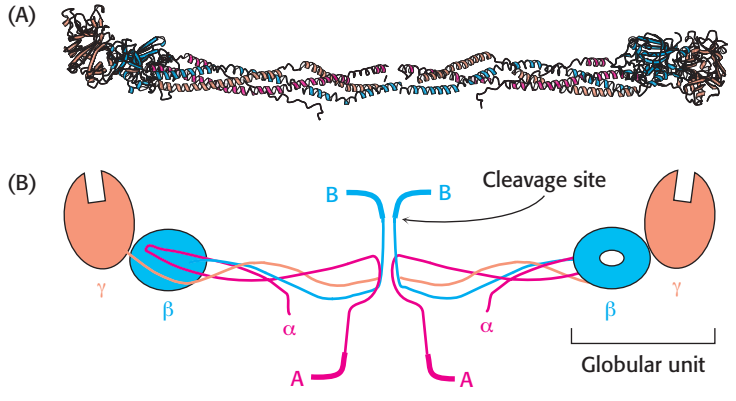


Figure 10.27 Structure of a fibrinogen molecule. (A) A ribbon diagram. The two rod regions are α-helical coiled coils, connected to a globular region at each end. The structure of the central globular region has not been determined. (B) A schematic representation showing the positions of the fibrinopeptides A and B. [Part A drawn from 1DEQ.pdb.]

Fibrin monomers spontaneously assemble into ordered fibrous arrays called *fibrin*. Electron micrographs and low-angle x-ray patterns show that fibrin has a periodic structure that repeats every 23 nm (Figure 10.28). Higher-resolution images reveal how the removal of the fibrinopeptides permits the fibrin monomers to come together to form fibrin. The homologous β and γ chains have globular domains at the carboxyl-terminal ends (Figure 10.29). These domains have binding "holes" that interact with peptides. The β domain is specific for sequences of the form H_3N^+-Gly-His-Arg-, whereas the γ domain binds H_3N^+-Gly-Pro-Arg-. Exactly these sequences (sometimes called "knobs") are exposed at the amino-terminal ends of the β and α chains, respectively, on thrombin cleavage. The knobs of the α subunits fit into the holes on the γ subunits of another monomer to form a protofibril. This protofibril is extended when the knobs of the β subunits fit into the holes of β subunits of other protofibrils. Thus, analogous to the activation of chymotrypsinogen, peptide-bond cleavage exposes new amino termini that can participate in specific interactions. The newly formed "soft clot" is stabilized by the formation of amide bonds between the side chains of lysine and glutamine residues in different monomers.

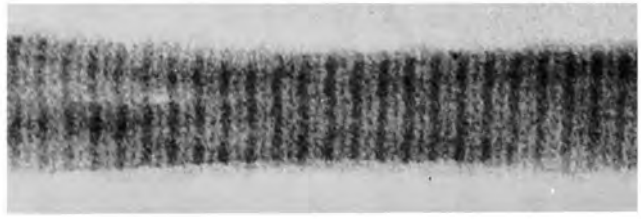
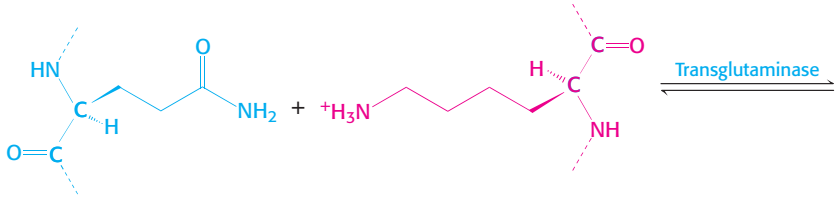
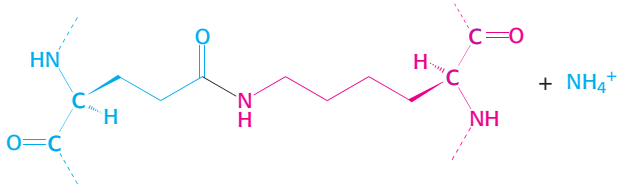


Figure 10.28 Electron micrograph of fibrin. The 23-nm period along the fiber axis is half the length of a fibrinogen molecule. [Courtesy of Dr. Henry Slayter.]



Glutamine **Lysine**



Cross-link

This cross-linking reaction is catalyzed by *transglutaminase (factor $XIII_a$)*, which itself is activated from the protransglutaminase form by thrombin.

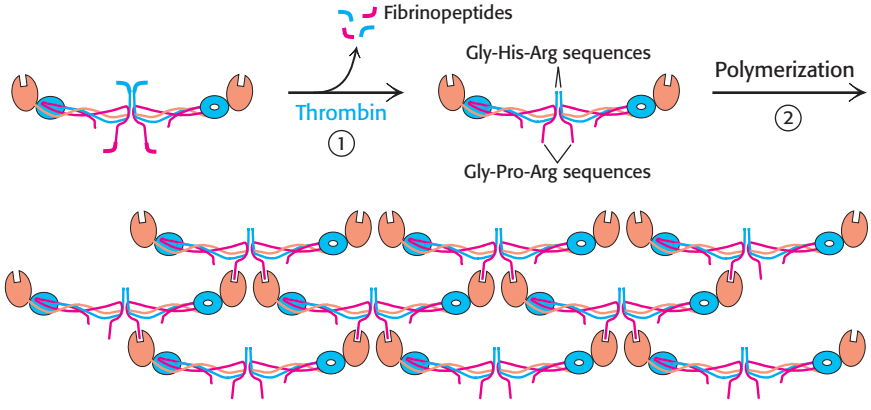


Figure 10.29 Formation of a fibrin clot. (1) Thrombin cleaves fibrinopeptides A and B from the central globule of fibrinogen. (2) Globular domains at the carboxyl-terminal ends of the β and γ chains interact with "knobs" exposed at the amino-terminal ends of the β and γ chains to form clots.

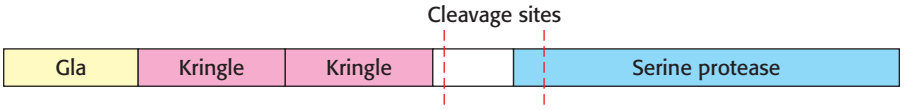
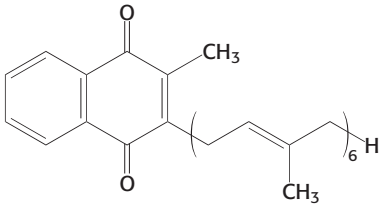
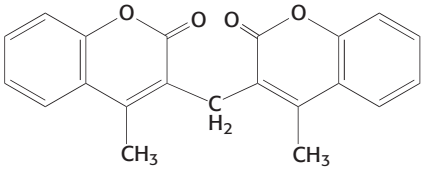


Figure 10.30 Modular structure of prothrombin. Cleavage of two peptide bonds yields thrombin. All the γ-carboxyglutamate residues are in the gla domain.

Prothrombin is readied for activation by a vitamin K-dependent modification

Thrombin is synthesized as a zymogen called *prothrombin*. The inactive molecule comprises four major domains, with the serine protease domain at its carboxyl terminus. The first domain is called a *gla domain* (a γ-carboxyglutamate-rich domain), and the second and third domains are called *kringle domains* (named after a Danish pastry that they resemble; Figure 10.30). These domains work in concert to keep prothrombin in an inactive form and to target it to appropriate sites for its activation by factor X_a (a serine protease) and factor V_a (a stimulatory protein). Activation is begun by proteolytic cleavage of the bond between arginine 274 and threonine 275 to release a fragment containing the first three domains. Cleavage of the bond between arginine 323 and isoleucine 324 (analogous to the key bond in chymotrypsinogen) yields active thrombin.





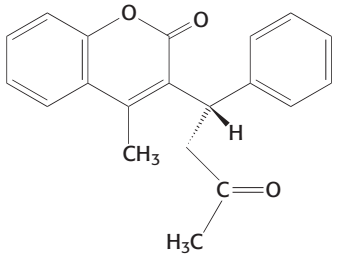
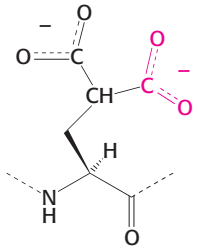


Figure 10.31 Structures of vitamin K and two antagonists, dicoumarol and warfarin.

Vitamin K (Figure 10.31) has been known for many years to be essential for the synthesis of prothrombin and several other clotting factors. Indeed, it is called vitamin K because a deficiency in this vitamin results in defective blood *k*oagulation (Scandinavian spelling). The results of studies of the abnormal prothrombin synthesized in the absence of vitamin K or in the presence of vitamin K antagonists, such as dicoumarol, revealed the vitamin's importance to proper clot formation. *Dicoumarol* is found in spoiled sweet clover and causes a fatal hemorrhagic disease in cattle fed on this hay. This coumarin derivative is used clinically as an *anticoagulant* to prevent thromboses in patients prone to clot formation. Dicoumarol and such related vitamin K antagonists as *warfarin* also serve as effective rat poisons. Cows fed dicoumarol synthesize an abnormal prothrombin that does not bind Ca^{2+}, in contrast with normal prothrombin. This difference was puzzling for some time because abnormal prothrombin has the same number of amino acid residues as that of normal prothrombin and gives the same amino acid analysis after acid hydrolysis.

Nuclear magnetic resonance studies revealed that normal prothrombin contains *γ-carboxyglutamate,* a formerly unknown residue that evaded detection because its second carboxyl group is lost on acid hydrolysis in the course of amino acid analysis. The abnormal prothrombin formed subsequent to the administration of anticoagulants lacks this modified amino acid. In fact, the first 10 glutamate residues in the amino-terminal region of prothrombin are carboxylated to γ-carboxyglutamate by a vitamin K-dependent enzyme system (Figure 10.32). *The vitamin K-dependent carboxylation reaction converts glutamate, a weak chelator of Ca^{2+}, into γ-carboxyglutamate, a much stronger chelator.* Prothrombin is thus able to bind Ca^{2+}, but what is the effect of this binding? The binding of Ca^{2+} by prothrombin anchors the zymogen to phospholipid membranes derived from blood platelets after injury. The binding of prothrombin to phospholipid surfaces is crucial because it brings prothrombin into close proximity to two clotting proteins that catalyze its conversion into thrombin. The



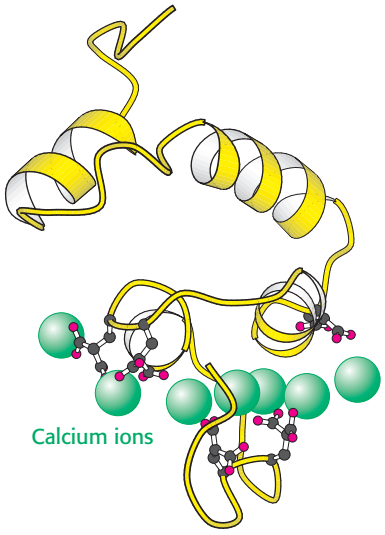


Figure 10.32 The calcium-binding region of prothrombin. Prothrombin binds calcium ions with the modified amino acid γ-carboxyglutamate (red). [Drawn from 2PF2.pdb.]

calcium-binding domain is removed during activation, freeing the thrombin from the membrane so that it can cleave fibrinogen and other targets.

Hemophilia revealed an early step in clotting

Some important breakthroughs in the elucidation of clotting pathways have come from studies of patients with bleeding disorders. *Classic hemophilia*, or *hemophilia A*, is the best-known clotting defect. This disorder is genetically transmitted as a sex-linked recessive characteristic. *In classic hemophilia, factor VIII (antihemophilic factor) of the intrinsic pathway is missing or has markedly reduced activity.* Although factor VIII is not itself a protease, it markedly stimulates the activation of factor X, the final protease of the intrinsic pathway, by factor IX_a, a serine protease (Figure 10.33). Thus, activation of the intrinsic pathway is severely impaired in hemophilia.

In the past, hemophiliacs were treated with transfusions of a concentrated plasma fraction containing factor VIII. This therapy carried the risk of infection. Indeed, many hemophiliacs contracted hepatitis and, more recently, AIDS. A safer source of factor VIII was urgently needed. With the use of biochemical purification and recombinant DNA techniques, the gene for factor VIII was isolated and expressed in cells grown in culture. Recombinant factor VIII purified from these cells has largely replaced plasma concentrates in treating hemophilia.

An account of a hemorrhagic disposition existing in certain families

"About seventy or eighty years ago, a woman by the name of Smith settled in the vicinity of Plymouth, New Hampshire, and transmitted the following idiosyncrasy to her descendants. It is one, she observed, to which her family is unfortunately subject and has been the source not only of great solicitude, but frequently the cause of death. If the least scratch is made on the skin of some of them, as mortal a hemorrhage will eventually ensue as if the largest wound is inflicted. . . . It is a surprising circumstance that the males only are subject to this strange affection, and that all of them are not liable to it. . . . Although the females are exempt, they are still capable of transmitting it to their male children."

John Otto (1803)

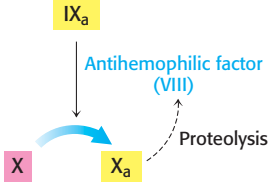


Figure 10.33 Action of antihemophilic factor. Antihemophilic factor (Factor VIII) stimulates the activation of factor X by factor IX_a. Interestingly, the activity of factor VIII is markedly increased by limited proteolysis by thrombin. This positive feedback amplifies the clotting signal and accelerates clot formation after a threshold has been reached.

The clotting process must be precisely regulated

There is a fine line between hemorrhage and thrombosis, the formation of blood clots in blood vessels. Clots must form rapidly yet remain confined to the area of injury. What are the mechanisms that normally limit clot

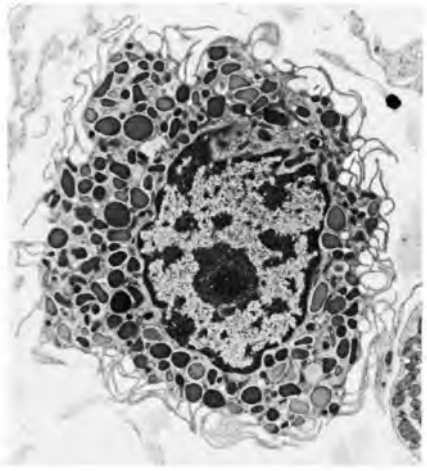


Figure 10.34 Electron micrograph of a mast cell. Heparin and other molecules in the dense granules are released into the extracellular space when the cell is triggered to secrete. [Courtesy of Lynne Mercer.]

formation to the site of injury? The lability of clotting factors contributes significantly to the control of clotting. Activated factors are short-lived because they are diluted by blood flow, removed by the liver, and degraded by proteases. For example, the stimulatory protein factors V_a and $VIII_a$ are digested by protein C, a protease that is switched on by the action of thrombin. *Thus, thrombin has a dual function: it catalyzes the formation of fibrin and it initiates the deactivation of the clotting cascade.*

Specific inhibitors of clotting factors are also critical in the termination of clotting. For instance, *tissue factor pathway inhibitor* (TFPI) inhibits the complex of TF–VII_a–X_a. Separate domains in TFPI inhibit VII_a and X_a. Another key inhibitor is *antithrombin III,* a plasma protein that inactivates thrombin by forming an irreversible complex with it. Antithrombin III resembles α_1-antitrypsin except that it inhibits thrombin much more strongly than it inhibits elastase (see Figure 10.24). Antithrombin III also blocks other serine proteases in the clotting cascade—namely, factors XII_a, XI_a, IX_a, and X_a. The inhibitory action of antithrombin III is enhanced by *heparin,* a negatively charged polysaccharide found in mast cells near the walls of blood vessels and on the surfaces of endothelial cells (Figure 10.34). Heparin acts as an *anticoagulant* by increasing the rate of formation of irreversible complexes between antithrombin III and the serine protease clotting factors. Antitrypsin and antithrombin are *serpins,* a family of *ser*ine *p*rotease *in*hibitors.

The importance of the ratio of thrombin to antithrombin is illustrated in the case of a 14-year-old boy who died of a bleeding disorder because of a mutation in his α_1-antitrypsin, which normally inhibits elastase. Methionine 358 in α_1-antitrypsin's binding pocket for elastase was replaced by arginine, resulting in a change in specificity from an elastase inhibitor to a thrombin inhibitor. α_1-Antitrypsin activity normally increases markedly after injury to counteract excess elastase arising from stimulated neutrophils. The mutant α_1-antitrypsin caused the patient's thrombin activity to drop to such a low level that hemorrhage ensued. We see here a striking example of how a change of a single residue in a protein can dramatically alter specificity and an example of the critical importance of having the right amount of a protease inhibitor.

Antithrombin limits the extent of clot formation, but what happens to the clots themselves? Clots are not permanent structures but are designed to dissolve when the structural integrity of damaged areas is restored. Fibrin is split by *plasmin,* a serine protease that hydrolyzes peptide bonds in the coiled-coil regions. Plasmin molecules can diffuse through aqueous channels in the porous fibrin clot to cut the accessible connector rods. Plasmin is formed by the proteolytic activation of *plasminogen,* an inactive precursor that has a high affinity for the fibrin clots. This conversion is carried out by *tissue-type plasminogen activator* (TPA), a 72-kd protein that has a domain structure closely related to that of prothrombin (Figure 10.35). However, a domain that targets TPA to fibrin clots replaces the membrane-targeting gla domain of prothrombin. The TPA bound to fibrin clots swiftly activates adhering plasminogen. In contrast, TPA activates free plasminogen very slowly. The gene for TPA has been cloned and expressed in cultured mammalian cells. Clinical studies have shown that TPA administered intravenously within an hour of the formation of a

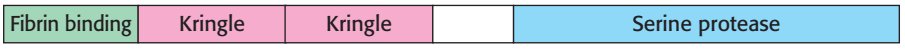


Figure 10.35 Modular structure of tissue-type plasminogen activator (TPA).

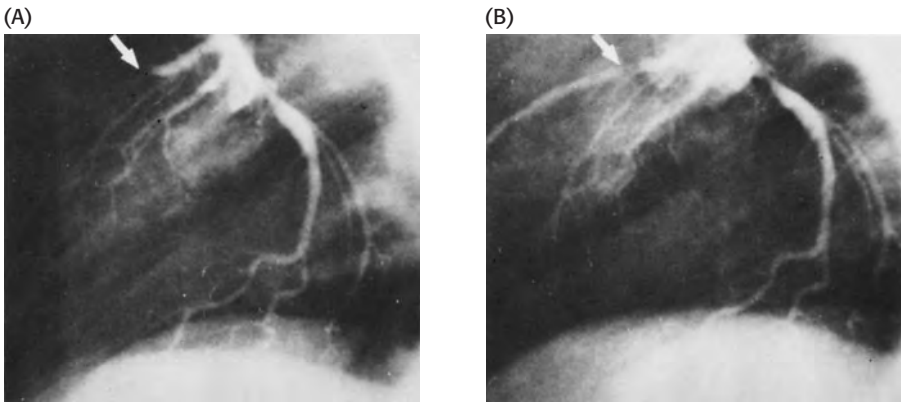


Figure 10.36 The effect of tissue-type plasminogen factor. TPA leads to the dissolution of blood clots, as shown by x-ray images of blood vessels in the heart (A) before and (B) 3 hours after the administration of TPA. The position of the clot is marked by the arrow in part A. [After F. Van de Werf, P. A. Ludbrook, S. R. Bergmann, A. J. Tiefenbrunn, K. A. A. Fox, H. de Geest, M. Verstraete, D. Collen, and B. E. Sobel. *New Engl. J. Med.* 310(1984):609–613.]

blood clot in a coronary artery markedly increases the likelihood of surviving a heart attack (Figure 10.36).

Summary

10.1 Aspartate Transcarbamoylase Is Allosterically Inhibited by the End Product of Its Pathway

Allosteric proteins constitute an important class of proteins whose biological activity can be regulated. Specific regulatory molecules can modulate the activity of allosteric proteins by binding to distinct regulatory sites, separate from the functional sites. These proteins have multiple functional sites, which display cooperation as evidenced by a sigmoidal dependence of function on substrate concentration. Aspartate transcarbamoylase (ATCase), one of the best-understood allosteric enzymes, catalyzes the synthesis of *N*-carbamoylaspartate, the first intermediate in the synthesis of pyrimidines. ATCase is feedback inhibited by cytidine triphosphate, the final product of the pathway. ATP reverses this inhibition. ATCase consists of separable catalytic (c_3) subunits (which bind the substrates) and regulatory (r_2) subunits (which bind CTP and ATP). The inhibitory effect of CTP, the stimulatory action of ATP, and the cooperative binding of substrates are mediated by large changes in quaternary structure. On binding substrates, the c_3 subunits of the c_6r_6 enzyme move apart and reorient themselves. This allosteric transition is highly concerted. All subunits of an ATCase molecule simultaneously interconvert from the T (low-affinity) to the R (high-affinity) state.

10.2 Isozymes Provide a Means of Regulation Specific to Distinct Tissues and Developmental Stages

Isozymes differ in structural characteristics but catalyze the same reaction. They provide a means of fine-tuning metabolism to meet the needs of a given tissue or developmental stage. The results of gene-duplication events provide the means for subtle regulation of enzyme function.

10.3 Covalent Modification Is a Means of Regulating Enzyme Activity

The covalent modification of proteins is a potent means of controlling the activity of enzymes and other proteins. Phosphorylation is the

most common type of reversible covalent modification. Signals can be highly amplified by phosphorylation because a single kinase can act on many target molecules. The regulatory actions of protein kinases are reversed by protein phosphatases, which catalyze the hydrolysis of attached phosphoryl groups.

Cyclic AMP serves as an intracellular messenger in the transduction of many hormonal and sensory stimuli. Cyclic AMP switches on protein kinase A, a major multifunctional kinase, by binding to the regulatory subunit of the enzyme, thereby releasing the active catalytic subunits of PKA. In the absence of cAMP, the catalytic sites of PKA are occupied by pseudosubstrate sequences of the regulatory subunit.

10.4 Many Enzymes Are Activated by Specific Proteolytic Cleavage

The activation of an enzyme by the proteolytic cleavage of one or a few peptide bonds is a recurring control mechanism seen in processes as diverse as the activation of digestive enzymes and blood clotting. The inactive precursor is a zymogen (proenzyme). Trypsinogen is activated by enteropeptidase or trypsin, and trypsin then activates a host of other zymogens, leading to the digestion of foodstuffs. For instance, trypsin converts chymotrypsinogen, a zymogen, into active chymotrypsin by hydrolyzing a single peptide bond.

A striking feature of the clotting process is that it is accomplished by a cascade of zymogen conversions, in which the activated form of one clotting factor catalyzes the activation of the next precursor. Many of the activated clotting factors are serine proteases. In the final step of clot formation, fibrinogen, a highly soluble molecule in the plasma, is converted by thrombin into fibrin by the hydrolysis of four arginine–glycine bonds. The resulting fibrin monomer spontaneously forms long, insoluble fibers called fibrin. Zymogen activation is also essential in the lysis of clots. Plasminogen is converted into plasmin, a serine protease that cleaves fibrin, by tissue-type plasminogen activator. Although zymogen activation is irreversible, specific inhibitors of some proteases exert control. The irreversible protein inhibitor antithrombin III holds blood clotting in check in the clotting cascade.

Key Terms

cooperativity (p. 289)
feedback (end-product) inhibition (p. 291)
allosteric (regulatory) site (p. 291)
homotropic effect (p. 294)
concerted mechanism (p. 294)
sequential model (p. 294)
heterotropic effect (p. 296)
isozyme (isoenzyme) (p. 296)
covalent modification (p. 297)
protein kinase (p. 298)
consensus sequence (p. 299)
protein phosphatase (p. 300)
protein kinase A (PKA) (p. 301)
pseudosubstrate sequence (p. 301)
zymogen (proenzyme) (p. 303)
enzymatic cascade (p. 307)
intrinsic pathway (p. 308)
extrinsic pathway (p. 308)

Problems

1. *Context please.* The allosteric properties of aspartate transcarbamoylase have been discussed in detail in this chapter. What is the function of aspartate transcarbamoylase?

2. *Activity profile.* A histidine residue in the active site of aspartate transcarbamoylase is thought to be important in stabilizing the transition state of the bound substrates.

Predict the pH dependence of the catalytic rate, assuming that this interaction is essential and dominates the pH-activity profile of the enzyme. (See equations on p. 16.)

3. *Knowing when to say when.* What is feedback inhibition? Why is it a useful property?

4. *Knowing when to get going.* What is the biochemical rationale for ATP serving as a positive regulator of ATCase?

5. *No T.* What would be the effect of a mutation in an allosteric enzyme that resulted in a T/R ratio of 0?

6. *Turned upside down.* An allosteric enzyme that follows the concerted mechanism has a T/R ratio of 300 in the absence of substrate. Suppose that a mutation reversed the ratio. How would this mutation affect the relation between the rate of the reaction and the substrate concentration?

7. *Partners.* As shown in Figure 10.2, CTP inhibits ATCase; however, the inhibition is not complete. Can you suggest another molecule that might enhance the inhibition of ATCase? Hint: See Figure 25.2.

8. *RT equilibrium.* Differentiate between homotropic and heterotropic effectors.

9. *Restoration project.* If isolated regulatory subunits and catalytic subunits of ATCase are mixed, the native enzyme is reconstituted. What is the biological significance of the observation?

10. *Because it's an enzyme.* X-ray crystallographic studies of ATCase in the R form required the use of the bisubstrate analog PALA. Why was this analog, a competitive inhibitor, used instead of the actual substrates?

11. *Allosteric switching.* A substrate binds 100 times as tightly to the R state of an allosteric enzyme as to its T state. Assume that the concerted (MWC) model applies to this enzyme. (See equations for the Concerted Model in the Appendix to Chapter 7.)

(a) By what factor does the binding of one substrate molecule per enzyme molecule alter the ratio of the concentrations of enzyme molecules in the R and T states?

(b) Suppose that L, the ratio of [T] to [R] in the absence of substrate, is 10^7 and that the enzyme contains four binding sites for substrate. What is the ratio of enzyme molecules in the R state to those in the T state in the presence of saturating amounts of substrate, assuming that the concerted model is obeyed?

12. *Allosteric transition.* Consider an allosteric protein that obeys the concerted model. Suppose that the ratio of T to R formed in the absence of ligand is 10^5, $K_T = 2$ mM, and $K_R = 5$ μM. The protein contains four binding sites for ligand. What is the fraction of molecules in the R form when 0, 1, 2, 3, and 4 ligands are bound? (See equations for the Concerted Model in the Appendix to Chapter 7.)

13. *Negative cooperativity.* You have isolated a dimeric enzyme that contains two identical active sites. The binding of substrate to one active site decreases the substrate affinity of the other active site. Can the concerted model account for this negative cooperativity?

14. *Paradoxical at first glance.* Recall that phosphonacetyl-L-aspartate (PALA) is a potent inhibitor of ATCase because it mimics the two physiological substrates. However, low concentrations of this unreactive bisubstrate analog *increase* the reaction velocity. On the addition of PALA, the reaction rate increases until an average of three molecules of PALA are bound per molecule of enzyme. This maximal velocity is 17-fold greater than it is in the absence of PALA. The reaction rate then decreases to nearly zero on the addition of three more molecules of PALA per molecule of enzyme. Why do low concentrations of PALA activate ATCase?

15. *Regulation energetics.* The phosphorylation and dephosphorylation of proteins is a vital means of regulation. Protein kinases attach phosphoryl groups, whereas only a phosphatase will remove the phosphoryl group from the target protein. What is the energy cost of this means of covalent regulation?

16. *Viva la difference.* What is an isozyme?

17. *Fine-tuning biochemistry.* What is the advantage for an organism to have isozymic forms of an enzyme?

18. *Making matches.*

(a) ATCase_____
(b) T state_____
(c) R state_____
(d) Phosphorylation_____
(e) Kinase_____
(f) Phosphatase_____
(g) cAMP_____
(h) Zymogen_____
(i) Enteropeptidase_____
(j) Vitamin K_____
(l) Tissue factor_____

1. Protein phosphorylation
2. Required to modify glutamate
3. Activates a particular kinase
4. Proenzyme
5. Activates trypsin
6. Common covalent modification
7. Inhibited by CTP
8. Less-active state of an allosteric protein
9. Initiates extrinsic pathway
10. Forms fibrin
11. More-active state of an allosteric protein
12. Removes phosphates

19. *Powering change.* Phosphorylation is a common covalent modification of proteins in all forms of life. What energetic advantages accrue from the use of ATP as the phosphoryl donor?

20. *No going back.* What is the key difference between regulation by covalent modification and specific proteolytic cleavage?

21. *Zymogen activation.* When very low concentrations of pepsinogen are added to acidic media, how does the half-time for activation depend on zymogen concentration?

22. *No protein shakes advised.* Predict the physiological effects of a mutation that resulted in a deficiency of enteropeptidase.

23. *A revealing assay.* Suppose that you have just examined a young boy with a bleeding disorder highly suggestive of classic hemophilia (factor VIII deficiency). Because of the late hour, the laboratory that carries out specialized coagulation assays is closed. However, you happen to have a sample of blood from a classic hemophiliac whom you admitted to the hospital an hour earlier. What is the simplest and most rapid test that you can perform to determine whether your present patient also is deficient in factor VIII activity?

24. *Counterpoint.* The synthesis of factor X, like that of prothrombin, requires vitamin K. Factor X also contains γ-carboxyglutamate residues in its amino-terminal region. However, activated factor X, in contrast with thrombin, retains this region of the molecule. What is a likely functional consequence of this difference between the two activated species?

25. *A discerning inhibitor.* Antithrombin III forms an irreversible complex with thrombin but not with prothrombin. What is the most likely reason for this difference in reactivity?

26. *Repeating heptads.* Each of the three types of fibrin chains contains repeating heptapeptide units *(abcdefg)* in which residues *a* and *d* are hydrophobic. Propose a reason for this regularity.

27. *Drug design.* A drug company has decided to use recombinant DNA methods to prepare a modified α_1-antitrypsin that will be more resistant to oxidation than is the naturally occurring inhibitor. Which single amino acid substitution would you recommend?

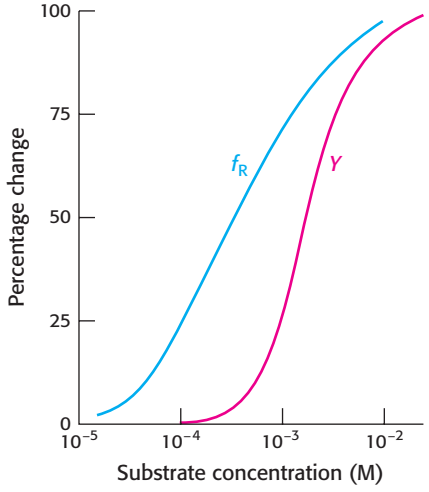
28. *Blood must flow.* Why is inappropriate blood-clot formation dangerous?

29. *Dissolution now.* What is tissue-type plasminogen activator and what is its role in preventing heart attacks.

30. *Joining together.* What differentiates a soft clot from a mature clot?

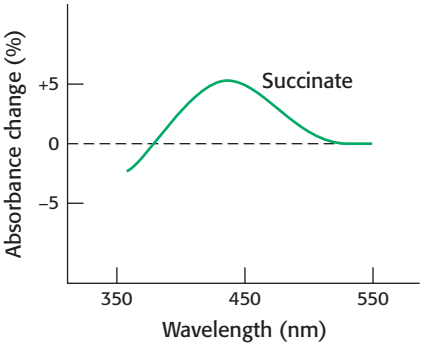
Data Interpretation Problems

31. *Distinguishing between models.* The following graph shows the fraction of an allosteric enzyme in the R state (f_R) and the fraction of active sites bound to substrate (Y) as a function of substrate concentration. Which model, the concerted or sequential, best explains these results?



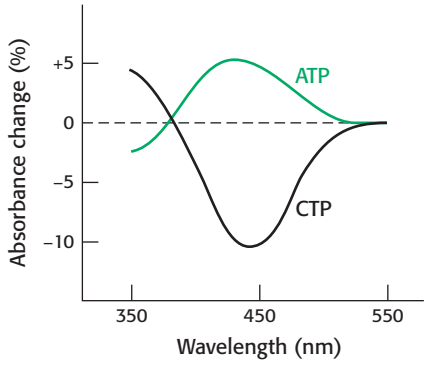
[After M. W. Kirschner and H. K. Schachman. *Biochemistry* 12:2997–3004, 1966.]

32. *Reporting live from ATCase 1.* ATCase underwent reaction with tetranitromethane to form a colored nitrotyrosine group (λ_{max} = 430 nm) in each of its catalytic chains. The absorption by this reporter group depends on its immediate environment. An essential lysine residue at each catalytic site also was modified to block the binding of substrate. Catalytic trimers from this doubly modified enzyme were then combined with native trimers to form a hybrid enzyme. The absorption by the nitrotyrosine group was measured on addition of the substrate analog succinate. What is the significance of the alteration in the absorbance at 430 nm?



[After H. K. Schachman. *J. Biol. Chem.* 263:18583–18586, 1988.]

33. *Reporting live from ATCase 2.* A different ATCase hybrid was constructed to test the effects of allosteric activators and inhibitors. Normal regulatory subunits were combined with nitrotyrosine-containing catalytic subunits. The addition of ATP in the absence of substrate increased the absorbance at 430 nm, the same change elicited by the addition of succinate (see the graph in Problem 32). Conversely, CTP in the absence of substrate decreased the absorbance at 430 nm. What is the significance of the changes in absorption of the reporter groups?



[After H. K. Schachman. *J. Biol. Chem.* 263:18583–18586, 1988.]

Chapter Integration Problems

34. *Density matters.* The sedimentation value of aspartate transcarbamoylase decreases when the enzyme switches to the R state. On the basis of the allosteric properties of the enzyme, explain why the sedimentation value decreases.

35. *Too tight a grip.* Trypsin cleaves proteins on the carboxyl side of lysine. Trypsin inhibitor has a lysine residue, and binds to trypsin, yet it is not a substrate. Explain.

Mechanism Problems

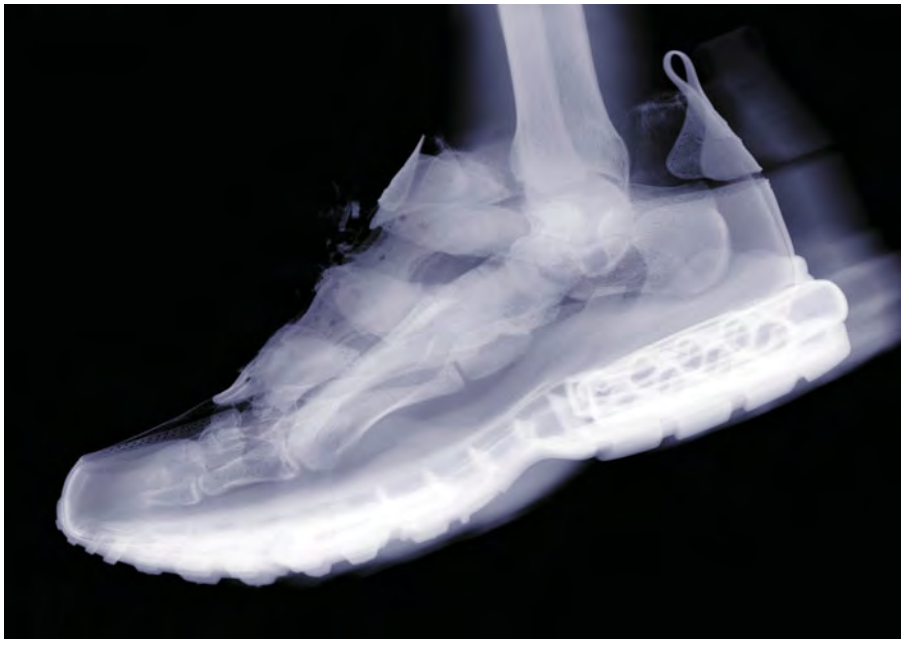
36. *Aspartate transcarbamoylase.* Write the mechanism (in detail) for the conversion of aspartate and carbamoyl phosphate into *N*-carbamoylaspartate. Include a role for the histidine residue present in the active site.

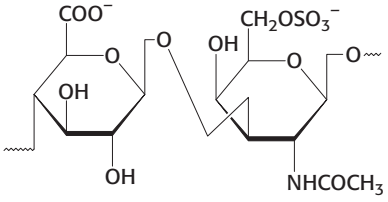
37. *Protein kinases.* Write a mechanism (in detail) for the phosphorylation of a serine residue by ATP catalyzed by a protein kinase. What groups might you expect to find in the enzyme's active site?

This page intentionally left blank

CHAPTER 11

Carbohydrates





Carbohydrates are important fuel molecules, but they play many other biochemical roles, including protection against high-impact forces. The cartilage of a runner's foot cushions the impact of each step she takes. A key component of cartilage are molecules called glycosaminoglycans, large polymers made up of many repeats of dimers such as the pair shown at the right. [Untitled x-ray/ Nick Veasey/Getty Images.]

For years, the study of carbohydrates was considered less exciting than many if not most topics of biochemistry. Carbohydrates were recognized as important fuels and structural components but were thought to be peripheral to most key activities of the cell. In essence, they were considered the underlying girders and fuel for a magnificent piece of biochemical architecture. This view has changed dramatically in the past few years. We have learned that cells of all organisms are coated in a dense and complex coat of carbohydrates. Secreted proteins are often extensively decorated with carbohydrates essential to a protein's function. The extracellular matrix in higher eukaryotes—the environment in which the cells live—is rich in secreted carbohydrates central to cell survival and cell-to-cell communication. Carbohydrates are crucial for the development and functioning of all organisms, not only as fuels, but also as information-rich molecules. Carbohydrates, carbohydrate-containing proteins, and specific carbohydrate-binding proteins are required for interactions that allow cells to form tissues, are the basis of human blood groups, and are used by a variety of pathogens

OUTLINE

to gain access to their hosts. Indeed, rather than mere infrastructure components, carbohydrates supply details and enhancements to the biochemical architecture of the cell, helping to define the beauty, functionality, and uniqueness of the cell.

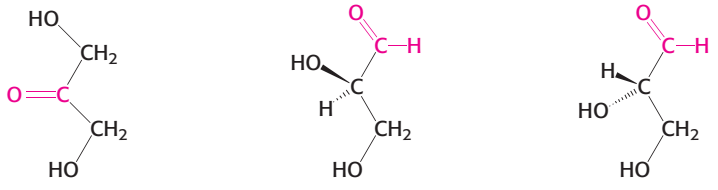
A key property of carbohydrates that allows their many functions is the tremendous structural *diversity* possible within this class of molecules. Carbohydrates are built from monosaccharides, which are small molecules—typically containing from three to nine carbon atoms that are bound to hydroxyl groups—that vary in size and in the stereochemical configuration at one or more carbon centers. These monosaccharides can be linked together to form a large variety of oligosaccharide structures. The sheer number of possible oligosaccharides makes this class of molecules information rich. This information, when attached to proteins, can augment the already immense diversity of proteins.

The realization of the importance of carbohydrates to so many aspects of biochemistry has spawned a field of study called *glycobiology*. Glycobiology is the study of the synthesis and structure of carbohydrates and how carbohydrates are attached to and recognized by other molecules such as proteins. Along with a new field comes a new "omics" to join genomics and proteomics—*glycomics*. Glycomics is the study of the glycome, all of the carbohydrates and carbohydrate-associated molecules that cells produce. Like the proteome, the glycome is not static and can change, depending on cellular and environmental conditions. Unraveling oligosaccharide structures and elucidating the effects of their attachment to other molecules constitute a tremendous challenge in the field of biochemistry.

11.1 Monosaccharides Are the Simplest Carbohydrates

Carbohydrates are carbon-based molecules that are rich in hydroxyl groups. Indeed, the empirical formula for many carbohydrates is $(CH_2O)_n$—literally, a carbon hydrate. Simple carbohydrates are called *monosaccharides*. These simple sugars serve not only as fuel molecules but also as fundamental constituents of living systems. For instance, DNA is built on simple sugars: its backbone consists of alternating phosphoryl groups and deoxyribose, a cyclic five-carbon sugar.

Monosaccharides are aldehydes or ketones that have two or more hydroxyl groups. The smallest monosaccharides, composed of three carbon atoms, are dihydroxyacetone and D- and L-glyceraldehyde.



Dihydroxyacetone (a ketose) | **D-Glyceraldehyde** (an aldose) | **L-Glyceraldehyde** (an aldose)

Dihydroxyacetone is called a *ketose* because it contains a keto group (in red above), whereas glyceraldehyde is called an *aldose* because it contains an aldehyde group. They are referred to as *trioses* (tri- for three, referring to the three carbon atoms that they contain). Similarly, simple monosaccharides with four, five, six, and seven carbon atoms are called *tetroses, pentoses, hexoses,* and *heptoses,* respectively. Perhaps the monosaccharides of which we are most aware are the hexoses, such as glucose and fructose. Glucose is

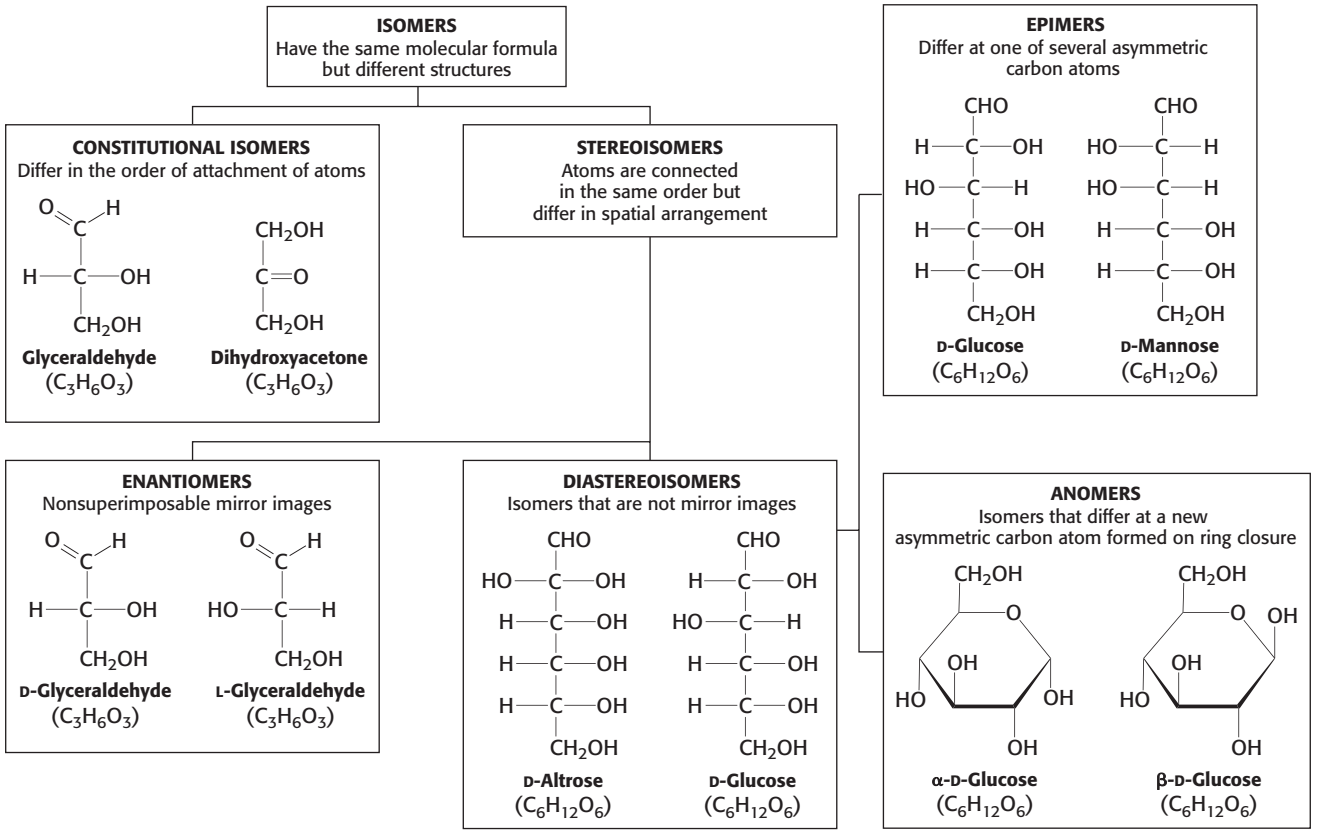


Figure 11.1 Isomeric forms of carbohydrates.

an essential energy source for virtually all forms of life. Fructose is commonly used as a sweetener that is converted into glucose derivatives inside the cell.

Carbohydrates can exist in a dazzling variety of isomeric forms (Figure 11.1). Dihydroxyacetone and glyceraldehyde are called *constitutional isomers* because they have identical molecular formulas but differ in how the atoms are ordered. *Stereoisomers* are isomers that differ in spatial arrangement. Recall from the discussion of amino acids (p. 27) that stereoisomers are designated as having either D or L configuration. Glyceraldehyde has a single asymmetric carbon atom and, thus, there are two stereoisomers of this sugar: D-glyceraldehyde and L-glyceraldehyde. These molecules are a type of stereoisomer called *enantiomers,* which are mirror images of each other. Most vertebrate monosaccharides have the D configuration. According to convention, the D and L isomers are determined by the configuration of the asymmetric carbon atom farthest from the aldehyde or keto group. Dihydroxyacetone is the only monosaccharide without at least one asymmetric carbon atom.

Monosaccharides made up of more than three carbon atoms have multiple asymmetric carbons, and so they can exist not only as enantiomers but also as *diastereoisomers,* isomers that are not mirror images of each other. The number of possible stereoisomers equals 2^n, where n is the number of asymmetric carbon atoms. Thus, a six-carbon aldose with 4 asymmetric carbon atoms can exist in 16 possible diastereoisomers, of which glucose is one such isomer.

Figure 11.2 shows the common sugars that we will see most frequently in our study of biochemistry. D-Ribose, the carbohydrate component of RNA, is a five-carbon aldose, as is deoxyribose, the monosaccharide component of deoxynucleotides. D-Glucose, D-mannose, and D-galactose are

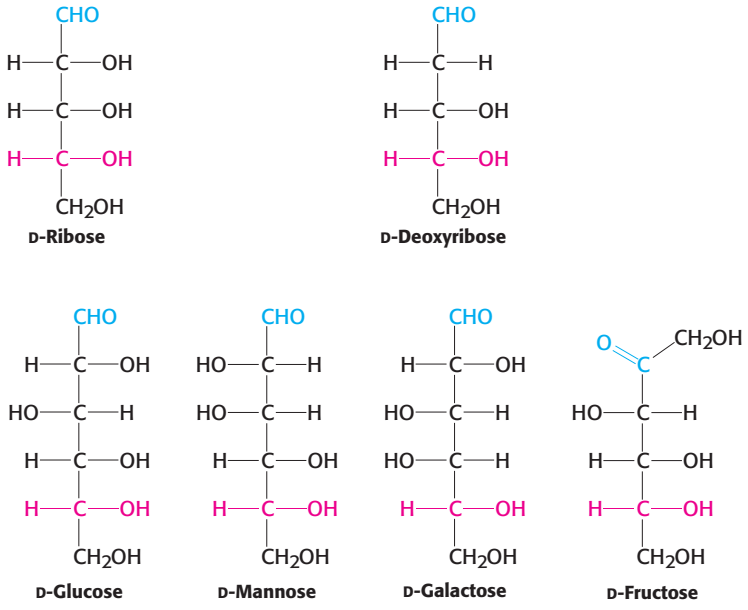


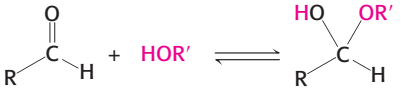
Figure 11.2 Common monosaccharides. Aldoses contain an aldehyde (shown in blue), whereas ketoses, such as fructose, contain a ketose (also shown in blue). The asymmetric carbon atom farthest from the aldehyde or ketone (shown in red) designates the structures as being in the D configuration.

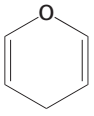
abundant six-carbon aldoses. Note that D-glucose and D-mannose differ in configuration only at C-2, the carbon atom in the second position. Sugars that are diastereoisomers differing in configuration at only a single asymmetric center are called *epimers*. Thus, D-glucose and D-mannose are epimeric at C-2; D-glucose and D-galactose are epimeric at C-4.

Note that ketoses have one less asymmetric center than aldoses with the same number of carbon atoms. D-Fructose is the most abundant ketohexose.

Many common sugars exist in cyclic forms

The predominant forms of ribose, glucose, fructose, and many other sugars in solution, as is the case inside the cell, are not open chains. Rather, the open-chain forms of these sugars cyclize into rings. The chemical basis for ring formation is that an aldehyde can react with an alcohol to form a *hemiacetal*.

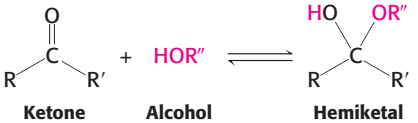


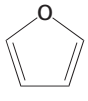


Pyran

For an aldohexose such as glucose, a single molecule provides both the aldehyde and the alcohol: the C-1 aldehyde in the open-chain form of glucose reacts with the C-5 hydroxyl group to form an *intramolecular hemiacetal* (Figure 11.3). The resulting cyclic hemiacetal, a six-membered ring, is called *pyranose* because of its similarity to *pyran*.

Similarly, a ketone can react with an alcohol to form a *hemiketal*.





Furan

The C-2 keto group in the open-chain form of a ketohexose, such as fructose, can form an *intramolecular hemiketal* by reacting with either the

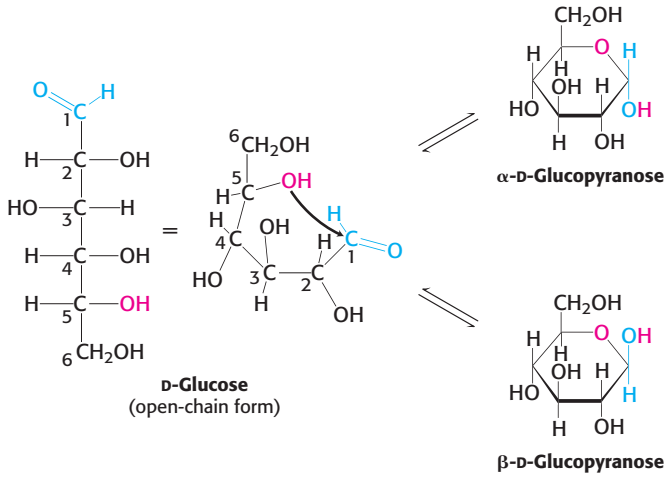


Figure 11.3 Pyranose formation. The open-chain form of glucose cyclizes when the C-5 hydroxyl group attacks the oxygen atom of the C-1 aldehyde group to form an intramolecular hemiacetal. Two anomeric forms, designated α and β, can result.

C-6 hydroxyl group to form a six-membered cyclic hemiketal or the C-5 hydroxyl group to form a five-membered cyclic hemiketal (Figure 11.4). The five-membered ring is called a *furanose* because of its similarity to *furan*.

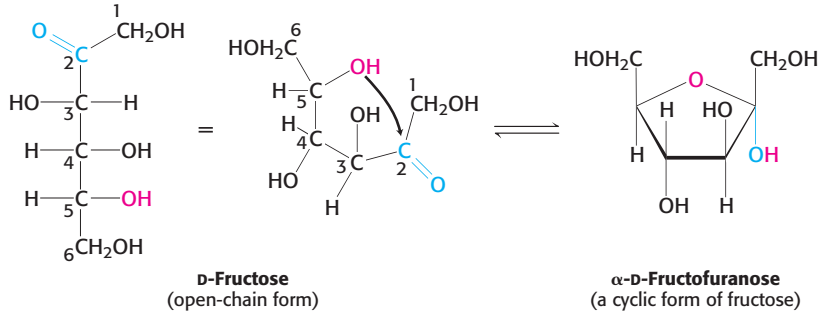


Figure 11.4 Furanose formation. The open-chain form of fructose cyclizes to a five-membered ring when the C-5 hydroxyl group attacks the C-2 ketone to form an intramolecular hemiketal. Two anomers are possible, but only the α anomer is shown.

The depictions of glucopyranose (glucose) and fructofuranose (fructose) shown in Figures 11.3 and 11.4 are *Haworth projections*. In such projections, the carbon atoms in the ring are not written out. The approximate plane of the ring is perpendicular to the plane of the paper, with the heavy line on the ring projecting toward the reader.

We have seen that carbohydrates can contain many asymmetric carbon atoms. An additional asymmetric center is created when a cyclic hemiacetal is formed, creating yet another diastereoisomeric form of sugars called anomers. In glucose, C-1 (the carbonyl carbon atom in the open-chain form) becomes an asymmetric center. Thus, two ring structures can be formed: α-D-glucopyranose and β-D-glucopyranose (see Figure 11.3). For D sugars drawn as Haworth projections in the standard orientation as shown in Figure 11.3, the *designation α means that the hydroxyl group attached to C-1 is on the opposite side of the ring as C-6; β means that the hydroxyl group is on the same side of the ring as C-6*. The C-1 carbon atom is called the *anomeric carbon atom*, and the α and β forms are called *anomers*. An equilibrium mixture of glucose contains approximately one-third α anomer, two-thirds β anomer, and <1% of the open-chain form.

The furanose-ring form of fructose also has anomeric forms, in which α and β refer to the hydroxyl groups attached to C-2, the anomeric carbon atom (see Figure 11.4). Fructose forms both pyranose and furanose rings.

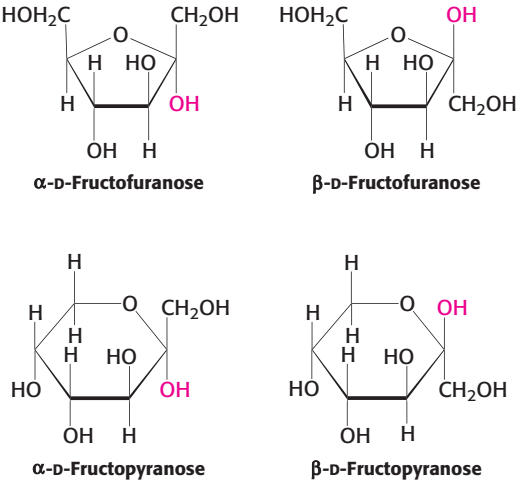


Figure 11.5 Ring structures of fructose. Fructose can form both five-membered furanose and six-membered pyranose rings. In each case, both α and β anomers are possible.

The pyranose form predominates in fructose free in solution, and the furanose form predominates in many fructose derivatives (Figure 11.5).

β-D-Fructopyranose, found in honey, is one of the sweetest chemicals known. The β-D-fructofuranose form is not nearly as sweet. Heating converts β-fructopyranose into the β-fructofuranose form, reducing the sweetness of the solution. For this reason, corn syrup with a high concentration of fructose in the β-D-pyranose form is used as a sweetener in cold, but not hot, drinks. Figure 11.6 shows the common sugars discussed previously in their ring forms.

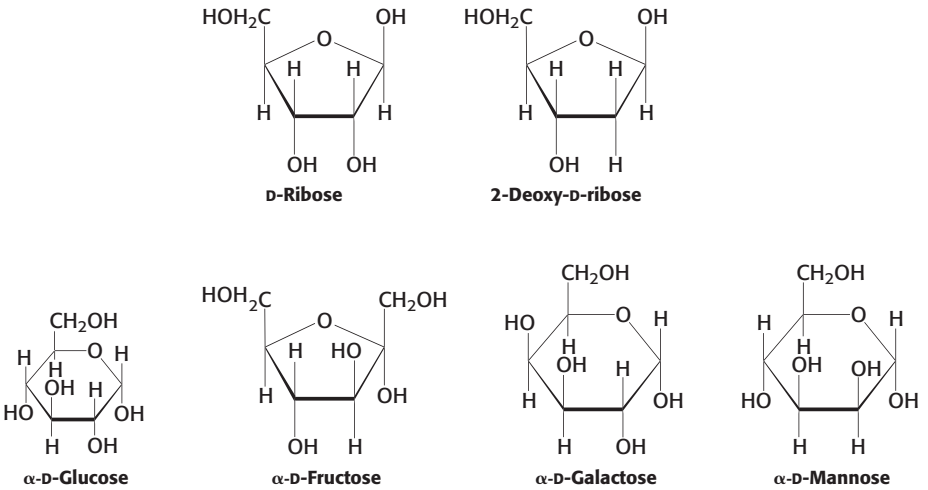
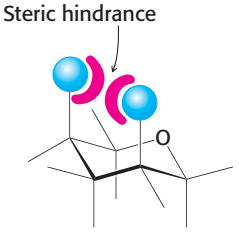


Figure 11.6 Common monosaccharides in their ring forms.

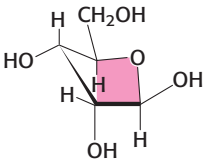
Pyranose and furanose rings can assume different conformations

The six-membered pyranose ring is not planar, because of the tetrahedral geometry of its saturated carbon atoms. Instead, pyranose rings adopt two classes of conformations, termed chair and boat because of the resemblance to these objects (Figure 11.7). In the chair form, the substituents on the ring carbon atoms have two orientations: axial and equatorial. *Axial* bonds are nearly perpendicular to the average plane of the ring, whereas *equatorial* bonds are nearly parallel to this plane. Axial substituents sterically hinder each other if they emerge on the same side of the ring (e.g., 1,3-diaxial groups). In contrast, equatorial substituents are less crowded.



The chair form of *β-D-glucopyranose predominates because all axial positions are occupied by hydrogen atoms.* The bulkier —OH and —CH_2OH groups emerge at the less-hindered periphery. The boat form of glucose is disfavored because it is quite sterically hindered.

Furanose rings, like pyranose rings, are not planar. They can be puckered so that four atoms are nearly coplanar and the fifth is about 0.5 Å away from this plane (Figure 11.8). This conformation is called an *envelope form* because the structure resembles an opened envelope with the back flap raised. In the ribose moiety of most biomolecules, either C-2 or C-3 is out of the plane on the same side as C-5. These conformations are called C-2-endo and C-3-endo, respectively.



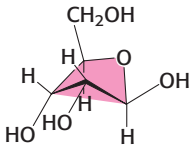


Figure 11.8 Envelope conformations of β-D-ribose. The C-3-endo and C-2-endo forms of β-D-ribose are shown. The color indicates the four atoms that lie approximately in a plane.

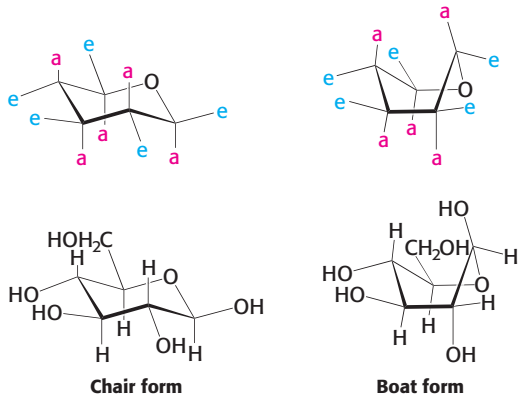
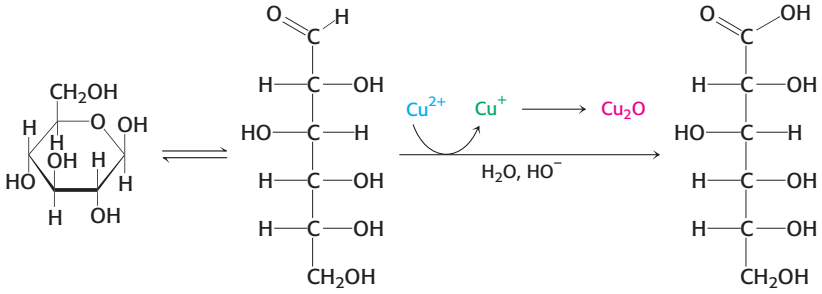


Figure 11.7 Chair and boat forms of β-D-glucopyranose. The chair form is more stable owing to less steric hindrance because the axial positions are occupied by hydrogen atoms. Abbreviations: a, axial; e, equatorial.

Glucose is a reducing sugar

Because the α and β isomers of glucose are in an equilibrium that passes through the open-chain form, glucose has some of the chemical properties of free aldehydes, such as the ability to react with oxidizing agents. For example, glucose can react with cupric ion (Cu^{2+}), reducing it to cuprous ion (Cu^{+}), while being oxidized to gluconic acid.



Solutions of cupric ion (known as Fehling's solution) provide a simple test for the presence of sugars such as glucose. Sugars that react are called *reducing sugars;* those that do not are called *nonreducing sugars.* Reducing sugars can often nonspecifically react with other molecules. For instance, as a reducing sugar, glucose can react with hemoglobin to form glycosylated hemoglobin. Monitoring changes in the amount of glycosylated hemoglobin is an especially useful means of assessing the effectiveness of treatments for diabetes mellitus, a condition characterized by high levels of blood glucose (Section 27.3). Because the glycosylated hemoglobin remains in circulation, the amount of the modified hemoglobin corresponds to the long-term regulation—over several months—of glucose levels. In nondiabetic people, less than 6% of the hemoglobin is glycosylated, whereas, in uncontrolled diabetics, almost 10% of the hemoglobin is glycosylated. Although the glycosylation of hemoglobin has no effect on oxygen binding and is thus benign, similar reducing reactions between sugars and other

proteins are often detrimental to the body because the glycosylations alter the normal biochemical function of the modified proteins. Modifications known as *advanced glycosylation end products* (AGE) have been implicated in aging, arteriosclerosis, and diabetes, as well as other pathological conditions. AGE is the name given to a series of reactions between an amino group not participating in a peptide bond in a protein and the aldehyde form of a carbohydrate.

Monosaccharides are joined to alcohols and amines through glycosidic bonds

The biochemical properties of monosaccharides can by modified by reaction with other molecules. These modifications increase the biochemical versatility of carbohydrates, enabling them to serve as signal molecules or rendering them more susceptible to combustion. Three common reactants are alcohols, amines, and phosphates. A bond formed between the anomeric carbon atom of glucose and the oxygen atom of an alcohol is called a *glycosidic bond*—specifically, an O-*glycosidic bond*. *O*-Glycosidic bonds are prominent when carbohydrates are linked together to form long polymers and when they are attached to proteins. In addition, the anomeric carbon atom of a sugar can be linked to the nitrogen atom of an amine to form an N-*glycosidic bond*, such as when nitrogenous bases are attached to ribose units to form nucleosides. Examples of modified carbohydrates are shown in Figure 11.9.

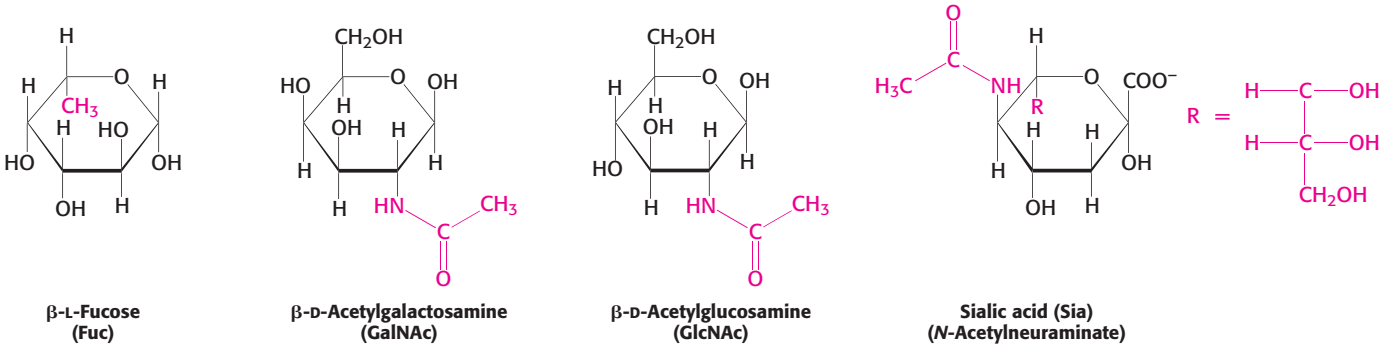
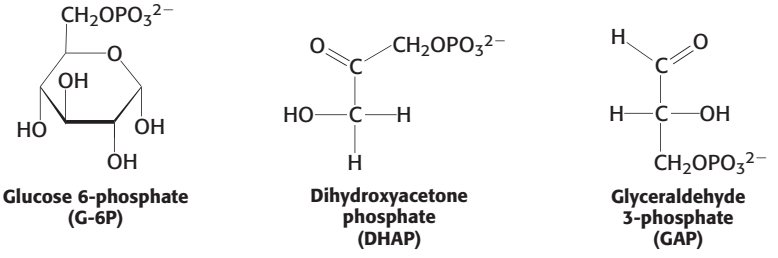


Figure 11.9 Modified monosaccharides. Carbohydrates can be modified by the addition of substituents (shown in red) other than hydroxyl groups. Such modified carbohydrates are often expressed on cell surfaces.

Phosphorylated sugars are key intermediates in energy generation and biosyntheses

One sugar modification deserves special note because of its prominence in metabolism. The addition of phosphoryl groups is a common modification of sugars. For instance, the first step in the breakdown of glucose to obtain energy is its conversion into glucose 6-phosphate. Several subsequent intermediates in this metabolic pathway, such as dihydroxyacetone phosphate and glyceraldehyde 3-phosphate, are phosphorylated sugars.



Phosphorylation makes sugars anionic; the negative charge prevents these sugars from spontaneously leaving the cell by crossing lipid-bilayer membranes. Phosphorylation also *creates reactive intermediates* that will more readily form linkages to other molecules. For example, a multiply phosphorylated derivative of ribose plays key roles in the biosyntheses of purine and pyrimidine nucleotides (Chapter 25).

11.2 Monosaccharides Are Linked to Form Complex Carbohydrates

Because sugars contain many hydroxyl groups, glycosidic bonds can join one monosaccharide to another. *Oligosaccharides* are built by the linkage of two or more monosaccharides by *O*-glycosidic bonds (Figure 11.10). In the disaccharide maltose, for example, two D-glucose residues are joined by a glycosidic linkage between the α-anomeric form of C-1 on one sugar and the hydroxyl oxygen atom on C-4 of the adjacent sugar. Such a linkage is called an α-1,4-glycosidic bond. Just as proteins have a polarity defined by the amino and carboxyl termini, oligosaccharides have a polarity defined by their reducing and nonreducing ends. The carbohydrate unit at the reducing end has a free anomeric carbon atom that has reducing activity because it can form the open-chain form, as discussed earlier (p. 325). By convention, this end of the oligosaccharide is still called the nonreducing end even when it is bound to another molecule such as a protein and thus no longer has reducing properties.

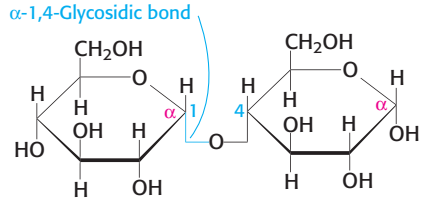
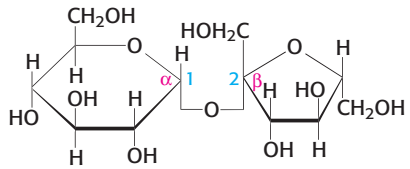


Figure 11.10 Maltose, a disaccharide. Two molecules of glucose are linked by an α-1,4-glycosidic bond to form the disaccharide maltose. The angles in the bonds to the central oxygen atom do not denote carbon atoms. The angles are added only for ease of illustration. The glucose molecule on the right is capable of assuming the open-chain form, which is capable of acting as a reducing agent. The glucose molecule on the left cannot assume the open-chain form, because the C-1 carbon atom is bound to another molecule.

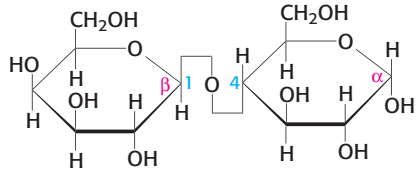
The fact that monosaccharides have multiple hydroxyl groups means that many different glycosidic linkages are possible. For example, consider three monosaccharides—glucose, mannose, and galactose. These molecules can be linked together in the laboratory to form more than 12,000 different structures differing in the order of the monosaccharides and the hydroxyl groups participating in the glycosidic linkages. For instance, the hydroxyl group on carbon 1 of one monosaccharide can link to carbons 4 or 6 of the next monosaccharide. In this section, we will look at some of the most common oligosaccharides found in nature.

Sucrose, lactose, and maltose are the common disaccharides

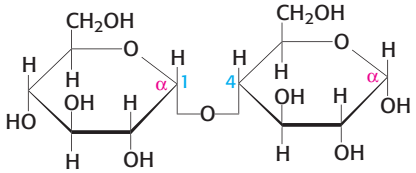
A *disaccharide* consists of two sugars joined by an *O*-glycosidic bond. Three abundant disaccharides that we encounter frequently are sucrose, lactose, and maltose (Figure 11.11). *Sucrose* (common table sugar) is obtained commercially from sugar cane or sugar beets. The anomeric carbon atoms of a glucose unit and a fructose unit are joined in this disaccharide; the configuration of this glycosidic linkage is α for glucose and β for fructose. Sucrose can be cleaved into its component monosaccharides by the enzyme *sucrase.*



Sucrose
(α-D-Glucopyranosyl-(1→2)-β-D-fructofuranose



Lactose
(β-D-Galactopyranosyl-(1→4)-α-D-glucopyranose



Maltose
(α-D-Glucopyranosyl-(1→4)-α-D-glucopyranose

Figure 11.11 Common disaccharides. Sucrose, lactose, and maltose are common dietary components. The angles in the bonds to the central oxygen atoms do not denote carbon atoms.

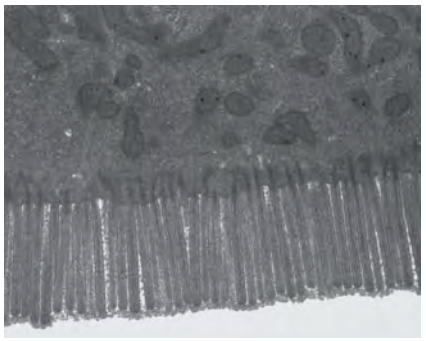


Figure 11.12 Electron micrograph of microvilli. Lactase and other enzymes that hydrolyze carbohydrates are present on microvilli that project from the outer face of the plasma membrane of intestinal epithelial cells. [From Louisa Howard and Katherine Connolly. Courtesy of Louisa Howard, Dartmouth College.]

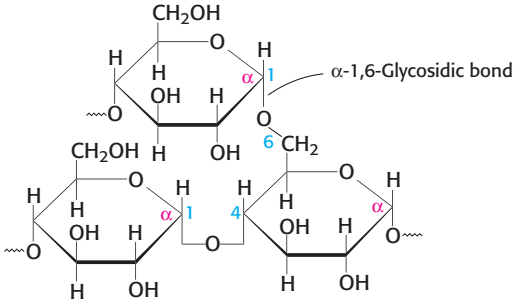


Figure 11.13 Branch point in glycogen. Two chains of glucose molecules joined by α-1,4-glycosidic bonds are linked by an α-1,6-glycosidic bond to create a branch point. Such an α-1,6-glycosidic bond forms at approximately every 10 glucose units, making glycogen a highly branched molecule.

Lactose, the disaccharide of milk, consists of galactose joined to glucose by a β-1,4-glycosidic linkage. Lactose is hydrolyzed to these monosaccharides by *lactase* in human beings and by *β-galactosidase* in bacteria. In *maltose*, two glucose units are joined by an α-1,4-glycosidic linkage. Maltose comes from the hydrolysis of large polymeric oligosaccharides such as starch and glycogen and is in turn hydrolyzed to glucose by *maltase*. Sucrase, lactase, and maltase are located on the outer surfaces of epithelial cells lining the small intestine (Figure 11.12). The cleavage products of sucrose, lactose, and maltose can be further processed to provide energy in the form of ATP.

Glycogen and starch are storage forms of glucose

Glucose is an important energy source in virtually all life forms. However, free glucose molecules cannot be stored because in high concentrations, glucose will disturb the osmotic balance of the cell, with the potential result being cell death. The solution is to store glucose as units in a large polymer, which is not osmotically active.

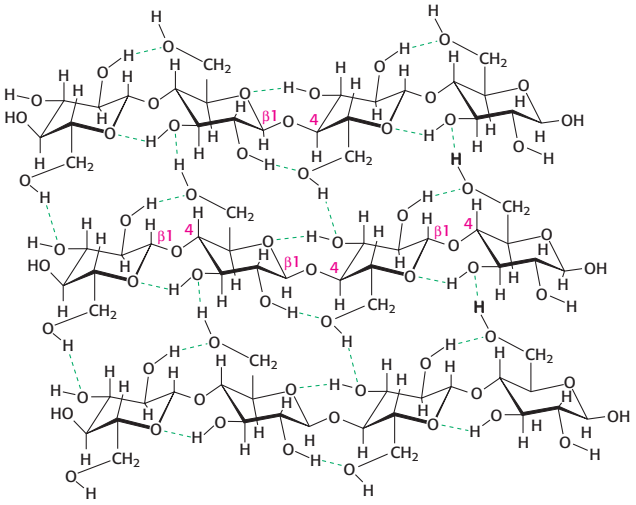
Large polymeric oligosaccharides, formed by the linkage of multiple monosaccharides, are called *polysaccharides* and play vital roles in energy storage and in maintaining the structural integrity of an organism. If all of the monosaccharide units in a polysaccharide are the same, the polymer is called a *homopolymer*. The most common homopolymer in animal cells is *glycogen,* the storage form of glucose. Glycogen is present in most of our tissues but is most common in muscle and liver. As will be considered in detail in Chapter 21, glycogen is a large, branched polymer of glucose residues. Most of the glucose units in glycogen are linked by α-1,4-glycosidic bonds. The branches are formed by α-1,6-glycosidic bonds, present about once in 10 units (Figure 11.13).

The nutritional reservoir in plants is the homopolymer *starch,* of which there are two forms. *Amylose,* the unbranched type of starch, consists of glucose residues in α-1,4 linkage. *Amylopectin,* the branched form, has about 1 α-1,6 linkage per 30 α-1,4 linkages, in similar fashion to glycogen except for its lower degree of branching. More than half the carbohydrate ingested by human beings is starch found in wheat, potatoes, and rice, to name just a few sources. Amylopectin, amylose, and glycogen are rapidly hydrolyzed by *α-amylase,* an enzyme secreted by the salivary glands and the pancreas.

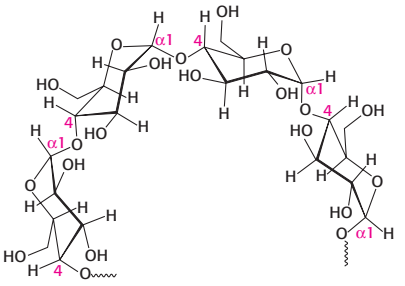
We have considered only homopolymers of glucose. However, given the variety of different monosaccharides that can be put together in any number of arrangements, the number of possible polysaccharides is huge. We will consider some of these polysaccharides shortly.

Cellulose, a structural component of plants, is made of chains of glucose

Cellulose, the other major polysaccharide of glucose found in plants, serves a structural rather than a nutritional role as an important component of the plant cell wall. *Cellulose is among the most abundant organic compounds in the biosphere.* Some 10^{15} kg of cellulose is synthesized and degraded on Earth each year, an amount 1000 times as great as the combined weight of the human race. Cellulose is an unbranched polymer of glucose residues joined by β-1,4 linkages, in contrast with the α-1,4 linkage seen in starch and glycogen. This simple difference in stereochemistry yields two molecules with vastly different properties and biological functions. The β configuration



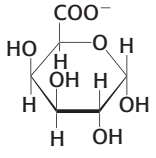
Cellulose
(β-1,4 linkages)



Starch and glycogen
(α-1,4 linkages)

Figure 11.14 Glycosidic bonds determine polysaccharide structure. The β-1,4 linkages favor straight chains, which are optimal for structural purposes. The α-1,4 linkages favor bent structures, which are more suitable for storage.

allows cellulose to form very long, straight chains. Fibrils are formed by parallel chains that interact with one another through hydrogen bonds, generating a rigid, supportive structure. The straight chains formed by β linkages are optimal for the construction of fibers having a high tensile strength. The α-1,4 linkages in glycogen and starch produce a very different molecular architecture: a hollow helix is formed instead of a straight chain (Figure 11.14). The hollow helix formed by α linkages is well suited to the formation of a more-compact, accessible store of sugar. Although mammals lack cellulases and therefore cannot digest wood and vegetable fibers, cellulose and other plant fibers are still an important constituent of the mammalian diet as a component of dietary fiber. Soluble fiber such as *pectin* (polygalacturonic acid) slows the movement of food through the gastrointestinal tract, allowing improved digestion and the absorption of nutrients. Insoluble fibers, such as cellulose, increase the rate at which digestion products pass through the large intestine. This increase in rate can minimize exposure to toxins in the diet. Cellulose is currently being investigated as potential source of ethanol for biofuels.

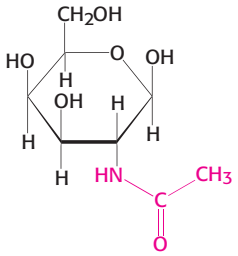


Galacturonic acid

11.3 Carbohydrates Can Be Linked to Proteins to Form Glycoproteins

A carbohydrate group can be covalently attached to a protein to form a *glycoprotein*. We will examine three classes of glycoproteins. The first class is simply referred to as glycoproteins. In glycoproteins of this class, the protein constituent is the largest component by weight. This versatile class plays a variety of biochemical roles. Many glycoproteins are components of cell membranes, where they take part in processes such as cell adhesion and the binding of sperm to eggs. Other glycoproteins are formed by linking carbohydrates to soluble proteins. In particular, many of the proteins secreted from cells are glycosylated, or modified by the attachment of carbohydrates, including most proteins present in the serum component of blood.

The second class of glycoproteins comprises the *proteoglycans*. The protein component of proteoglycans is conjugated to a particular type of polysaccharide called a *glycosaminoglycan*. Carbohydrates make up a much



β-D-Acetylgalactosamine (GalNAc)

larger percentage by weight of the proteoglycan compared with simple glycoproteins. Proteoglycans function as structural components and lubricants.

Mucins, or *mucoproteins,* are, like proteoglycans, predominately carbohydrate. *N*-Acetylgalactosamine is usually the carbohydrate moiety bound to the protein in mucins. *N*-Acetylgalactosamine is an example of an *amino sugar,* so named because an amino group replaces a hydroxyl group. Mucins, a key component of mucus, serve as lubricants.

Glycosylation greatly increases the complexity of the proteome. A given protein with several potential glycosylation sites can have many different glycosylated forms (sometimes called *glycoforms*), each of which can be generated only in a specific cell type or developmental stage.

Carbohydrates can be linked to proteins through asparagine (*N*-linked) or through serine or threonine (*O*-linked) residues

Sugars in glycoproteins are attached either to the amide nitrogen atom in the side chain of asparagine (termed an N-*linkage*) or to the oxygen atom in the side chain of serine or threonine (termed an O-*linkage*), as shown in Figure 11.15. An asparagine residue can accept an oligosaccharide only if the residue is part of an Asn-X-Ser or Asn-X-Thr sequence, in which X can be any residue, except proline. Thus, *potential glycosylation sites can be detected within amino acid sequences.* However, not all potential sites are glycosylated. Which sites are glycosylated depends on other aspects of the protein structure and on the cell type in which the protein is expressed. All *N*-linked oligosaccharides have in common a pentasaccharide core consisting of three mannose and two *N*-acetylglucosamine residues. Additional sugars are attached to this core to form the great variety of oligosaccharide patterns found in glycoproteins (Figure 11.16).

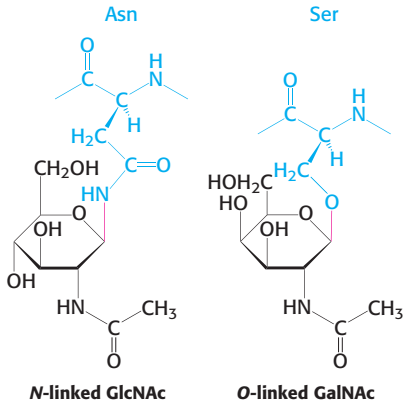


Figure 11.15 Glycosidic bonds between proteins and carbohydrates. A glycosidic bond links a carbohydrate to the side chain of asparagine (*N*-linked) or to the side chain of serine or threonine (*O*-linked). The glycosidic bonds are shown in red.

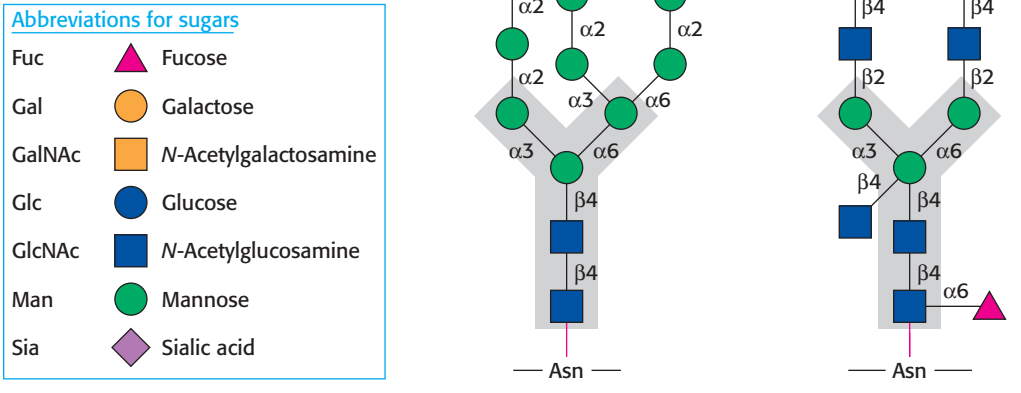


Figure 11.16 *N*-linked oligosaccharides. A pentasaccharide core (shaded gray) is common to all *N*-linked oligosaccharides and serves as the foundation for a wide variety of *N*-linked oligosaccharides, two of which are illustrated: (A) high-mannose type; (B) complex type.

The glycoprotein erythropoietin is a vital hormone

Let us look at a glycoprotein present in the blood serum that has dramatically improved treatment for anemia, particularly that induced by cancer chemotherapy. The glycoprotein hormone *erythropoietin*

(EPO) is secreted by the kidneys and stimulates the production of red blood cells. EPO is composed of 165 amino acids and is *N*-glycosylated at three asparagine residues and *O*-glycosylated on a serine residue (Figure 11.17). The mature EPO is 40% carbohydrate by weight, and glycosylation enhances the stability of the protein in the blood. Unglycosylated protein has only about 10% of the bioactivity of the glycosylated form because the protein is rapidly removed from the blood by the kidneys. The availability of recombinant human EPO has greatly aided the treatment of anemias. However, some endurance athletes have used recombinant human EPO to increase the red-blood-cell count and hence their oxygen-carrying capacity. Drug-testing laboratories are able to distinguish some forms of prohibited human recombinant EPO from natural EPO in athletes by detecting differences in their glycosylation patterns through the use of isoelectric focusing (p. 75).

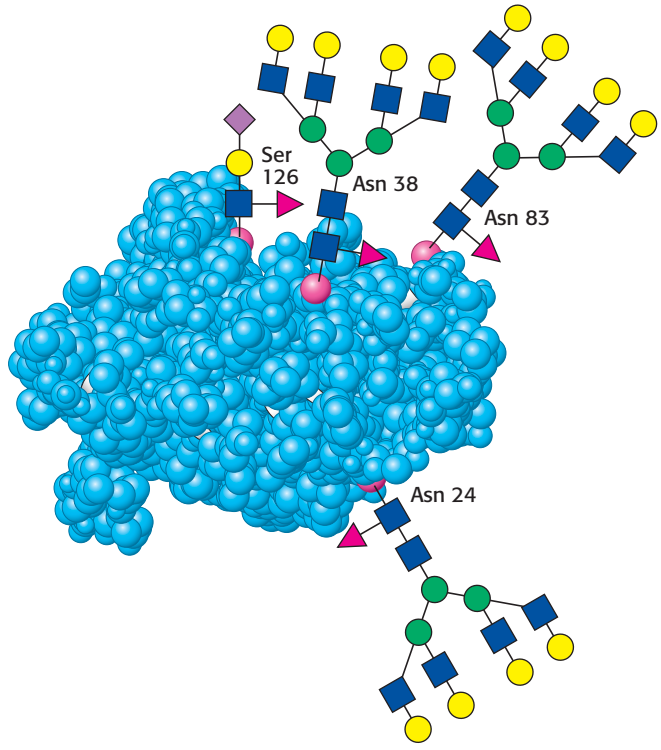


Figure 11.17 Oligosaccharides attached to erythropoietin. Erythropoietin has oligosaccharides linked to three asparagine residues and one serine residue. The structures shown are approximately to scale. See Figure 11.16 for the carbohydrate key. [Drawn from 1BUY.pdf.]

Proteoglycans, composed of polysaccharides and protein, have important structural roles

As stated earlier, proteoglycans are proteins attached to glycosaminoglycans. The glycosaminoglycan makes up as much as 95% of the biomolecule by weight, and so the proteoglycan resembles a polysaccharide more than a protein. Proteoglycans not only function as lubricants and structural components in connective tissue, but also mediate the adhesion of cells to the extracellular matrix, and bind factors that stimulate cell proliferation.

The properties of proteoglycans are determined primarily by the glycosaminoglycan component. Many glycosaminoglycans are made of repeating units of disaccharides containing a derivative of an amino sugar, either glucosamine or galactosamine (Figure 11.18).

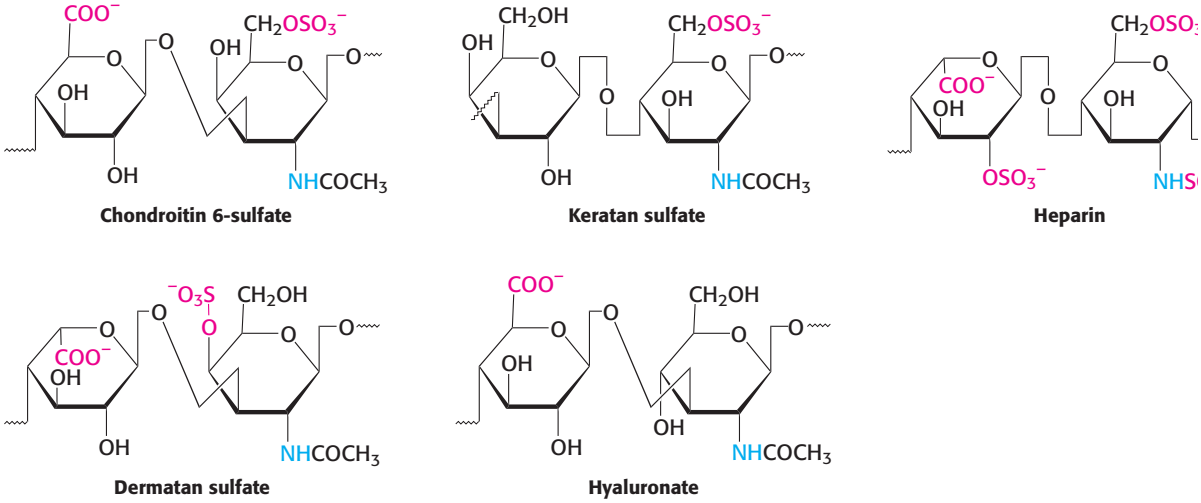


Figure 11.18 Repeating units in glycosaminoglycans. Structural formulas for five repeating units of important glycosaminoglycans illustrate the variety of modifications and linkages that are possible. Amino groups are shown in blue and negatively charged groups in red. Hydrogen atoms have been omitted for clarity. The right-hand structure is a glucosamine derivative in each case.

Figure 11.19 Hurler disease. Formerly called gargoylism, Hurler disease is a mucopolysaccharidosis having symptoms that include wide nostrils, a depressed nasal bridge, thick lips and earlobes, and irregular teeth. In Hurler disease, glycosaminoglycans cannot be degraded. The excess of these molecules are stored in the soft tissue of the facial regions, resulting in the characteristic facial features. [Courtesy National MPS Society, www.mpssociety.org.]

At least one of the two sugars in the repeating unit has a *negatively charged carboxylate or sulfate group.* The major glycosaminoglycans in animals are chondroitin sulfate, keratan sulfate, heparin, heparan sulfate, dermatan sulfate, and hyaluronate. *Mucopolysaccharidoses* are a collection of diseases, such as Hurler disease, that result from the inability to degrade glycosaminoglycans (Figure 11.19). Although precise clinical features vary with the disease, all mucopolysaccharidoses result in skeletal deformities and reduced life expectancies.

Proteoglycans are important components of cartilage

Among the best-characterized members of this diverse class is the proteoglycan in the extracellular matrix of cartilage. The proteoglycan *aggrecan* and the protein *collagen* are key components of cartilage. The triple helix of collagen (p. 43) provides structure and tensile strength, whereas aggrecan serves as a shock absorber. The protein component of aggrecan is a large molecule composed of 2397 amino acids. The protein has three globular domains, and the site of glycosaminoglycan attachment is the extended region between globular domains 2 and 3. This linear region contains highly repetitive amino acid sequences, which are sites for the attachment of keratan sulfate and chondroitin sulfate. Many molecules of aggrecan are in turn noncovalently bound through the first globular domain to a very long filament formed by linking together molecules of the glycosaminoglycan hyaluronan (Figure 11.20). Water is bound to the glycosaminoglycans, attracted by the many negative charges. Aggrecan can cushion compressive forces because the absorbed water enables it to spring back after having been deformed. When pressure is exerted, as when the foot hits the ground while walking, water is squeezed from the glycosaminoglycan, cushioning the impact. When the pressure is released, the water rebinds. *Osteoarthritis* can result from the proteolytic degradation of aggrecan and collagen in the cartilage.

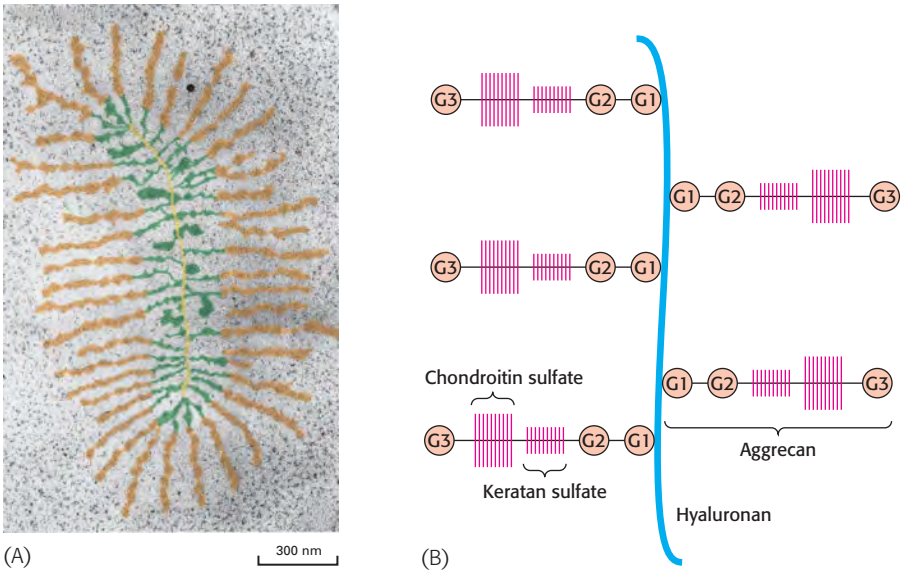


Figure 11.20 Structure of proteoglycan from cartilage. (A) Electron micrograph of a proteoglycan from cartilage (with false color added). Proteoglycan monomers emerge laterally at regular intervals from opposite sides of a central filament of hyaluronan. (B) Schematic representation. G = globular domain. [(A) Courtesy of Dr. Lawrence Rosenberg. From J. A. Buckwalter and L. Rosenberg. *Collagen Relat. Res.* 3:489–504, 1983.]

In addition to being a key component of structural tissues, glycosaminoglycans are common throughout the biosphere. Chitin is a glycosaminoglycan found in the exoskeleton of insects, crustaceans, and arachnids and is, next to cellulose, the second most abundant polysaccharide in nature (Figure 11.21).

Figure 11.21 Chitin, a glycosaminoglycan, is present in insect wings and the exoskeleton. Glycosaminoglycans are components of the exoskeletons of insects, crustaceans, and arachnids. [FLPA/Alamy.]

Mucins are glycoprotein components of mucus

As stated earlier, another class of glycoproteins consists of the *mucins* (mucoproteins). In mucins, the protein component is extensively glycosylated to serine or threonine residues by *N*-acetylgalactosamine (see Figure 11.9). Mucins are capable of forming large polymeric structures and are common in mucous secretions. These glycoproteins are synthesized by specialized cells in the tracheobronchial, gastrointestinal, and genitourinary tracts. Because a key function of mucins is to act as a lubricant, mucins are abundant in saliva.

A model of a mucin is shown in Figure 11.22A. The defining feature of the mucins is a region of the protein backbone termed the *variable number of tandem repeats* (VNTR) region, which is rich in serine and threonine residues that are *O*-glycosylated. Indeed, the carbohydrate moiety can account for as much as 80% of the molecule by weight. A number of core carbohydrate structures are conjugated to the protein component of mucin. Figure 11.22B shows one such structure.

Mucins adhere to epithelial cells and act as a protective barrier; they also hydrate the underlying cells. In addition to protecting cells from environmental insults, such as stomach acid, inhaled chemicals in the lungs, and bacterial infections, mucins have roles in fertilization, the immune response, and cell adhesion. Mucins are overexpressed in bronchitis and cystic fibrosis, and the overexpression of mucins is characteristic of adenocarcinomas—cancers of the glandular cells of epithelial origin.

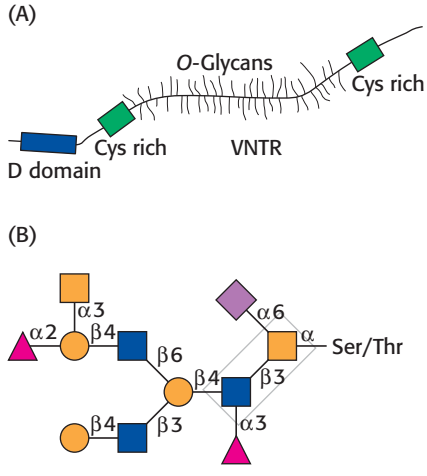


Figure 11.22 Mucin structure. (A) A schematic representation of a mucoprotein. The VNTR region is highly glycosylated, forcing the molecule into an extended conformation. The Cys-rich domains and the D domain facilitate the polymerization of many such molecules. (B) An example of an oligosaccharide that is bound to the VNTR region of the protein. [After A. Varki et al. (Eds.), *Essentials of Glycobiology*, 2d ed. (Cold Spring Harbor Press, 2009), pp. 117, 118.]

Protein glycosylation takes place in the lumen of the endoplasmic reticulum and in the Golgi complex

The major pathway for protein glycosylation takes place inside the lumen of the *endoplasmic reticulum* (ER) and in the *Golgi complex*, organelles that play central roles in protein trafficking (Figure 11.23). The protein is synthesized by ribosomes attached to the cytoplasmic face of the ER membrane, and the

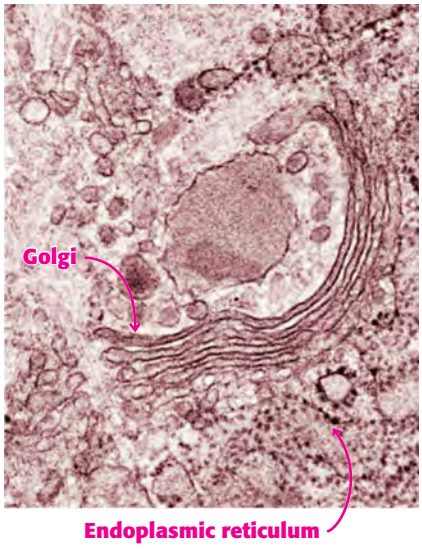
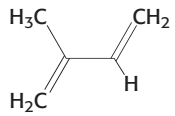


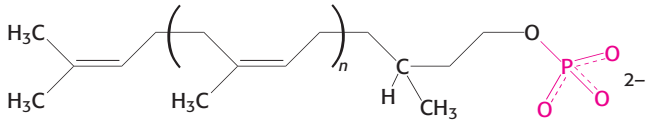
Figure 11.23 Golgi complex and endoplasmic reticulum. The electron micrograph shows the Golgi complex and adjacent endoplasmic reticulum. The black dots on the cytoplasmic surface of the ER membrane are ribosomes. [Micrograph courtesy of Lynne Mercer.]



Isoprene

peptide chain is inserted into the lumen of the ER (Section 30.6). The *N*-linked glycosylation begins in the ER and continues in the Golgi complex, whereas the *O*-linked glycosylation takes place exclusively in the Golgi complex.

A large oligosaccharide destined for attachment to the asparagine residue of a protein is assembled on *dolichol phosphate*, a specialized lipid molecule located in the ER membrane and containing about 20 isoprene (C_5) units.



n = 15–19

Dolichol phosphate

The terminal phosphate group of the dolichol phosphate is the site of attachment of the activated oligosaccharide, which is subsequently transferred to a specific asparagine residue of the growing polypeptide chain. Both the activated sugars and the complex enzyme that is responsible for transferring the oligosaccharide to the protein are located on the lumenal side of the ER. Thus, proteins in the cytoplasm are not glycosylated by this pathway.

Proteins in the lumen of the ER and in the ER membrane are transported to the Golgi complex, which is a stack of flattened membranous sacs. *Carbohydrate units of glycoproteins are altered and elaborated in the Golgi complex.* The *O*-linked sugar units are fashioned there, and the *N*-linked sugars, arriving from the ER as a component of a glycoprotein, are modified in many different ways. *The Golgi complex is the major sorting center of the cell.* Proteins proceed from the Golgi complex to lysosomes, secretory granules, or the plasma membrane, according to signals encoded within their amino acid sequences and three-dimensional structures (Figure 11.24).

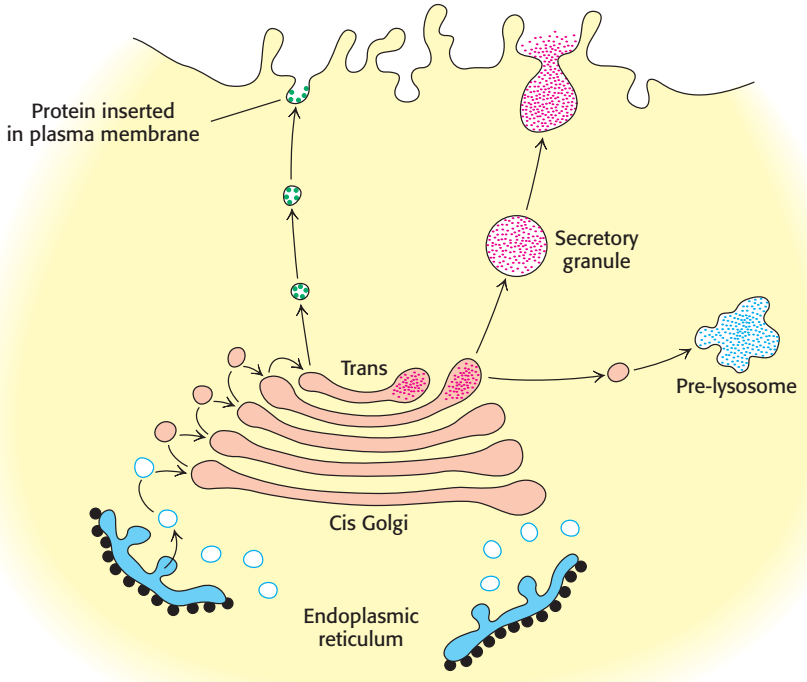


Figure 11.24 Golgi complex as sorting center. The Golgi complex is the sorting center in the targeting of proteins to lysosomes, secretory vesicles, and the plasma membrane. The cis face of the Golgi complex receives vesicles from the endoplasmic reticulum, and the trans face sends a different set of vesicles to target sites. Vesicles also transfer proteins from one compartment of the Golgi complex to another. [Courtesy of Dr. Marilyn Farquhar.]

Specific enzymes are responsible for oligosaccharide assembly

How are the complex carbohydrates formed, be they unconjugated molecules such as glycogen or components of glycoproteins? Complex carbohydrates are synthesized through the action of specific enzymes, *glycosyltransferases,* which catalyze the formation of glycosidic bonds. Given the diversity of known glycosidic linkages, many different enzymes are required. Indeed, glycosyltransferases account for 1% to 2% of gene products in all organisms examined.

The general form of the reaction catalyzed by a glycosyltransferase is shown in Figure 11.25. The sugar to be added comes in the form of an activated (energy-rich) sugar nucleotide, such as UDP-glucose (UDP is the abbreviation for uridine diphosphate). The attachment of a nucleotide to enhance the energy content of a molecule is a common strategy in biosynthesis that we will see many times in our study of biochemistry. The acceptor substrates for glycosyltransferases are quite varied and include carbohydrates, serine, threonine and asparagine residues of proteins, lipids, and even nucleic acids.

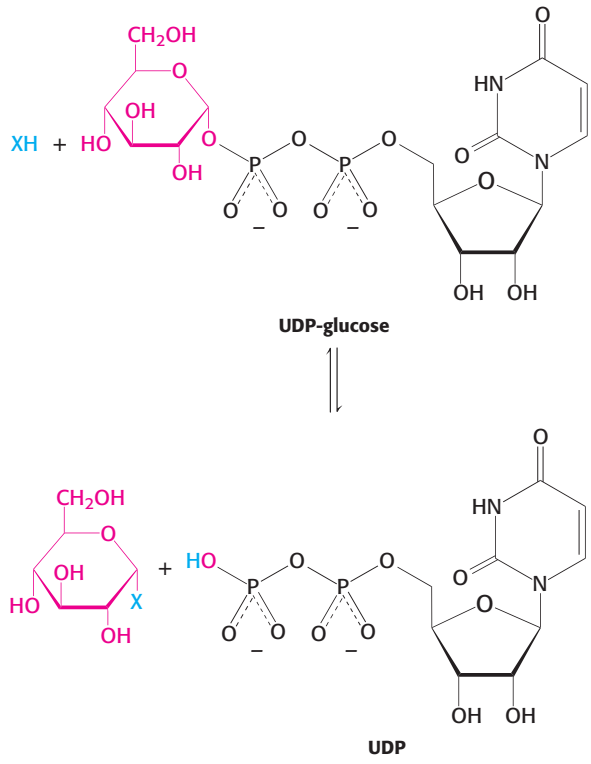


Figure 11.25 General form of a glycosyltransferase reaction. The sugar to be added comes from a sugar nucleotide—in this case, UDP-glucose. The acceptor, designated X in this illustration, can be one of a variety of biomolecules, including other carbohydrates or proteins.

Blood groups are based on protein glycosylation patterns

The human ABO blood groups illustrate the effects of glycosyltransferases on the formation of glycoproteins. Each blood group is designated by the presence of one of the three different carbohydrates, termed A, B, or O, attached to glycoproteins and glycolipids on the surfaces of red blood cells (Figure 11.26). These structures have in common an oligosaccharide foundation called the O (or sometimes H) antigen. The A and B antigens differ from the O antigen by the addition of one extra monosaccharide, either *N*-acetylgalactosamine (for A) or galactose (for B) through an α-1,3 linkage to a galactose moiety of the O antigen.

Specific glycosyltransferases add the extra monosaccharide to the O antigen. Each person inherits the gene for one glycosyltransferase of this type from each parent. The type A transferase specifically adds *N*-acetylgalactosamine, whereas the type B transferase adds galactose. These enzymes are identical in all but 4 of 354 positions. The O phenotype is the result of a mutation that leads to premature termination of translation and, hence, to the production of neither of the required glycosyltransferases.

These structures have important implications for blood transfusions and other transplantation procedures. If an antigen not normally present in a person is introduced, the person's immune system recognizes it as foreign. Red-blood-cell lysis occurs rapidly, leading to a severe drop in blood pressure (hypotension), shock, kidney failure, and death from circulatory collapse.

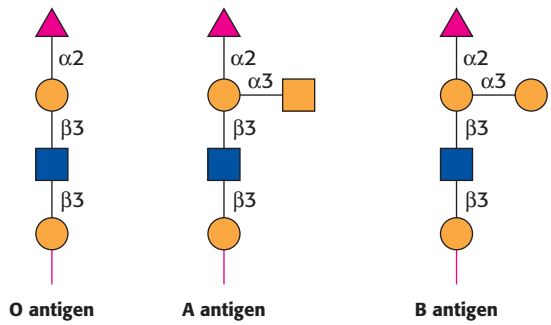


Figure 11.26 Structures of A, B, and O oligosaccharide antigens. The carbohydrate structures shown are depicted symbolically by employing a scheme (see the key in Figure 11.16) that is becoming widely used.

Why are different blood types present in the human population? Suppose that a pathogenic organism such as a parasite expresses on its cell surface a carbohydrate antigen similar to one of the blood-group antigens. This antigen may not be readily detected as foreign in a person whose blood type matches the parasite antigen, and the parasite will flourish. However, other people with different blood types will be protected. Hence,

there will be selective pressure on human beings to vary blood type to prevent parasitic mimicry and a corresponding selective pressure on parasites to enhance mimicry. The constant "arms race" between pathogenic microorganisms and human beings drives the evolution of diversity of surface antigens within the human population.

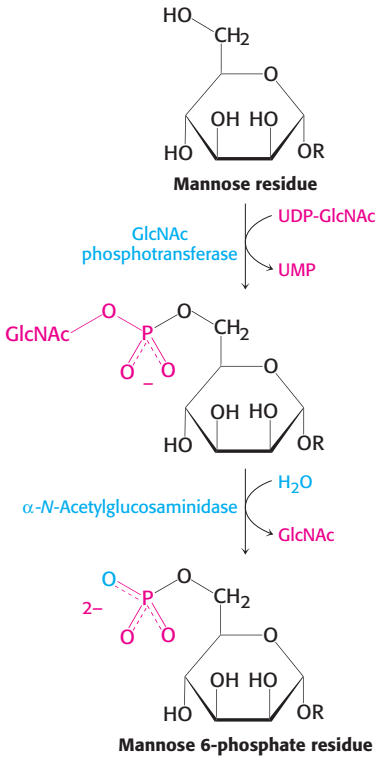


Figure 11.27 Formation of a mannose 6-phosphate marker. A glycoprotein destined for delivery to lysosomes acquires a phosphate marker in the Golgi compartment in a two-step process. First, GlcNAc phosphotransferase adds a phospho-*N*-acetylglucosamine unit to the 6-OH group of a mannose, and then an *N*-acetylglucosaminidase removes the added sugar to generate a mannose 6-phosphate residue in the core oligosaccharide.

Errors in glycosylation can result in pathological conditions

Although the role of carbohydrate attachment to proteins is not known in detail in most cases, data indicate that this glycosylation is important for the processing and stability of these proteins, as it is for EPO. For instance, certain types of muscular dystrophy can be traced to improper glycosylation of membrane proteins. Indeed, an entire family of severe inherited human disease called *congenital disorders of glycosylation* has been identified. These pathological conditions reveal the importance of proper modification of proteins by carbohydrates and their derivatives.

An especially clear example of the role of glycosylation is provided by *I-cell disease* (also called *mucolipidosis II*), a lysosomal storage disease. Normally, a carbohydrate marker directs certain digestive enzymes from the Golgi complex to lysosomes where they normally function. *Lysosomes* are organelles that degrade and recycle damaged cellular components or material brought into the cell by endocytosis. In patients with I-cell disease, lysosomes contain large *inclusions* of undigested glycosaminoglycans and glycolipids—hence the "I" in the name of the disease. These inclusions are present because the enzymes normally responsible for the degradation of glycosaminoglycans are missing from affected lysosomes. Remarkably, the enzymes are present at very high levels in the blood and urine. Thus, active enzymes are synthesized, but, in the absence of appropriate glycosylation, they are exported instead of being sequestered in lysosomes. In other words, *in I-cell disease, a whole series of enzymes are incorrectly addressed and delivered to the wrong location.* Normally, these enzymes contain a mannose 6-phosphate residue, a component of an *N*-oligosaccharide attached to proteins bound for the lysosome. In I-cell disease, however, the attached mannose lacks a phosphate (Figure 11.27). Mannose 6-phosphate is in fact the marker that normally directs many hydrolytic enzymes from the Golgi complex to lysosomes. I-cell patients are deficient in the *N*-acetylglucosamine phosphotransferase catalyzing the first step in the addition of the phosphoryl group; the consequence is the mistargeting of eight essential enzymes. I-cell disease causes the patient to suffer severe psychomotor retardation and skeletal deformities, similar to those in Hurler disease.

Oligosaccharides can be "sequenced"

How is it possible to determine the structure of a glycoprotein—the oligosaccharide structures and their points of attachment? Most approaches make use of enzymes that cleave oligosaccharides at specific types of linkages.

The first step is to detach the oligosaccharide from the protein. For example, *N*-linked oligosaccharides can be released from proteins by an enzyme such as *peptide N-glycosidase F,* which cleaves the *N*-glycosidic bonds linking the oligosaccharide to the protein. The oligosaccharides can then be isolated and analyzed. MALDI-TOF or other mass spectrometric techniques (Section 3.4) provide the mass of an oligosaccharide fragment. However, many possible oligosaccharide structures are consistent with a given mass. More-complete information can be obtained by cleaving the oligosaccharide with enzymes of varying specificities. For example, *β-1,4-galactosidase* cleaves β-glycosidic bonds exclusively at galactose residues.

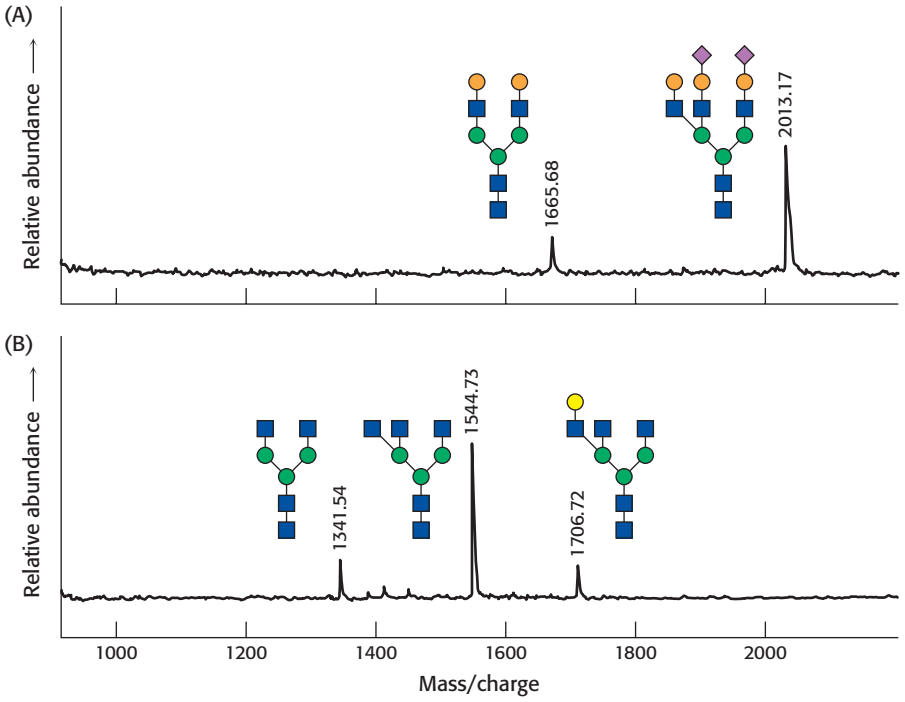


Figure 11.28 Mass spectrometric "sequencing" of oligosaccharides. Carbohydrate-cleaving enzymes were used to release and specifically cleave the oligosaccharide component of the glycoprotein fetuin from bovine serum. Parts A and B show the masses obtained with MALDI-TOF spectrometry as well as the corresponding structures of the oligosaccharide-digestion products (using the same scheme as that in Figure 11.16): (A) digestion with peptide *N*-glycosidase F (to release the oligosaccharide from the protein) and neuraminidase; (B) digestion with peptide *N*-glycosidase F, neuraminidase, and β-1,4-galactosidase. Knowledge of the enzyme specificities and the masses of the products permits the characterization of the oligosaccharide. See Figure 11.16 for the carbohydrate key. [After A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, G. W. Hart, and J. Marth (Eds.), *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 1999), p. 596.]

The products can again be analyzed by mass spectrometry (Figure 11.28). The repetition of this process with the use of an array of enzymes of different specificity will eventually reveal the structure of the oligosaccharide.

Proteases applied to glycoproteins can reveal the points of oligosaccharide attachment. Cleavage by a specific protease yields a characteristic pattern of peptide fragments that can be analyzed chromatographically. Fragments attached to oligosaccharides can be picked out because their chromatographic properties will change on glycosidase treatment. Mass spectrometric analysis or direct peptide sequencing can reveal the identity of the peptide in question and, with additional effort, the exact site of oligosaccharide attachment.

Now that the sequencing of the human genome is complete, the characterization of the much more complex proteome, including the biological roles of specifically modified proteins, can begin in earnest.

11.4 Lectins Are Specific Carbohydrate-Binding Proteins

The diversity and complexity of the carbohydrate units and the variety of ways in which they can be joined in oligosaccharides and polysaccharides suggest that they are functionally important. Nature does not construct complex patterns when simple ones suffice. Why all this intricacy and diversity? It is now clear that these carbohydrate structures are the recognition sites for a special class of proteins. Such proteins, termed *glycan-binding*

proteins, bind specific carbohydrate structures on neighboring cell surfaces. Originally discovered in plants, glycan-binding proteins are ubiquitous, and no living organisms have been found that lack these key proteins. We will focus on a particular class of glycan-binding proteins termed *lectins* (from Latin *legere,* "to select"). The interaction of lectins with their carbohydrate partners is another example of carbohydrates being information-rich molecules that guide many biological processes. The diverse carbohydrate structures displayed on cell surfaces are well suited to serving as sites of interaction between cells and their environments. Interestingly, the partners for lectin binding are often the carbohydrate moiety of glycoproteins.

Lectins promote interactions between cells

Cell–cell contact is a vital interaction in a host of biochemical functions, ranging from building a tissue from isolated cells to facilitating the transmission of information. The chief function of lectins, carbohydrate-binding proteins, is to facilitate cell–cell contact. A lectin usually contains two or more binding sites for carbohydrate units. These carbohydrate-binding sites on the surface of one cell interact with arrays of carbohydrates displayed on the surface of another cell. Lectins and carbohydrates are linked by a number of weak noncovalent interactions that ensure specificity yet permit unlinking as needed. The weak interactions between one cell surface and another resemble the action of Velcro; each interaction is weak, but the composite is strong.

We have already met a lectin obliquely. Recall that, in I-cell disease, lysosomal enzymes lack the appropriate mannose 6-phosphate, a molecule that directs the enzymes to the lysosome. Under normal circumstance, the *mannose 6-phosphate receptor,* a lectin, binds the enzymes in the Golgi apparatus and directs them to the lysosome.

Lectins are organized into different classes

Lectins can be divided into classes on the basis of their amino acid sequences and biochemical properties. One large class is the C type (for *c*alcium-requiring) found in animals. These proteins each have a homologous domain of 120 amino acids that is responsible for carbohydrate binding. The structure of one such domain bound to a carbohydrate target is shown in Figure 11.29.

A calcium ion on the protein acts as a bridge between the protein and the sugar through direct interactions with sugar OH groups. In addition, two glutamate residues in the protein bind to both the calcium ion and the

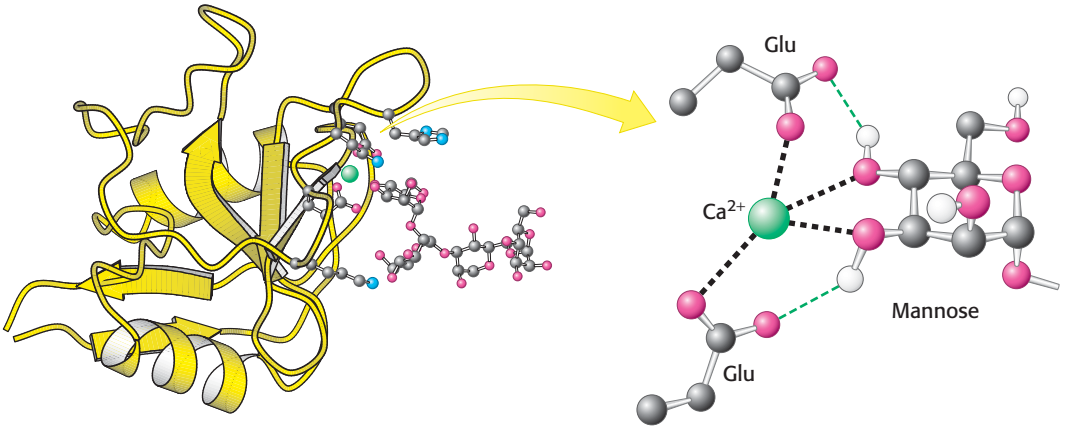


Figure 11.29 Structure of a C-type carbohydrate-binding domain of an animal lectin. *Notice* that a calcium ion links a mannose residue to the lectin. Selected interactions are shown, with some hydrogen atoms omitted for clarity. [Drawn from 2MSC.pdb.]

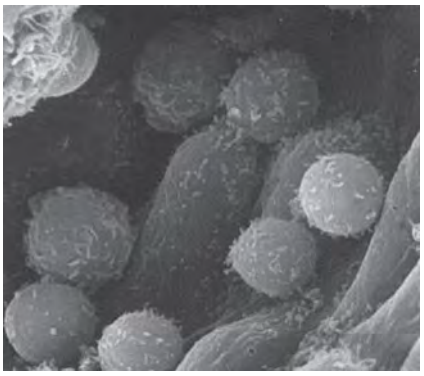


Figure 11.30 Selectins mediate cell–cell interactions. The scanning electron micrograph shows lymphocytes adhering to the endothelial lining of a lymph node. The L selectins on the lymphocyte surface bind specifically to carbohydrates on the lining of the lymph-node vessels [Courtesy of Dr. Eugene Butcher.]

sugar, and other protein side chains form hydrogen bonds with other OH groups on the carbohydrate. The carbohydrate-binding specificity of a particular lectin is determined by the amino acid residues that bind the carbohydrate.

Proteins termed *selectins* are members of the C-type family. Selectins bind immune-system cells to sites of injury in the inflammatory response (Figure 11.30). The L, E, and P forms of selectins bind specifically to carbohydrates on *l*ymph-node vessels, *e*ndothelium, or activated blood *p*latelets, respectively. New therapeutic agents that control inflammation may emerge from a deeper understanding of how selectins bind and distinguish different carbohydrates. L-Selectin, originally thought to participate only in the immune response, is produced by embryos when they are ready to attach to the endometrium of the mother's uterus. For a short period of time, the endometrial cells present an oligosaccharide on the cell surface. When the embryo attaches through lectins, the attachment activates signal pathways in the endometrium to make implantation of the embryo possible.

Another large class of lectins comprises the L-lectins. These lectins are especially rich in the seeds of leguminous plants, and many of the initial biochemical characterizations of lectins were performed on this readily available lectin. Although the exact role of lectins in plants is unclear, they can serve as potent insecticides. Other L-type lectins, such as *calnexin* and *calreticulin,* are prominent chaperones in the eukaryotic endoplasmic reticulum. Recall that chaperones are proteins that facilitate the folding of other proteins.

Influenza virus binds to sialic acid residues

Many pathogens gain entry into specific host cells by adhering to cell-surface carbohydrates. For example, influenza virus recognizes sialic acid residues linked to galactose residues that are present on cell-surface glycoproteins. The viral protein that binds to these sugars is called *hemagglutinin* (Figure 11.31).

After binding hemagglutinin, the virus is engulfed by the cell and begins to replicate. To exit the cell, the new virions must bind to hemaglutinin in what is essentially the reverse of viral entry. Another viral protein, neuraminidase (sialidase), cleaves the glycosidic bonds to the sialic acid residues of hemagglutinin, freeing the virus to infect new cells, spreading the infection throughout the respiratory tract. Inhibitors of this enzyme such as oseltamivir (Tamiflu) and zanamivir (Relenza) are important anti-influenza agents.

Viral hemagglutinin's carbohydrate-binding specificity may play an important role in species specificity of infection and ease of transmission. For instance, avian influenza H5N1 (bird flu) is especially lethal and is

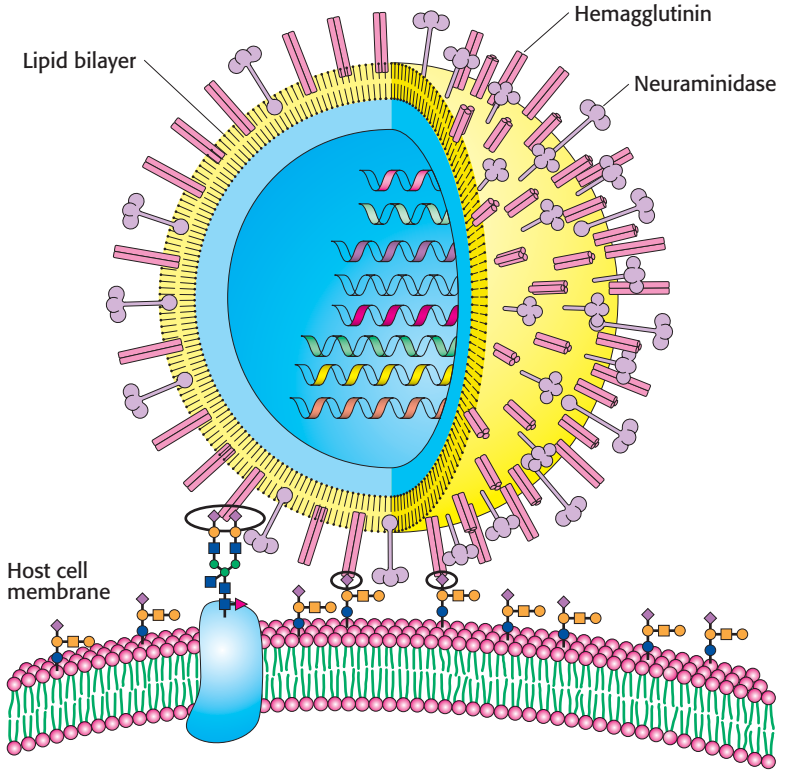


Figure 11.31 Viral receptors. Influenza virus targets cells by binding to sialic acid residues (purple diamonds) located at the termini of oligosaccharides present on cell-surface glycoproteins and glycolipids. These carbohydrates are bound by hemagglutinin (interaction circles), one of the major proteins expressed on the surface of the virus. The other major viral-surface protein, neuraminidase, is an enzyme that cleaves oligosaccharide chains to release the viral particle at a later stage of the viral life cycle.

readily spread from bird to bird. Although human beings can be infected by this virus, infection is rare and human-to-human transmission is rarer still. The biochemical basis of these characteristics is that the avian-virus hemagglutinin recognizes a different carbohydrate sequence from that recognized in human influenza. Although human beings have the sequence to which the avian virus binds, it is located deep in the lungs. Infection by the avian virus is thus difficult, and, when it does occur, the avian virus is not readily transmitted by sneezing or coughing.

Plasmodium falciparum, the parasitic protozoan that causes malaria, also relies on glycan binding to infect and colonize its host. Glycan-binding proteins of the parasitic form initially injected by the mosquito bind to the glycosaminoglycan heparin sulfate on the liver, initiating the parasite's entry into the cell. On exiting from the liver later in its life cycle, the parasite invades red blood cells by using another glycan-binding protein to bind to the carbohydrate moiety of glycophorin, a prominent membrane glycoprotein in red blood cells. Developing means to disrupt the carbohydrate interactions between pathogens and host cells may prove to be clinically useful.

Summary

11.1 Monosaccharides Are the Simplest Carbohydrates

Carbohydrates are aldoses or ketoses that are rich in hydroxyl groups. An aldose is a carbohydrate with an aldehyde group (as in glyceraldehyde and glucose), whereas a ketose contains a keto group (as in dihydroxyacetone and fructose). A sugar belongs to the D series if the absolute configuration of its asymmetric carbon atom farthest from the aldehyde or keto group is the same as that of D-glyceraldehyde. Most naturally occurring sugars belong to the D series. The C-1 alde-

hyde in the open-chain form of glucose reacts with the C-5 hydroxyl group to form a six-membered pyranose ring. The C-2 keto group in the open-chain form of fructose reacts with the C-5 hydroxyl group to form a five-membered furanose ring. Pentoses such as ribose and deoxyribose also form furanose rings. An additional asymmetric center is formed at the anomeric carbon atom (C-1 in aldoses and C-2 in ketoses) in these cyclizations. The hydroxyl group attached to the anomeric carbon atom is on the opposite side of the ring from the CH_2OH group attached to the chiral center in the α anomer, whereas it is on the same side of the ring as the CH_2OH group in the β anomer. Not all atoms in the ring lie in the same plane. Rather, pyranose rings usually adopt the chair conformation, and furanose rings usually adopt the envelope conformation. Sugars are joined to alcohols and amines by glycosidic bonds from the anomeric carbon atom. For example, *N*-glycosidic bonds link sugars to purines and pyrimidines in nucleotides, RNA, and DNA.

11.2 Monosaccharides Are Linked to Form Complex Carbohydrates

Sugars are linked to one another in disaccharides and polysaccharides by *O*-glycosidic bonds. Sucrose, lactose, and maltose are the common disaccharides. Sucrose (common table sugar) consists of α-glucose and β-fructose joined by a glycosidic linkage between their anomeric carbon atoms. Lactose (in milk) consists of galactose joined to glucose by a β-1,4 linkage. Maltose (in starch) consists of two glucoses joined by an α-1,4 linkage. Starch is a polymeric form of glucose in plants, and glycogen serves a similar role in animals. Most of the glucose units in starch and glycogen are in α-1,4 linkage. Cellulose, the major structural polymer of plant cell walls, consists of glucose units joined by β-1,4 linkages. These β linkages give rise to long straight chains that form fibrils with high tensile strength. In contrast, the α linkages in starch and glycogen lead to open helices, in keeping with their roles as mobilizable energy stores.

11.3 Carbohydrates Can Be Linked to Proteins to Form Glycoproteins

Carbohydrates are commonly conjugated to proteins. If the protein component is predominant, the conjugate of protein and carbohydrate is called a glycoprotein. Most secreted proteins are glycoproteins. The signal molecule erythropoietin is a glycoprotein. Glycoproteins are also prominent on the external surface of the plasma membrane. Proteins bearing covalently linked glycosaminoglycans are proteoglycans. Glycosaminoglycans are polymers of repeating disaccharides. One of the units in each repeat is a derivative of glucosamine or galactosamine. These highly anionic carbohydrates have a high density of carboxylate or sulfate groups. Proteoglycans are found in the extracellular matrices of animals and are key components of cartilage. Mucoproteins, like proteoglycans, are predominantly carbohydrate by weight. The protein component is heavily *O*-glycosylated with *N*-acetylgalactosamine joining the oligosaccharide to the protein. Mucoproteins serve as lubricants.

Specific enzymes link the oligosaccharide units on proteins either to the side-chain oxygen atom of a serine or threonine residue or to the side-chain amide nitrogen atom of an asparagine residue. Protein glycosylation takes place in the lumen of the endoplasmic reticulum. The *N*-linked oligosaccharides are synthesized on dolichol phosphate and subsequently transferred to the protein acceptor. Additional sugars are attached in the Golgi complex to form diverse patterns.

11.4 Lectins Are Specific Carbohydrate-Binding Proteins

Carbohydrates on cell surfaces are recognized by proteins called lectins. In animals, the interplay of lectins and their sugar targets guides cell–cell contact. The viral protein hemagglutinin on the surface of the influenza virus recognizes sialic acid residues on the surfaces of cells invaded by the virus. A small number of carbohydrate residues can be joined in many different ways to form highly diverse patterns that can be distinguished by the lectin domains of protein receptors.

Key Terms

glycobiology (p. 320)
glycomics (p. 320)
monosaccharide (p. 320)
ketose (p. 320)
aldose (p. 320)
triose (p. 320)
tetrose (p. 320)
pentose (p. 320)
hexose (p. 320)
heptose (p. 320)
constitutional isomer (p. 321)
stereoisomer (p. 321)
enantiomer (p. 321)
diastereoisomer (p. 321)
epimer (p. 322)
hemiacetal (p. 322)
pyranose (p. 322)
hemiketal (p. 322)
furanose (p. 323)
anomer (p. 323)
reducing sugar (p. 325)
nonreducing sugar (p. 325)
advanced glycosylation product (AGE) (p. 326)
glycosidic bond (p. 326)
oligosaccharide (p. 327)
disaccharide (p. 327)
polysaccharide (p. 328)
glycogen (p. 328)
starch (p. 328)
cellulose (p. 328)
glycoprotein (p. 329)
proteoglycan (p. 329)
glycosaminoglycan (p. 329)
mucin (mucoprotein) (p. 330)
glycoform (p. 330)
endoplasmic reticulum (p. 333)
Golgi complex (p. 333)
dolichol phosphate (p. 334)
glycosyltransferase (p. 335)
glycan-binding protein (p. 337)
lectin (p. 338)
selectin (p. 339)

Problems

1. *Word origin.* Account for the origin of the term *carbohydrate.*

2. *Diversity.* How many different oligosaccharides can be made by linking one glucose, one mannose, and one galactose? Assume that each sugar is in its pyranose form. Compare this number with the number of tripeptides that can be made from three different amino acids.

3. *Couples.* Indicate whether each of the following pairs of sugars consists of anomers, epimers, or an aldose–ketose pair:

(a) D-glyceraldehyde and dihydroxyacetone

(b) D-glucose and D-mannose

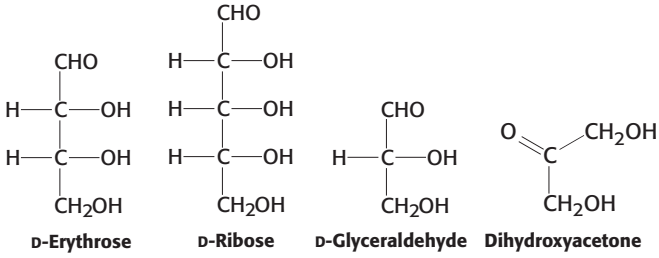
(c) D-glucose and D-fructose

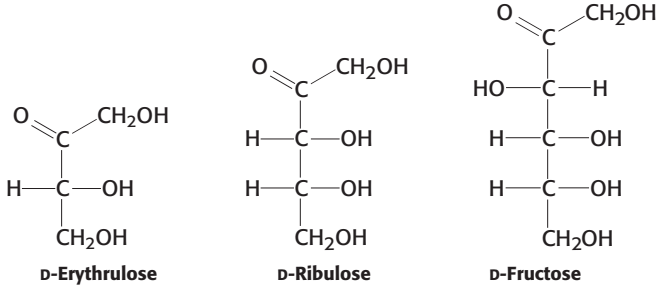
(d) α-D-glucose and β-D-glucose

(e) D-ribose and D-ribulose

(f) D-galactose and D-glucose

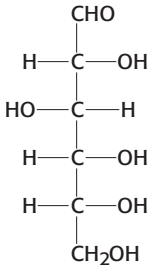
4. *Carbons and carbonyls.* To which classes of sugars do the monosaccharides shown here belong?





5. *Chemical cousins.* Although an aldose with 4 asymmetric carbon atoms is capable of forming 16 diastereoisomers, only 8 of the isomers are commonly observed, including glucose. They are listed below with their structural relation to glucose. Using the structure of glucose as a reference, draw the structures.

D-Allose: Epimeric at C-3
D-Altrose: Isomeric at C-2 and C-3
D-Mannose: Epimeric at C-2
D-Glucose: Isomeric at C-3 and C-4
D-Idose: Isomeric at C-2, C-3 and C-4
D-Galactose: Epimeric at C-4
D-Talose: Isomeric at C-2 and C-4



D-Glucose

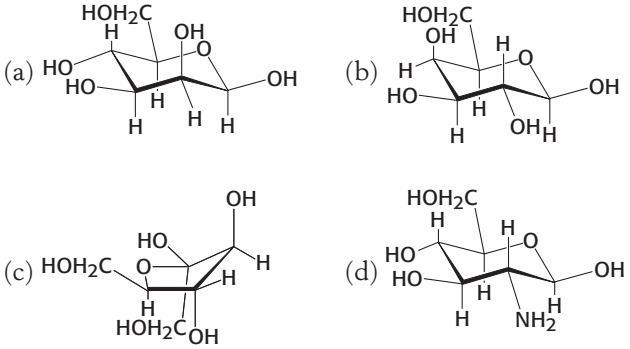
6. *Mutarotation.* The specific rotations of the α and β anomers of D-glucose are +112 degrees and +18.7 degrees, respectively. Specific rotation, $[\alpha]_D$, is defined as the observed rotation of light of wavelength 589 nm (the D line of a sodium lamp) passing through 10 cm of a 1 g ml^{-} solution of a sample. When a crystalline sample of α-D-glucopyranose is dissolved in water, the specific rotation decreases from 112 degrees to an equilibrium value of 52.7 degrees. On the basis of this result, what are the proportions of the α and β anomers at equilibrium? Assume that the concentration of the open-chain form is negligible.

7. *Telltale marker.* Glucose reacts slowly with hemoglobin and other proteins to form covalent compounds. Why is glucose reactive? What is the nature of the adduct formed?

8. *Periodate cleavage.* Compounds containing hydroxyl groups on adjacent carbon atoms undergo carbon–carbon bond cleavage when treated with periodate ion (IO_4^-). How can this reaction be used to distinguish between pyranosides and furanosides?

9. *Oxygen source.* Does the oxygen atom attached to C-1 in methyl α-D-glucopyranoside come from glucose or methanol?

10. *Sugar lineup.* Identify the following four sugars.



11. *Cellular glue.* A trisaccharide unit of a cell-surface glycoprotein is postulated to play a critical role in mediating cell–cell adhesion in a particular tissue. Design a simple experiment to test this hypothesis.

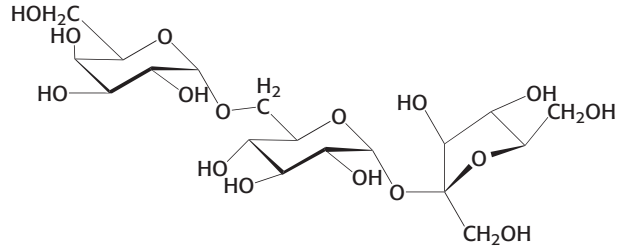
12. *Mapping the molecule.* Each of the hydroxyl groups of glucose can be methylated with reagents such as dimethylsulfate under basic conditions. Explain how exhaustive methylation followed by the complete digestion of a known amount of glycogen would enable you to determine the number of branch points and reducing ends.

13. *Component parts.* Raffinose is a trisaccharide and a minor constituent in sugar beets.

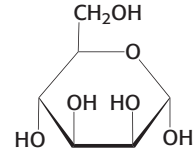
(a) Is raffinose a reducing sugar? Explain.

(b) What are the monosaccharides that compose raffinose?

(c) β-Galactosidase is an enzyme that will remove galactose residues from an oligosaccharide. What are the products of β-galactosidase treatment of raffinose?


Raffinose

14. *Anomeric differences.* α-D-Mannose is a sweet-tasting sugar. β-D-Mannose, on the other hand, tastes bitter. A pure solution of α-D-mannose loses its sweet taste with time as it is converted into the β anomer. Draw the β anomer and explain how it is formed from the α anomer.


α-D-Mannose

15. *A taste of honey.* Fructose in its β-D-pyranose form accounts for the powerful sweetness of honey. The β-D-furanose form, although sweet, is not as sweet as the pyranose form. The furanose form is the more stable form. Draw the two forms and explain why it may not always be wise to cook with honey.

16. *Making ends meet.* (a) Compare the number of reducing ends to nonreducing ends in a molecule of glycogen. (b) As we will see in Chapter 21, glycogen is an important fuel-storage form that is rapidly mobilized. At which end—the reducing or nonreducing—would you expect most metabolism to take place?

17. *A lost property.* Glucose and fructose are reducing sugars. Sucrose, or table sugar, is a disaccharide consisting of both fructose and glucose. Is sucrose a reducing sugar? Explain.

18. *Meat and potatoes*. Compare the structures of glycogen and starch.

19. *Straight or with a twist?* Account for the different structures of glycogen and cellulose.

20. *Sweet proteins*. List the key classes of glycoprotein, their defining characteristics, and their biological functions.

21. *Life extender*. What is the function of the carbohydrate moiety that is attached to EPO?

22. *Cushioning*. What is the role of the glycosaminoglycan in the cushioning provided by cartilage?

23. *Undelivered mail. Not returned to sender*. I-cell disease results when proteins normally destined to the lysosomes lack the appropriate carbohydrate-addressing molecule (p. 337). Suggest another possible means by which I-cell disease might arise.

24. *From one, many*. What is meant by a glycoform?

25. *Ome*. What is meant by the glycome?

26. *Exponential expansion?* Compare the amount of information inherent in the genome, the proteome, and the glycome.

27. *Locks and keys*. What does the fact that all organisms contain lectins suggest about the role of carbohydrates?

28. *Carbohydrates—not just for breakfast anymore*. Differentiate between a glycoprotein and a lectin.

29. *Carbohydrates and proteomics*. Suppose that a protein contains six potential *N*-linked glycosylation sites. How many possible proteins can be generated, depending on which of these sites is actually glycosylated? Do not include the effects of diversity within the carbohydrate added.

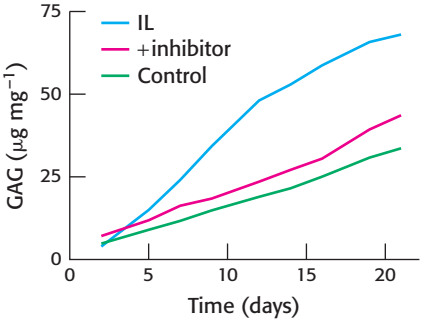
Chapter Integration Problems

30. *Stereospecificity*. Sucrose, a major product of photosynthesis in green leaves, is synthesized by a battery of enzymes. The substrates for sucrose synthesis, D-glucose and D-fructose, are a mixture of α and β anomers as well as acyclic compounds in solution. Nonetheless, sucrose consists of α-D-glucose linked by its carbon-1 atom to the carbon-2 atom of β-D-fructose. How can the specificity of sucrose be explained in light of the potential substrates?

31. *Specific recognition*. How might the technique of affinity chromatography be used to purify lectins?

Data Interpretation Problem

32. *Sore joints*. A contributing factor to the development of arthritis is the inappropriate proteolytic destruction of the aggrecan component of cartilage by the proteolytic enzyme aggrecanase. The immune-system signal molecule interleukin 2 (IL-2) activates aggrecanase; in fact, IL-2 blockers are sometimes used to treat arthritis. Studies were undertaken to determine whether inhibitors of aggrecanase can counteract the effects of IL-2. Pieces of cartilage were incubated in media with various additions and the amount of aggrecan destruction was measured as a function of time.



[After M. A. Pratta et al. *J. Biol. Chem.* 278:45539–45545, 2003, Fig. 7B.]

(a) Aggrecan degradation was measured by the release of glycosaminoglycan. What is the rational for this assay?

(b) Why might glycosaminoglycan release not indicate aggrecan degradation?

(c) What is the purpose of the control—cartilage incubated with no additions?

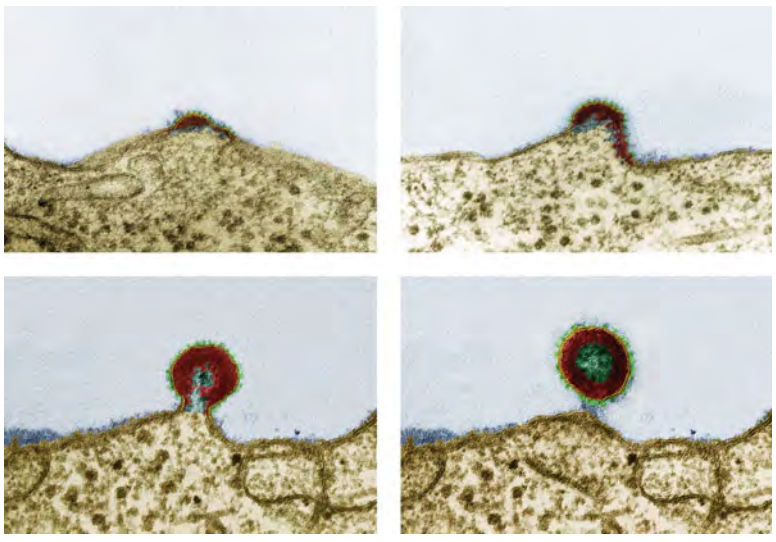
(d) What is the effect of adding IL-2 to the system?

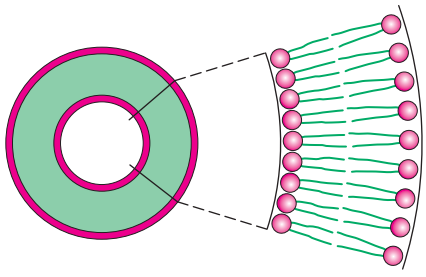
(e) What is the response when an aggrecanase inhibitor (ST154) is added in addition to IL-2?

(f) Why is there some aggrecan destruction in the control with the passage of time?

CHAPTER 12

Lipids and Cell Membranes





An HIV particle exits an infected cell by membrane budding. Cellular membranes are highly dynamic structures that spontaneously self-assemble. Driven by hydrophobic interactions, as shown in the diagram at right the fatty acid tails of membrane lipids pack together (green), while the polar heads (red) remain exposed on the surfaces. [Micrographs from Eye of Science/Photo Researchers.]

The boundaries of all cells are defined by *biological membranes* (Figure 12.1). These barriers prevent molecules generated inside the cell from leaking out and unwanted molecules from diffusing in; yet they also contain transport systems that allow the cell to take up specific molecules and remove unwanted ones. Such transport systems confer on membranes the important property of *selective permeability*.

Membranes are dynamic structures in which proteins float in a sea of lipids. The lipid components of the membrane form the barrier to permeability, and protein components act as a transport system of pumps and channels that allow selected molecules into and out of the cell. This transport system will be considered in the next chapter.

In addition to an external cell membrane (called the *plasma membrane*), eukaryotic cells also contain internal membranes that form the boundaries of organelles such as mitochondria, chloroplasts, peroxisomes, and lysosomes. Functional specialization in the course of evolution has been closely linked to the formation of such compartments. Specific systems have evolved to allow the targeting of selected proteins into or through particular internal membranes and, hence, into specific organelles. External and internal membranes share essential properties; these features are the subject of this chapter.

Biological membranes serve several additional functions indispensable for life, such as energy storage and information transduction, that are dictated by the proteins associated with them. In this chapter, we will examine

OUTLINE

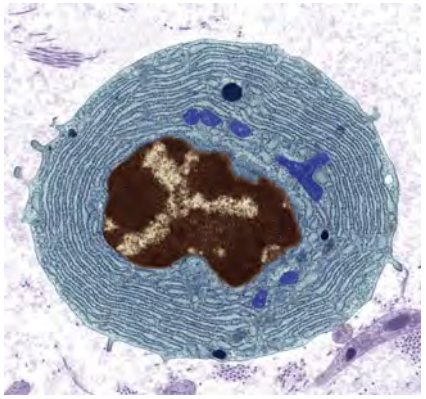


Figure 12.1 Electron micrograph of a plasma cell. This image has been colored to indicate the distinct boundary of the cell, formed by its plasma membrane. [Steve Gschmeissner/Photo Researchers.]

the properties of membrane proteins that enable them to exist in the hydrophobic environment of the membrane while connecting two hydrophilic environments, and defer a discussion of the functions of these proteins until later chapters.

Many Common Features Underlie the Diversity of Biological Membranes

Membranes are as diverse in structure as they are in function. However, they do have in common a number of important attributes:

1. Membranes are *sheetlike structures,* only two molecules thick, that form *closed boundaries* between different compartments. The thickness of most membranes is between 60 Å (6 nm) and 100 Å (10 nm).

2. Membranes consist mainly of *lipids* and *proteins*. The mass ratio of lipids to proteins ranges from 1:4 to 4:1. Membranes also contain *carbohydrates* that are linked to lipids and proteins.

3. Membrane lipids are small molecules that have both *hydrophilic* and *hydrophobic* moieties. These lipids spontaneously form *closed bimolecular sheets* in aqueous media. These *lipid bilayers* are barriers to the flow of polar molecules.

4. *Specific proteins mediate distinctive functions of membranes*. Proteins serve as pumps, channels, receptors, energy transducers, and enzymes. Membrane proteins are embedded in lipid bilayers, which create suitable environments for their action.

5. Membranes are *noncovalent assemblies*. The constituent protein and lipid molecules are held together by many noncovalent interactions, which act cooperatively.

6. Membranes are *asymmetric*. The two faces of biological membranes always differ from each other.

7. Membranes are *fluid structures*. Lipid molecules diffuse rapidly in the plane of the membrane, as do proteins, unless they are anchored by specific interactions. In contrast, lipid molecules and proteins do not readily rotate across the membrane. Membranes can be regarded as *two-dimensional solutions of oriented proteins and lipids*.

8. Most cell membranes are *electrically polarized,* such that the inside is negative [typically −60 millivolts (mV)]. Membrane potential plays a key role in transport, energy conversion, and excitability (Chapter 13).

12.1 Fatty Acids Are Key Constituents of Lipids

The hydrophobic properties of lipids are essential to their ability to form membranes. Most lipids owe their hydrophobic properties to one component, their fatty acids.

Fatty acid names are based on their parent hydrocarbons

Fatty acids are long hydrocarbon chains of various lengths and degrees of unsaturation terminated with carboxylic acid groups. The systematic name for a fatty acid is derived from the name of its parent hydrocarbon by the substitution of *oic* for the final *e.* For example, the C_{18} saturated fatty acid is called *octadecanoic acid* because the parent hydrocarbon is octadecane. A C_{18} fatty acid with one double bond is called octadec*enoic* acid; with two

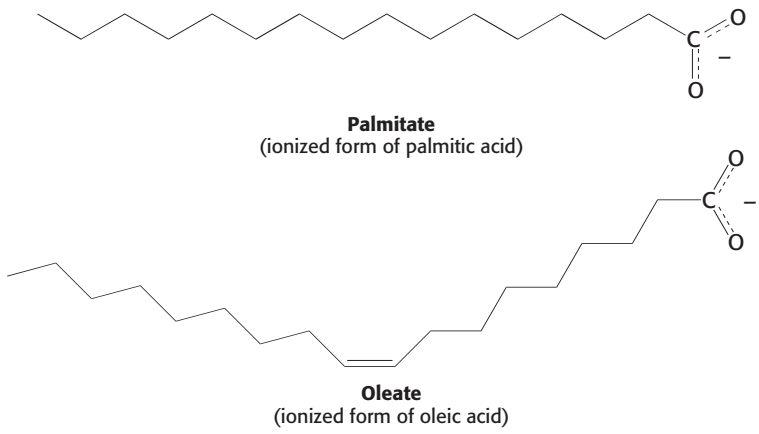
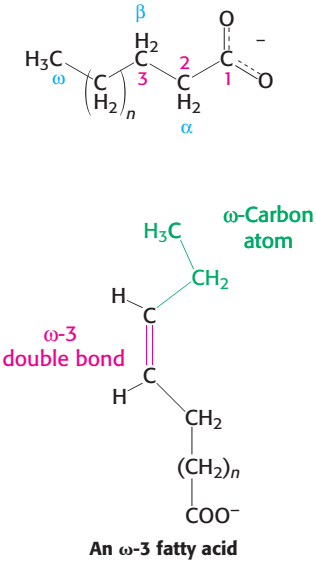


Figure 12.2 Structures of two fatty acids. Palmitate is a 16-carbon, saturated fatty acid, and oleate is an 18-carbon fatty acid with a single cis double bond.

double bonds, octadeca*dienoic* acid; and with three double bonds, octadeca*trienoic* acid. The notation 18:0 denotes a C_{18} fatty acid with no double bonds, whereas 18:2 signifies that there are two double bonds. The structures of the ionized forms of two common fatty acids—palmitic acid (16:0) and oleic acid (18:1)—are shown in Figure 12.2.

Fatty acid carbon atoms are numbered starting at the carboxyl terminus, as shown in the margin. Carbon atoms 2 and 3 are often referred to as α and β, respectively. The methyl carbon atom at the distal end of the chain is called the *ω-carbon atom*. The position of a double bond is represented by the symbol Δ followed by a superscript number. For example, *cis*-Δ^9 means that there is a cis double bond between carbon atoms 9 and 10; *trans*-Δ^2 means that there is a trans double bond between carbon atoms 2 and 3. Alternatively, the position of a double bond can be denoted by counting from the distal end, with the ω-carbon atom (the methyl carbon) as number 1. An ω-3 fatty acid, for example, has the structure shown in the margin. Fatty acids are ionized at physiological pH, and so it is appropriate to refer to them according to their carboxylate form: for example, palmitate or hexadecanoate.

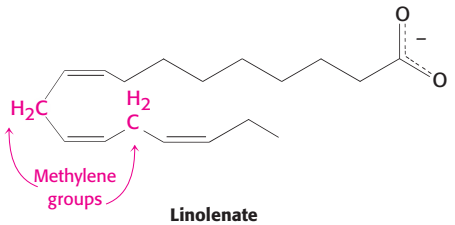


Fatty acids vary in chain length and degree of unsaturation

Fatty acids in biological systems usually contain an even number of carbon atoms, typically between 14 and 24 (Table 12.1). The 16- and 18-carbon fatty acids are most common. The dominance of fatty acid chains containing an even number of carbon atoms is in accord with the way in which fatty

Table 12.1 Some naturally occurring fatty acids in animals

Number of carbons	Number of double bonds	Common name	Systematic name	Formula
12	0	Laurate	*n*-Dodecanoate	$CH_3(CH_2)_{10}COO^-$
14	0	Myristate	*n*-Tetradecanoate	$CH_3(CH_2)_{12}COO^-$
16	0	Palmitate	*n*-Hexadecanoate	$CH_3(CH_2)_{14}COO^-$
18	0	Stearate	*n*-Octadecanoate	$CH_3(CH_2)_{16}COO^-$
20	0	Arachidate	*n*-Eicosanoate	$CH_3(CH_2)_{18}COO^-$
22	0	Behenate	*n*-Docosanoate	$CH_3(CH_2)_{20}COO^-$
24	0	Lignocerate	*n*-Tetracosanoate	$CH_3(CH_2)_{22}COO^-$
16	1	Palmitoleate	*cis*-Δ^9-Hexadecenoate	$CH_3(CH_2)_5CH{=}CH(CH_2)_7COO^-$
18	1	Oleate	*cis*-Δ^9-Octadecenoate	$CH_3(CH_2)_7CH{=}CH(CH_2)_7COO^-$
18	2	Linoleate	*cis,cis*-Δ^9, Δ^{12}-Octadecadienoate	$CH_3(CH_2)_4(CH{=}CHCH_2)_2(CH)_6COO^-$
18	3	Linolenate	all-*cis*-Δ^9, Δ^{12}, Δ^{15}-Octadecatrienoate	$CH_3CH_2(CH{=}CHCH_2)_3(CH_2)_6COO^-$
20	4	Arachidonate	all-*cis* Δ^5, Δ^8, Δ^{11}, -Δ^{14} Eicosatetraenoate	$CH_3(CH_2)_4(CH{=}CHCH_2)_4(CH_2)_2COO^-$



Linolenate

acids are biosynthesized (Chapter 26). The hydrocarbon chain is almost invariably unbranched in animal fatty acids. The alkyl chain may be saturated or it may contain one or more double bonds. The configuration of the double bonds in most unsaturated fatty acids is cis. The double bonds in polyunsaturated fatty acids are separated by at least one methylene group.

The properties of fatty acids and of lipids derived from them are markedly dependent on chain length and degree of saturation. Unsaturated fatty acids have lower melting points than do saturated fatty acids of the same length. For example, the melting point of stearic acid is 69.6°C, whereas that of oleic acid (which contains one cis double bond) is 13.4°C. The melting points of polyunsaturated fatty acids of the C_{18} series are even lower. Chain length also affects the melting point, as illustrated by the fact that the melting temperature of palmitic acid (C_{16}) is 6.5 degrees lower than that of stearic acid (C_{18}). Thus, *short chain length and unsaturation enhance the fluidity of fatty acids and of their derivatives.*

12.2 There Are Three Common Types of Membrane Lipids

By definition, *lipids are water-insoluble biomolecules that are highly soluble in organic solvents such as chloroform.* Lipids have a variety of biological roles: they serve as fuel molecules, highly concentrated energy stores, signal molecules and messengers in signal-transduction pathways, and components of membranes. The first three roles of lipids will be considered in later chapters. Here, our focus is on lipids as membrane constituents. The three major kinds of membrane lipids are *phospholipids, glycolipids,* and *cholesterol.* We begin with lipids found in eukaryotes and bacteria. The lipids in archaea are distinct, although they have many features related to membrane formation in common with lipids of other organisms.

Phospholipids are the major class of membrane lipids

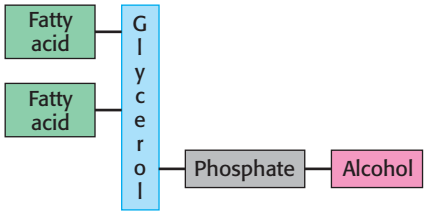
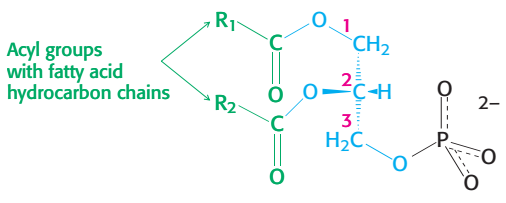


Figure 12.3 Schematic structure of a phospholipid.

Phospholipids are abundant in all biological membranes. A phospholipid molecule is constructed from four components: one or more fatty acids, a platform to which the fatty acids are attached, a phosphate, and an alcohol attached to the phosphate (Figure 12.3). The fatty acid components provide a hydrophobic barrier, whereas the remainder of the molecule has hydrophilic properties that enable interaction with the aqueous environment.

The platform on which phospholipids are built may be *glycerol,* a three-carbon alcohol, or *sphingosine,* a more complex alcohol. Phospholipids derived from glycerol are called *phosphoglycerides.* A phosphoglyceride consists of a glycerol backbone to which are attached two fatty acid chains and a phosphorylated alcohol.

In phosphoglycerides, the hydroxyl groups at C-1 and C-2 of glycerol are esterified to the carboxyl groups of the two fatty acid chains. The C-3 hydroxyl group of the glycerol backbone is esterified to phosphoric acid. When no further additions are made, the resulting compound is *phosphatidate (diacylglycerol 3-phosphate),* the simplest phosphoglyceride. Only small amounts of phosphatidate are present in membranes. However, the molecule is a key intermediate in the biosynthesis of the other phosphoglycerides (Section 26.1). The absolute configuration of the glycerol 3-phosphate moiety of membrane lipids is shown in Figure 12.4.

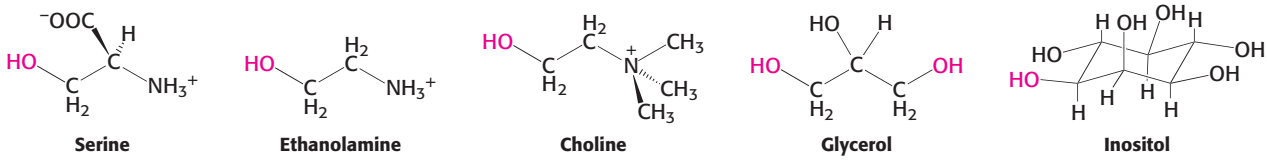


Phosphatidate
(Diacylglycerol 3-phosphate)

Figure 12.4 Structure of phosphatidate (diacylglycerol 3-phosphate). The absolute configuration of the center carbon (C-2) is shown.

The major phosphoglycerides are derived from phosphatidate by the formation of an ester bond between the phosphate group of phosphatidate and the hydroxyl group of one of several alcohols. The

common alcohol moieties of phosphoglycerides are the amino acid serine, ethanolamine, choline, glycerol, and inositol.



The structural formulas of phosphatidylcholine and the other principal phosphoglycerides—namely, phosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, and diphosphatidylglycerol—are given in Figure 12.5.

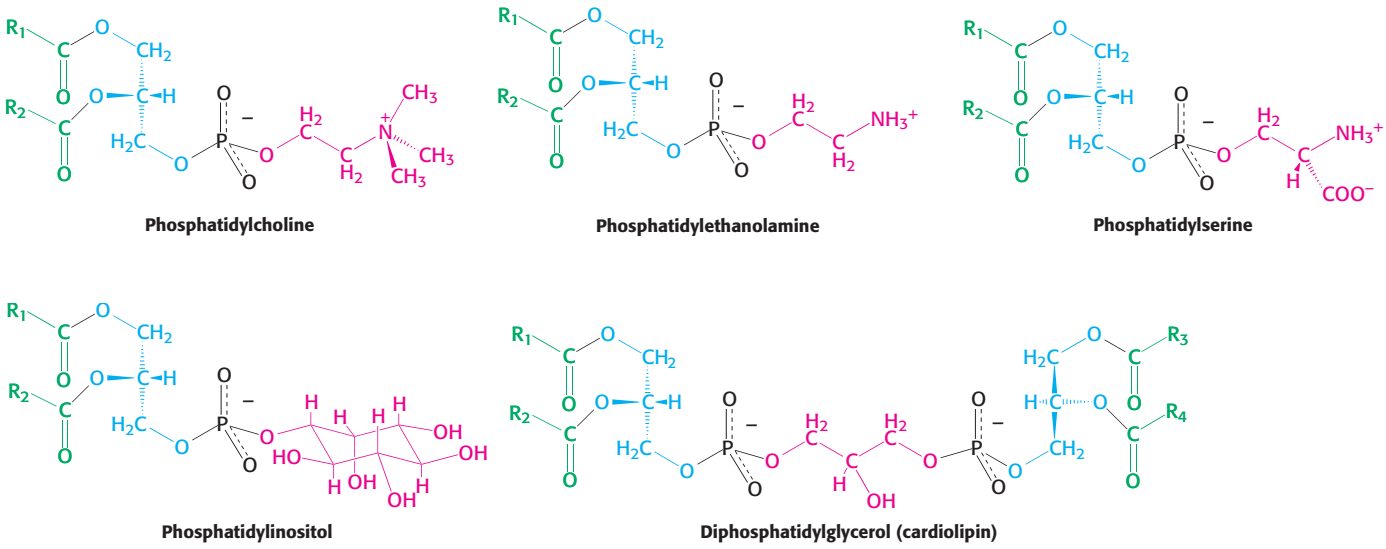


Figure 12.5 Some common phosphoglycerides found in membranes.

Sphingomyelin is a phospholipid found in membranes that is not derived from glycerol. Instead, the backbone in sphingomyelin is *sphingosine,* an amino alcohol that contains a long, unsaturated hydrocarbon chain (Figure 12.6). In sphingomyelin, the amino group of the sphingosine backbone is linked to a fatty acid by an amide bond. In addition, the primary hydroxyl group of sphingosine is esterified to phosphorylcholine.

Membrane lipids can include carbohydrate moieties

The second major class of membrane lipids, *glycolipids,* are *sugar-containing lipids.* Like sphingomyelin, the glycolipids in animal cells are derived from sphingosine. The amino group of the sphingosine backbone is acylated by a fatty acid, as in sphingomyelin. Glycolipids differ

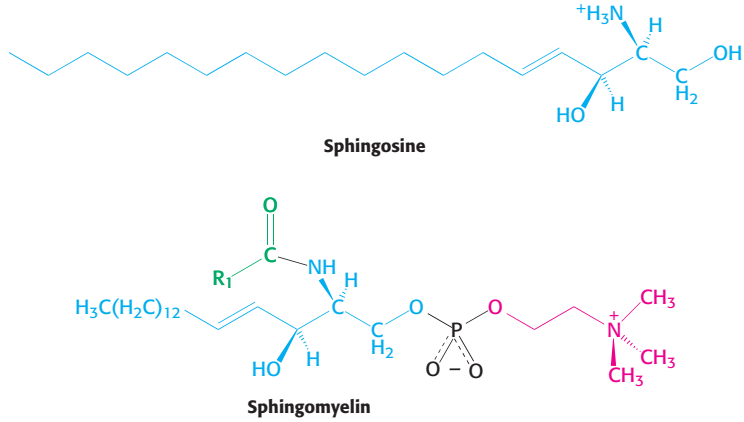
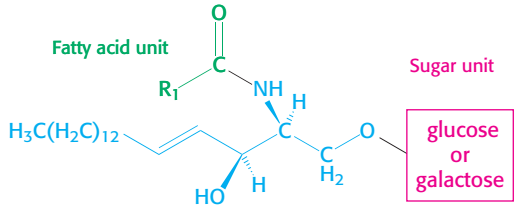


Figure 12.6 Structures of sphingosine and sphingomyelin. The sphingosine moiety of sphingomyelin is highlighted in blue.

from sphingomyelin in the identity of the unit that is linked to the primary hydroxyl group of the sphingosine backbone. In glycolipids, one or more sugars (rather than phosphorylcholine) are attached to this group. The simplest glycolipid, called a *cerebroside,* contains a single sugar residue, either glucose or galactose.

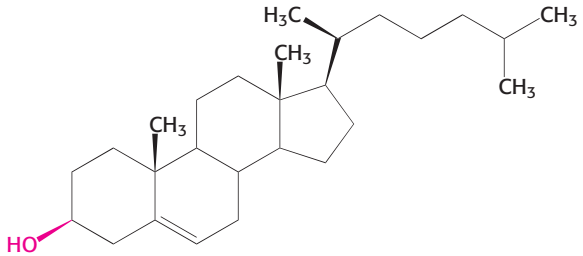


Cerebroside
(a glycolipid)

More-complex glycolipids, such as *gangliosides,* contain a branched chain of as many as seven sugar residues. Glycolipids are oriented in a completely asymmetric fashion with the *sugar residues always on the extracellular side of the membrane.*

Cholesterol Is a Lipid Based on a Steroid Nucleus

Cholesterol, the third major type of membrane lipid, has a structure that is quite different from that of phospholipids. It is a steroid, built from four linked hydrocarbon rings.



Cholesterol

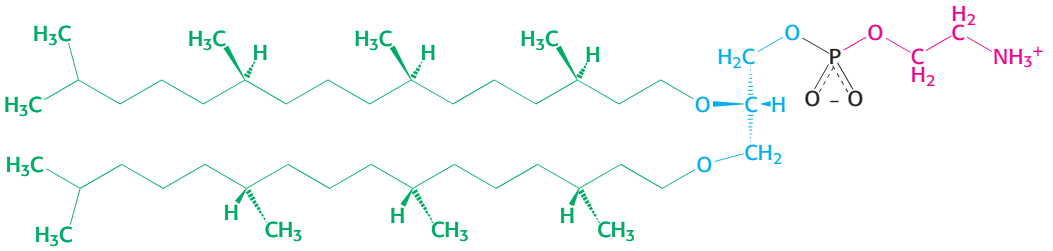
A hydrocarbon tail is linked to the steroid at one end, and a hydroxyl group is attached at the other end. In membranes, the orientation of the molecule is parallel to the fatty acid chains of the phospholipids, and the hydroxyl group interacts with the nearby phospholipid head groups. Cholesterol is absent from prokaryotes but is found to varying degrees in virtually all animal membranes. It constitutes almost 25% of the membrane lipids in certain nerve cells but is essentially absent from some intracellular membranes.

Figure 12.7 An archaeon and its environment. Archaea can thrive in habitats as harsh as a volcanic vent. Here, the archaea form an orange mat surrounded by yellow sulfurous deposits. [Krafft-Explorer/Photo Researchers.]

Archaeal membranes are built from ether lipids with branched chains

The membranes of archaea differ in composition from those of eukaryotes or bacteria in three important ways. Two of these differences clearly relate to the hostile living conditions of many archaea (Figure 12.7). First, the nonpolar chains are joined to a glycerol backbone by ether rather than ester linkages. The ether linkage is more resistant to hydrolysis. Second, the alkyl chains are branched rather than linear. They are built up from repeats of a fully saturated five-carbon fragment. These branched, saturated hydrocarbons are more resistant to oxidation. The ability of archaeal lipids to resist hydrolysis and oxidation may help these organisms to withstand the extreme conditions, such as high temperature,

low pH, or high salt concentration, under which some of these archaea grow. Finally, the stereochemistry of the central glycerol is inverted compared with that shown in Figure 12.4.



Membrane lipid from the archaeon *Methanococcus jannaschii*

A membrane lipid is an amphipathic molecule containing a hydrophilic and a hydrophobic moiety

The repertoire of membrane lipids is extensive. However, these lipids possess a critical common structural theme: *membrane lipids are amphipathic molecules* (amphiphilic molecules). A membrane lipid contains both a *hydrophilic* and a *hydrophobic* moiety.

Let us look at a model of a phosphoglyceride, such as phosphatidylcholine. Its overall shape is roughly rectangular (Figure 12.8A). The two hydrophobic fatty acid chains are approximately parallel to each other, whereas the hydrophilic phosphorylcholine moiety points in the opposite direction. Sphingomyelin has a similar conformation, as does the archaeal lipid depicted. Therefore, the following shorthand has been adopted to represent these membrane lipids: the hydrophilic unit, also called the *polar head group,* is represented by a circle, and the hydrocarbon tails are depicted by straight or wavy lines (Figure 12.8B).

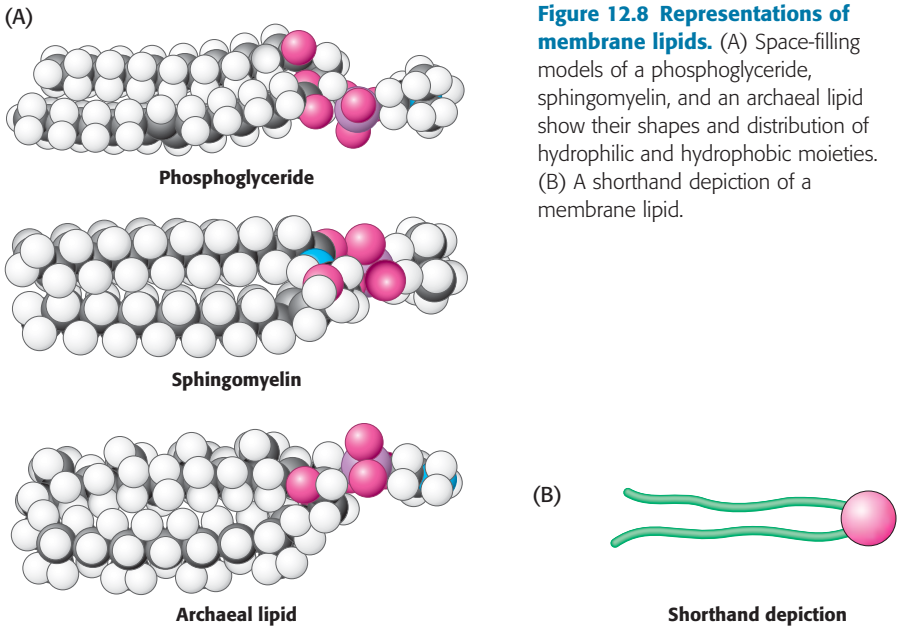


Figure 12.8 Representations of membrane lipids. (A) Space-filling models of a phosphoglyceride, sphingomyelin, and an archaeal lipid show their shapes and distribution of hydrophilic and hydrophobic moieties. (B) A shorthand depiction of a membrane lipid.

12.3 Phospholipids and Glycolipids Readily Form Bimolecular Sheets in Aqueous Media

What properties enable phospholipids to form membranes? *Membrane formation is a consequence of the amphipathic nature of the molecules.* Their polar head groups favor contact with water, whereas their hydrocarbon tails interact with one another in preference to water. How can molecules with these preferences arrange themselves in aqueous solutions? One way is to form a globular structure called a *micelle*. The polar head groups form the outside surface of the micelle, which is surrounded by water, and the hydrocarbon tails are sequestered inside, interacting with one another (Figure 12.9).

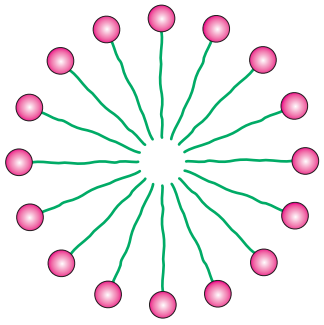


Figure 12.9 Diagram of a section of a micelle. Ionized fatty acids readily form such structures, but most phospholipids do not.

Alternatively, the strongly opposed preferences of the hydrophilic and hydrophobic moieties of membrane lipids can be satisfied by forming a *lipid bilayer,* composed of two lipid sheets (Figure 12.10). A lipid bilayer is also called a *bimolecular sheet*. The hydrophobic tails of each individual sheet interact with one another, forming a hydrophobic interior that acts as a permeability barrier. The hydrophilic head groups interact with the aqueous medium on each side of the bilayer. The two opposing sheets are called leaflets.

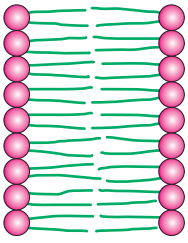


Figure 12.10 Diagram of a section of a bilayer membrane.

The favored structure for most phospholipids and glycolipids in aqueous media is a bimolecular sheet rather than a micelle. The reason is that the two fatty acid chains of a phospholipid or a glycolipid are too bulky to fit into the interior of a micelle. In contrast, salts of fatty acids (such as sodium palmitate, a constituent of soap) readily form micelles because they contain only one chain. *The formation of bilayers instead of micelles by phospholipids is of critical biological importance.* A micelle is a limited structure, usually less than 200 Å (20 nm) in diameter. In contrast, a bimolecular sheet can extend to macroscopic dimensions, as much as a millimeter (10^7 Å, or 10^6 nm) or more. Phospholipids and related molecules are important membrane constituents because they readily form extensive bimolecular sheets (Figure 12.11).

Lipid bilayers form spontaneously by a *self-assembly process.* In other words, the structure of a bimolecular sheet is inherent in the structure of the constituent lipid molecules. The growth of lipid bilayers from phospholipids is rapid and spontaneous in water. *Hydrophobic interactions are the major driving force for the formation of lipid bilayers.* Recall that hydrophobic

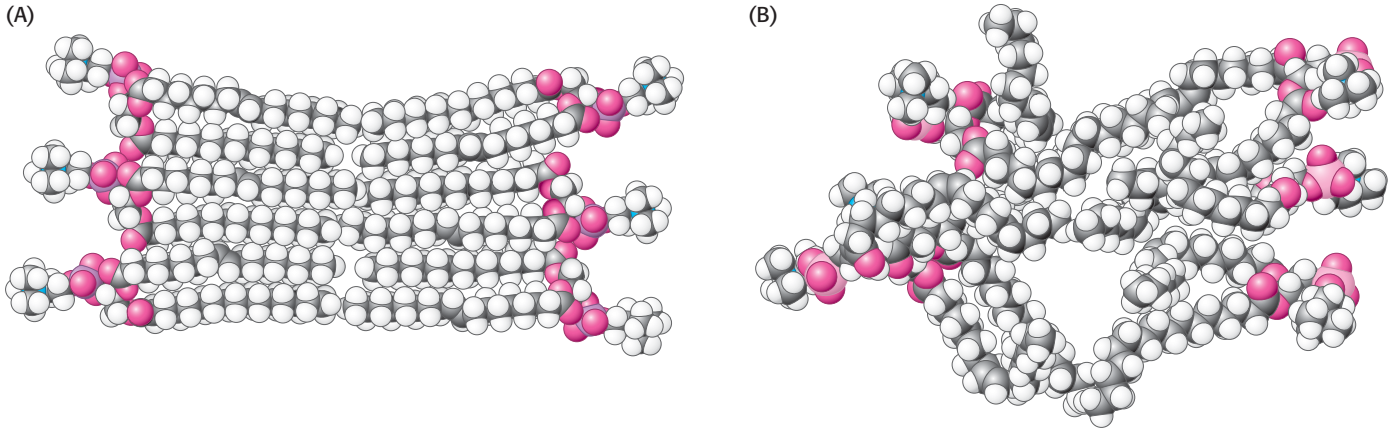


Figure 12.11 Space-filling model of a section of phospholipid bilayer membrane. (A) An idealized view showing regular structures. (B) A more realistic view of a fluid bilayer showing more irregular structures of the fatty acid chains.

interactions also play a dominant role in the stacking of bases in nucleic acids and in the folding of proteins (Sections 1.3 and 2.4). Water molecules are released from the hydrocarbon tails of membrane lipids as these tails become sequestered in the nonpolar interior of the bilayer. Furthermore, *van der Waals attractive forces between the hydrocarbon tails favor close packing of the tails*. Finally, there are *electrostatic and hydrogen-bonding attractions between the polar head groups and water molecules*. Thus, lipid bilayers are stabilized by the full array of forces that mediate molecular interactions in biological systems. Because lipid bilayers are held together by many *reinforcing, noncovalent interactions (predominantly hydrophobic)*, they are *cooperative structures*. These hydrophobic interactions have three significant biological consequences: (1) lipid bilayers have an inherent tendency to be *extensive;* (2) lipid bilayers will tend to *close on themselves* so that there are no edges with exposed hydrocarbon chains, and so they form compartments; and (3) lipid bilayers are *self-sealing* because a hole in a bilayer is energetically unfavorable.

Lipid vesicles can be formed from phospholipids

The propensity of phospholipids to form membranes has been used to create an important experimental and clinical tool. *Lipid vesicles,* or *liposomes,* are aqueous compartments enclosed by a lipid bilayer (Figure 12.12). These structures can be used to study membrane permeability or to deliver chemicals to cells. Liposomes are formed by suspending a suitable lipid, such as phosphatidylcholine, in an aqueous medium, and then *sonicating* (i.e., agitating by high-frequency sound waves) to give a dispersion of closed vesicles that are quite uniform in size. Vesicles formed by this method are nearly spherical and have a diameter of about 500 Å (50 nm). Larger vesicles (of the order of 1 μm or 10^4 Å in diameter) can be prepared by slowly evaporating the organic solvent from a suspension of phospholipid in a mixed-solvent system.

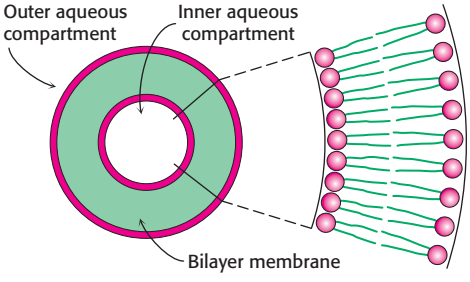


Figure 12.12 Liposome. A liposome, or lipid vesicle, is a small aqueous compartment surrounded by a lipid bilayer.

Ions or molecules can be trapped in the aqueous compartments of lipid vesicles by forming the vesicles in the presence of these substances (Figure 12.13). For example, 500-Å-diameter vesicles formed in a 0.1 M glycine solution will trap about 2000 molecules of glycine in each inner aqueous compartment. These glycine-containing vesicles can be separated from the surrounding solution of glycine by dialysis or by gel-filtration chromatography. The permeability of the bilayer membrane to glycine can then be determined by measuring the rate of efflux of glycine from the inner compartment of the vesicle to the ambient solution. Liposomes can be formed with specific membrane proteins embedded in them by solubilizing

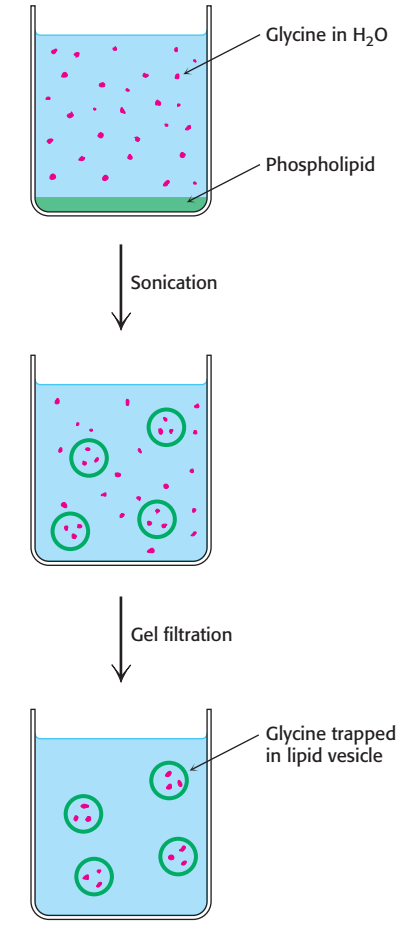


Figure 12.13 Preparation of glycine-containing liposomes. Liposomes containing glycine are formed by the sonication of phospholipids in the presence of glycine. Free glycine is removed by gel filtration.

the proteins in the presence of detergents and then adding them to the phospholipids from which liposomes will be formed. Protein–liposome complexes provide valuable experimental tools for examining a range of membrane-protein functions.

Therapeutic applications for liposomes are currently under active investigation. For example, liposomes containing drugs or DNA for gene-therapy experiments can be injected into patients. These liposomes fuse with the plasma membrane of many kinds of cells, introducing into the cells the molecules that they contain. Drug delivery with liposomes often lessens its toxicity. Less of the drug is distributed to normal tissues because long-circulating liposomes concentrate in regions of increased blood circulation, such as solid tumors and sites of inflammation. Moreover, the selective fusion of lipid vesicles with particular kinds of cells is a promising means of controlling the delivery of drugs to target cells.

Another well-defined synthetic membrane is a *planar bilayer membrane.* This structure can be formed across a 1-mm hole in a partition between two aqueous compartments by dipping a fine paintbrush into a membrane-forming solution, such as phosphatidylcholine in decane, and stroking the tip of the brush across the hole. The lipid film across the hole thins spontaneously into a lipid bilayer. The electrical conduction properties of this macroscopic bilayer membrane are readily studied by inserting electrodes into each aqueous compartment (Figure 12.14). For example, the permeability of the membrane to ions is determined by measuring the current across the membrane as a function of the applied voltage.

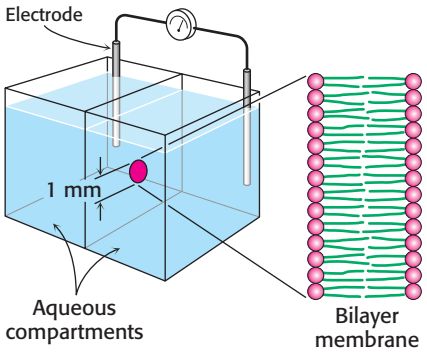


Figure 12.14 Experimental arrangement for the study of a planar bilayer membrane. A bilayer membrane is formed across a 1-mm hole in a septum that separates two aqueous compartments. This arrangement permits measurements of the permeability and electrical conductance of lipid bilayers.

Lipid bilayers are highly impermeable to ions and most polar molecules

Permeability studies of lipid vesicles and electrical-conductance measurements of planar bilayers have shown that *lipid bilayer membranes have a very low permeability for ions and most polar molecules.* Water is a conspicuous exception to this generalization; it traverses such membranes relatively easily because of its low molecular weight, high concentration, and lack of a complete charge. The range of measured permeability coefficients is very wide (Figure 12.15). For example, Na^+ and K^+ traverse these membranes 10^9 times as slowly as does H_2O. Tryptophan, a zwitterion at pH 7, crosses the membrane 10^3 times as slowly as does indole, a structurally related molecule that lacks ionic groups. In fact, *the permeability of small molecules is correlated with their solubility in a nonpolar solvent relative to their solubility in water.* This relation suggests that a small molecule might traverse a lipid bilayer membrane in the following way: first, it sheds its solvation shell of water; then, it is dissolved in the hydrocarbon core of the membrane; and, finally, it diffuses through this core to the other side of the membrane,

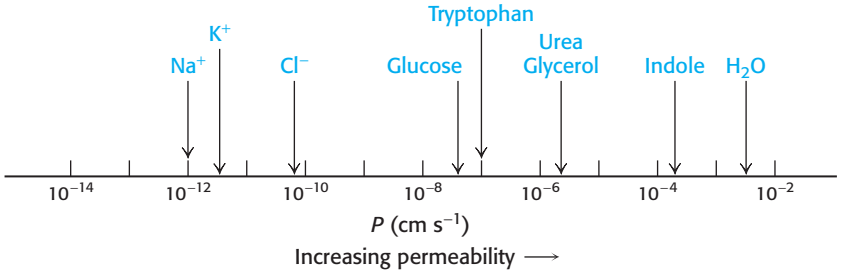


Figure 12.15 Permeability coefficients (*P*) of ions and molecules in a lipid bilayer. The ability of molecules to cross a lipid bilayer spans a wide range of values.

where it becomes resolvated by water. An ion such as Na^+ traverses membranes very slowly because the replacement of its coordination shell of polar water molecules by nonpolar interactions with the membrane interior is highly unfavorable energetically.

12.4 Proteins Carry Out Most Membrane Processes

We now turn to membrane proteins, which are responsible for most of the dynamic processes carried out by membranes. Membrane lipids form a permeability barrier and thereby establish compartments, whereas *specific proteins mediate nearly all other membrane functions.* In particular, proteins transport chemicals and information across a membrane. Membrane lipids create the appropriate environment for the action of such proteins.

Membranes differ in their protein content. Myelin, a membrane that serves as an electrical insulator around certain nerve fibers, has a low content of protein (18%). Relatively pure lipids are well suited for insulation. In contrast, the plasma membranes, or exterior membranes, of most other cells are much more metabolically active. They contain many pumps, channels, receptors, and enzymes. The protein content of these plasma membranes is typically 50%. Energy-transduction membranes, such as the internal membranes of mitochondria and chloroplasts, have the highest content of protein, typically 75%.

The protein components of a membrane can be readily visualized by *SDS–polyacrylamide gel electrophoresis.* As stated earlier (p. 71), the electrophoretic mobility of many proteins in SDS-containing gels depends on the mass rather than on the net charge of the protein. The gel-electrophoresis patterns of three membranes—the plasma membrane of erythrocytes, the photoreceptor membrane of retinal rod cells, and the sarcoplasmic reticulum membrane of muscle—are shown in Figure 12.16. It is evident that each of these three membranes contains many proteins but has a distinct protein composition. In general, *membranes performing different functions contain different repertoires of proteins.*

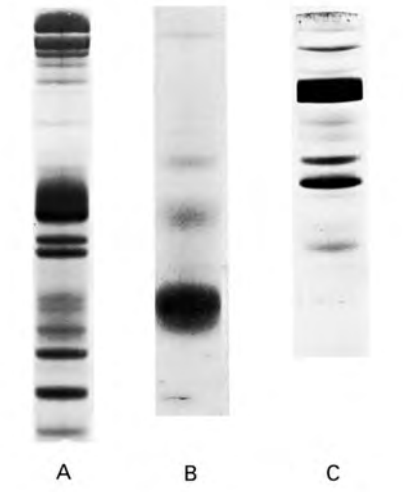


Figure 12.16 SDS–acrylamide gel patterns of membrane proteins.
(A) The plasma membrane of erythrocytes. (B) The photoreceptor membranes of retinal rod cells. (C) The sarcoplasmic reticulum membrane of muscle cells. [Courtesy of Dr. Theodore Steck (part A) and Dr. David MacLennan (part C).]

Proteins associate with the lipid bilayer in a variety of ways

The ease with which a protein can be dissociated from a membrane indicates how intimately it is associated with the membrane. Some membrane proteins can be solubilized by relatively mild means, such as extraction by a solution of high ionic strength (e.g., 1 M NaCl). Other membrane proteins are bound much more tenaciously; they can be solubilized only by using a detergent or an organic solvent. Membrane proteins can be classified as being either *peripheral* or *integral* on the basis of this difference in dissociability (Figure 12.17). *Integral membrane proteins* interact extensively with the hydrocarbon chains of membrane lipids, and they can be released only by agents that compete for these nonpolar interactions. In fact, most integral membrane proteins span the lipid bilayer. In contrast, *peripheral membrane proteins* are bound to membranes primarily by electrostatic and hydrogen-bond interactions with the head groups of lipids. These polar interactions can be disrupted by adding salts or by changing the pH. Many peripheral membrane proteins are bound to the surfaces of integral proteins, on

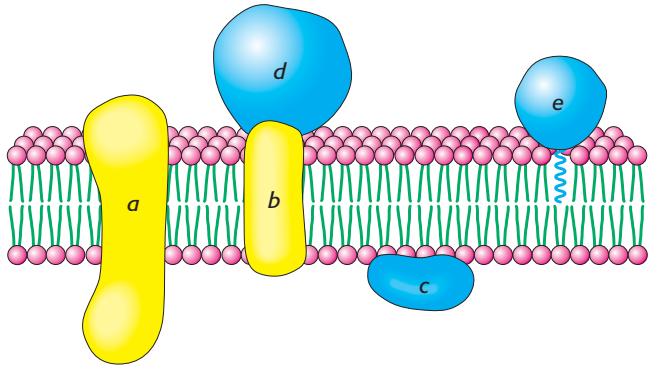


Figure 12.17 Integral and peripheral membrane proteins.
Integral membrane proteins (*a* and *b*) interact extensively with the hydrocarbon region of the bilayer. Most known integral membrane proteins traverse the lipid bilayer. Peripheral membrane proteins interact with the polar head groups of the lipids (*c*) or bind to the surfaces of integral proteins (*d*). Other proteins are tightly anchored to the membrane by a covalently attached lipid molecule (*e*).

either the cytoplasmic or the extracellular side of the membrane. Others are anchored to the lipid bilayer by a covalently attached hydrophobic chain, such as a fatty acid.

Proteins interact with membranes in a variety of ways

Membrane proteins are more difficult to purify and crystallize than are water-soluble proteins. Nonetheless, researchers using x-ray crystallographic or electron microscopic methods have determined the three-dimensional structures of more than 200 such proteins at sufficiently high resolution to discern the molecular details. As noted in Chapter 2, membrane proteins differ from soluble proteins in the distribution of hydrophobic and hydrophilic groups. We will consider the structures of three membrane proteins in some detail.

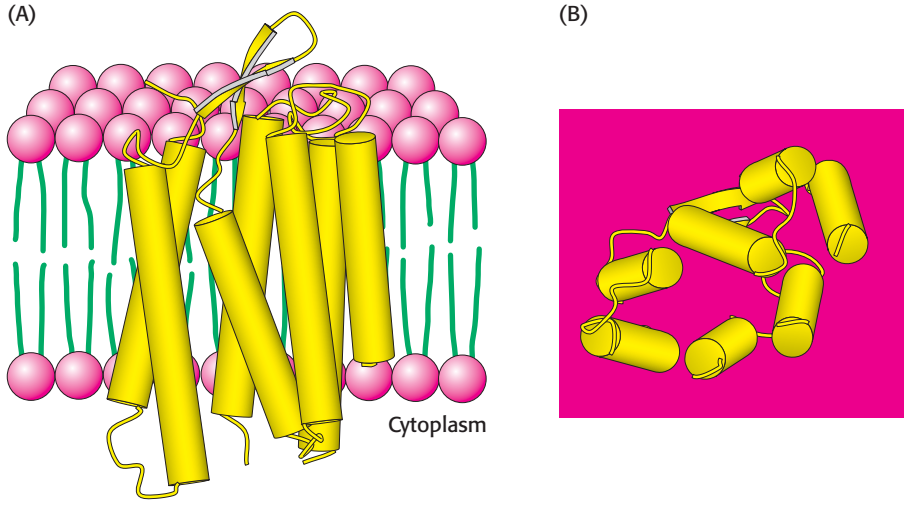


Figure 12.18 Structure of bacteriorhodopsin. *Notice* that bacteriorhodopsin consists largely of membrane-spanning α helices (represented by yellow cylinders). (A) View through the membrane bilayer. The interior of the membrane is green and the head groups are red. (B) View from the cytoplasmic side of the membrane. [Drawn from 1BRX.pdb.]

Proteins can span the membrane with alpha helices. The first membrane protein that we consider is the archaeal protein *bacteriorhodopsin*, shown in Figure 12.18. This protein uses light energy to transport protons from inside to outside the cell, generating a proton gradient used to form ATP. Bacteriorhodopsin is built almost entirely of α helices; seven closely packed α helices, arranged almost perpendicularly to the plane of the cell membrane, span its 45-Å width. Examination of the primary structure of bacteriorhodopsin reveals that most of the amino acids in these membrane-spanning α helices are nonpolar and only a very few are charged (Figure 12.19). This distribution of nonpolar amino acids is sensible because these residues are either in contact with the hydrocarbon core of the membrane or with one another. *Membrane-spanning α helices are the most common structural motif in membrane proteins.* As will be considered in Section 12.5, such regions can often be detected by examining amino acid sequence alone.

A Q I T G R P E W I W L A L G T A L M G L G T L Y F L V K G M G V S D P D A K K F Y A I T T L V P A
I A F T M Y L S M L L G Y G L T M V P F G G E Q N P I Y W A R Y A D W L F T T P L L L L D L A L L V
D A D Q G T I L A L V G A D G I M I G T G L V G A L T K V Y S Y R F V W W A I S T A A M L Y I L Y V
L F F G F T S K A E S M R P E V A S T F K V L R N V T V V L W S A Y V V V W L I G S E G A G I V P L
N I E T L L F M V L D V S A K V G F G L I L L R S R A I F G E A E A P E P S A D G A A A T S

Figure 12.19 Amino acid sequence of bacteriorhodopsin. The seven helical regions are highlighted in yellow and the charged residues in red.

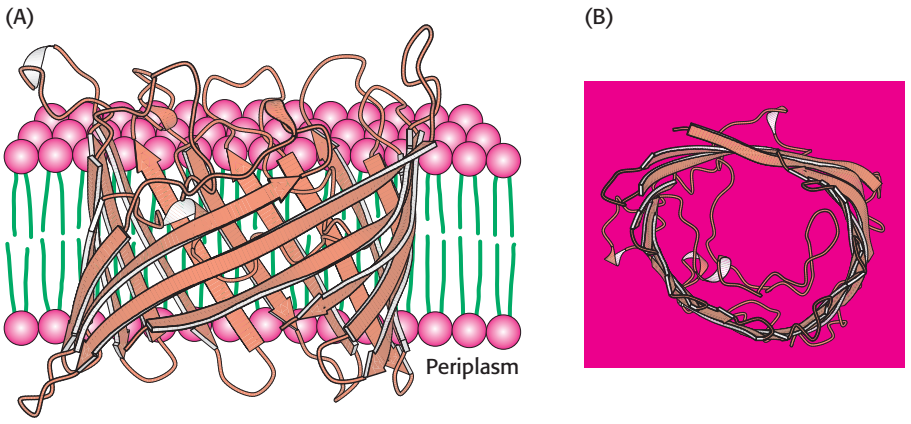


Figure 12.20 Structure of bacterial porin (from *Rhodopseudomonas blastica*). *Notice* that this membrane protein is built entirely of β strands. (A) Side view. (B) View from the periplasmic space. Only one monomer of the trimeric protein is shown. [Drawn from 1PRN.pdb.]

A channel protein can be formed from beta strands. Porin, a protein from the outer membranes of bacteria such as *E. coli* and *Rhodobacter capsulatus,* represents a class of membrane proteins with a completely different type of structure. Structures of this type are built from β strands and contain essentially no α helices (Figure 12.20).

The arrangement of β strands is quite simple: each strand is hydrogen bonded to its neighbor in an antiparallel arrangement, forming a single β sheet. The β sheet curls up to form a hollow cylinder that, as its name suggests, forms a pore, or channel, in the membrane. The outside surface of porin is appropriately nonpolar, given that it interacts with the hydrocarbon core of the membrane. In contrast, the inside of the channel is quite hydrophilic and is filled with water. This arrangement of nonpolar and polar surfaces is accomplished by the alternation of hydrophobic and hydrophilic amino acids along each β strand (Figure 12.21).

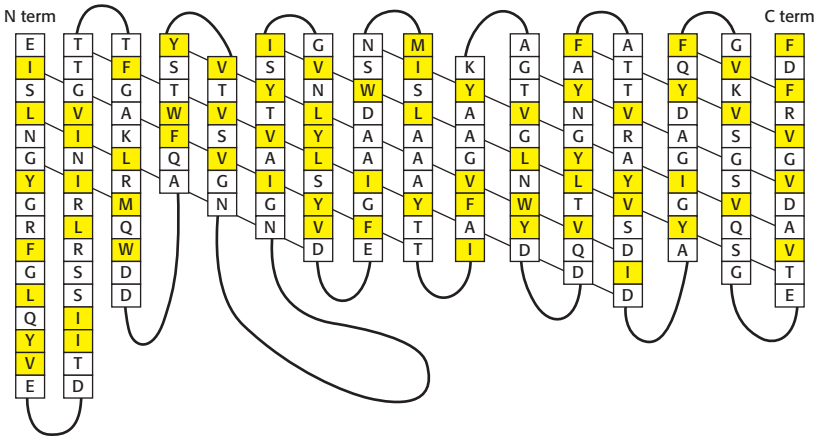


Figure 12.21 Amino acid sequence of a porin. Some membrane proteins, such as porins, are built from β strands that tend to have hydrophobic and hydrophilic amino acids in adjacent positions. The secondary structure of porin from *Rhodopseudomonas blastica* is shown, with the diagonal lines indicating the direction of hydrogen bonding along the β sheet. Hydrophobic residues (F, I, L, M, V, W, and Y) are shown in yellow. These residues tend to lie on the outside of the structure, in contact with the hydrophobic core of the membrane.

Embedding part of a protein in a membrane can link the protein to the membrane surface. The structure of the endoplasmic reticulum membrane-bound enzyme prostaglandin H_2 synthase-1 reveals a rather different role for α helices in protein–membrane associations. This enzyme

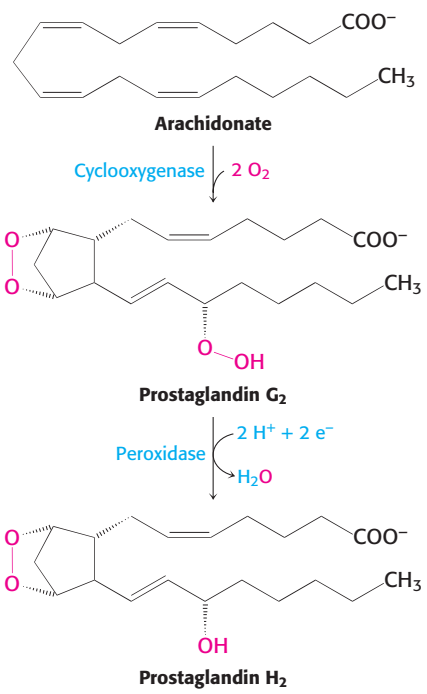


Figure 12.22 Formation of prostaglandin H_2. Prostaglandin H_2 synthase-1 catalyzes the formation of prostaglandin H_2 from arachidonic acid in two steps.

catalyzes the conversion of arachidonic acid into prostaglandin H_2 in two steps: (1) a cyclooxygenase reaction and (2) a peroxidase reaction (Figure 12.22). Prostaglandin H_2 promotes inflammation and modulates gastric acid secretion. The enzyme that produces prostaglandin H_2 is a homodimer with a rather complicated structure consisting primarily of α helices. Unlike bacteriorhodopsin, this protein is not largely embedded in the membrane. Instead, it lies along the outer surface of the membrane, firmly bound by a set of α helices with hydrophobic surfaces that extend from the bottom of the protein into the membrane (Figure 12.23). This linkage is sufficiently strong that only the action of detergents can release the protein from the membrane. Thus, this enzyme is classified as an integral membrane protein, although it does not span the membrane.

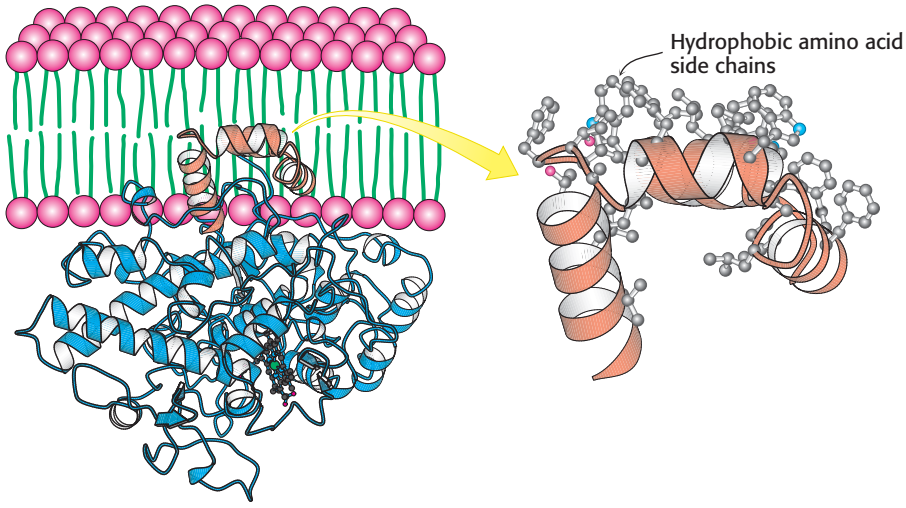


Figure 12.23 Attachment of prostaglandin H_2 synthase-1 to the membrane. *Notice* that prostaglandin H_2 synthase-1 is held in the membrane by a set of α helices (orange) coated with hydrophobic side chains. One monomer of the dimeric enzyme is shown. [Drawn from 1PTH.pdb.]

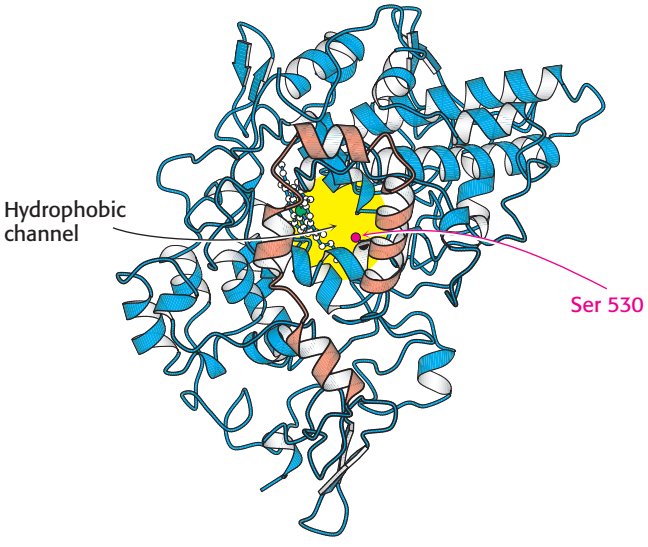


Figure 12.24 Hydrophobic channel of prostaglandin H_2 synthase-1. A view of prostaglandin H_2 synthase-1 from the membrane shows the hydrophobic channel that leads to the active site. The membrane-anchoring helices are shown in orange. [Drawn from 1PTH.pdb.]

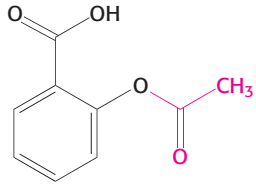
The localization of prostaglandin H_2 synthase-l in the membrane is crucial to its function. The substrate for this enzyme, arachidonic acid, is a hydrophobic molecule generated by the hydrolysis of membrane lipids. Arachidonic acid reaches the active site of the enzyme from the membrane without entering an aqueous environment by traveling through a hydrophobic channel in the protein (Figure 12.24). Indeed, nearly all of us have experienced the importance of this channel: drugs such as aspirin and ibuprofen block the channel and prevent prostaglandin synthesis by inhibiting the cyclooxygenase activity of the synthase. In particular, aspirin acts through the transfer of its acetyl group to a serine residue (Ser 530) that lies along the path to the active site (Figure 12.25).

Two important features emerge from our examination of these three examples of membrane-protein structure. First, the parts of the protein that interact with the hydrophobic parts of the membrane are coated with

nonpolar amino acid side chains, whereas those parts that interact with the aqueous environment are much more hydrophilic. Second, the structures positioned within the membrane are quite regular and, in particular, all backbone hydrogen-bond donors and acceptors participate in hydrogen bonds. *Breaking a hydrogen bond within a membrane is quite unfavorable, because little or no water is present to compete for the polar groups.*

Some proteins associate with membranes through covalently attached hydrophobic groups

The membrane proteins considered thus far associate with the membrane through surfaces generated by hydrophobic amino acid side chains. However, even otherwise soluble proteins can associate with membranes if hydrophobic groups are attached to the proteins. Three such groups are shown in Figure 12.26: (1) a palmitoyl group attached to a specific cysteine residue by a thioester bond, (2) a farnesyl group attached to a cysteine residue at the carboxyl terminus, and (3) a glycolipid structure termed a glycosylphosphatidylinositol (GPI) anchor attached to the carboxyl terminus. These modifications are attached by enzyme systems that recognize specific signal sequences near the site of attachment.



Aspirin
(Acetylsalicyclic acid)

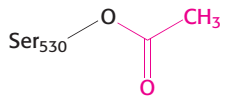
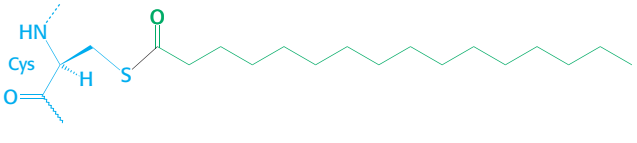
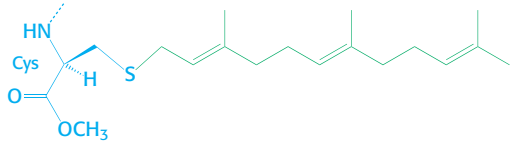


Figure 12.25 Aspirin's effects on prostaglandin H_2 synthase-1. Aspirin acts by transferring an acetyl group to a serine residue in prostaglandin H_2 synthase-1.



***S*-Palmitoylcysteine**



C-terminal *S*-farnesylcysteine methyl ester

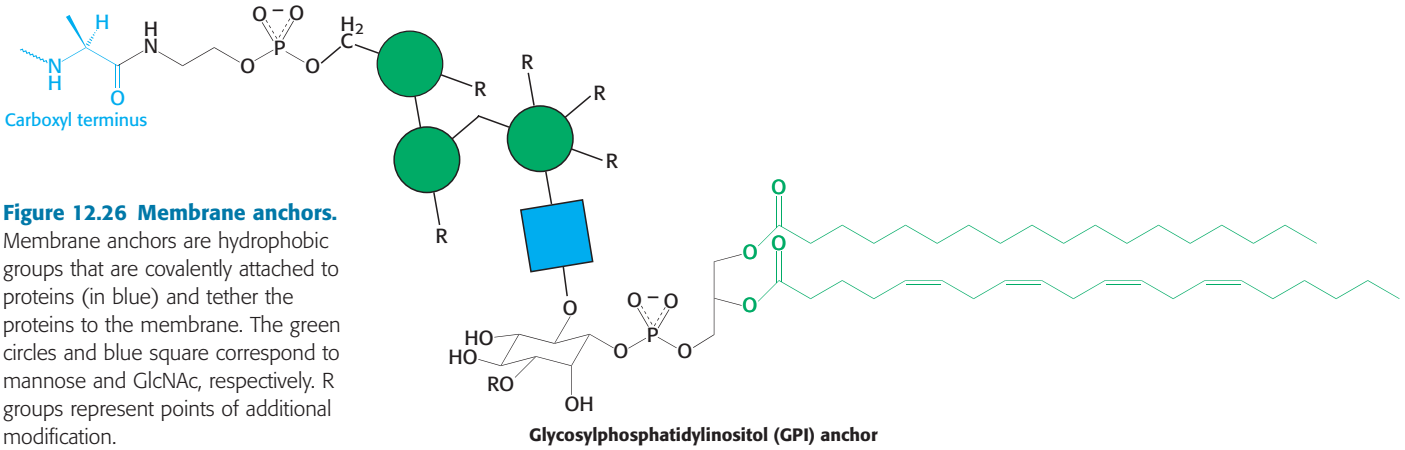


Figure 12.26 Membrane anchors. Membrane anchors are hydrophobic groups that are covalently attached to proteins (in blue) and tether the proteins to the membrane. The green circles and blue square correspond to mannose and GlcNAc, respectively. R groups represent points of additional modification.

Glycosylphosphatidylinositol (GPI) anchor

Transmembrane helices can be accurately predicted from amino acid sequences

Many membrane proteins, like bacteriorhodopsin, employ α helices to span the hydrophobic part of a membrane. As noted earlier, typically most of the residues in these α helices are nonpolar and almost none of them are charged. Can we use this information to identify likely membrane-spanning regions from sequence data alone? One approach to identifying transmembrane helices is to ask whether a postulated helical segment is likely to be more stable in a hydrocarbon environment or in water. Specifically, we want to estimate the free-energy change when a helical segment is transferred

Table 12.2 Polarity scale for identifying transmembrane helices

Amino acid residue	Transfer free energy in kJ mol^{-1} (kcal mol^{-1})
Phe	15.5 (3.7)
Met	14.3 (3.4)
Ile	13.0 (3.1)
Leu	11.8 (2.8)
Val	10.9 (2.6)
Cys	8.4 (2.0)
Trp	8.0 (1.9)
Ala	6.7 (1.6)
Thr	5.0 (1.2)
Gly	4.2 (1.0)
Ser	2.5 (0.6)
Pro	−0.8 (−0.2)
Tyr	−2.9 (−0.7)
His	−12.6 (−3.0)
Gln	−17.2 (−4.1)
Asn	−20.2 (−4.8)
Glu	−34.4 (−8.2)
Lys	−37.0 (−8.8)
Asp	−38.6 (−9.2)
Arg	−51.7 (−12.3)

Source: After D. M. Engelman, T. A. Steitz, and A. Goldman. *Annu. Rev. Biophys. Biophys. Chem.* 15(1986):321–353.
Note: The free energies are for the transfer of an amino acid residue in an α helix from the membrane interior (assumed to have a dielectric constant of 2) to water.

from the interior of a membrane to water. Free-energy changes for the transfer of individual amino acid residues from a hydrophobic to an aqueous environment are given in Table 12.2. For example, the transfer of a helix formed entirely of L-arginine residues, a positively charged amino acid, from the interior of a membrane to water would be highly favorable [−51.5 kJ mol^{-1} (−12.3 kcal mol^{-1}) per arginine residue in the helix]. In contrast, the transfer of a helix formed entirely of L-phenylalanine, a hydrophobic amino acid, would be unfavorable [+15.5 kJ mol^{-1} (+3.7 kcal mol^{-1}) per phenylalanine residue in the helix].

The hydrocarbon core of a membrane is typically 30 Å wide, a length that can be traversed by an α helix consisting of 20 residues. We can take the amino acid sequence of a protein and estimate the free-energy change that takes place when a hypothetical α helix formed of residues 1 through 20 is transferred from the membrane interior to water. The same calculation can be made for residues 2 through 21, 3 through 22, and so forth, until we reach the end of the sequence. The span of 20 residues chosen for this calculation is called a *window*. The free-energy change for each window is plotted against the first amino acid at the window to create a *hydropathy plot*. Empirically, a peak of +84 kJ mol^{-1} (+20 kcal mol^{-1}) or more in a hydropathy plot based on a window of 20 residues indicates that a polypeptide segment could be a membrane-spanning α helix. For example, glycophorin, a protein found in the membranes of red blood cells, is predicted by this criterion to have one membrane-spanning helix, in agreement with experimental findings (Figure 12.27). Note, however, that a peak in the hydropathy plot does not prove that a segment is a transmembrane helix. Even soluble proteins may have highly nonpolar regions. Conversely, some membrane proteins contain membrane-spanning features (such as a set of cylinder-forming β strands) that escape detection by these plots (Figure 12.28).

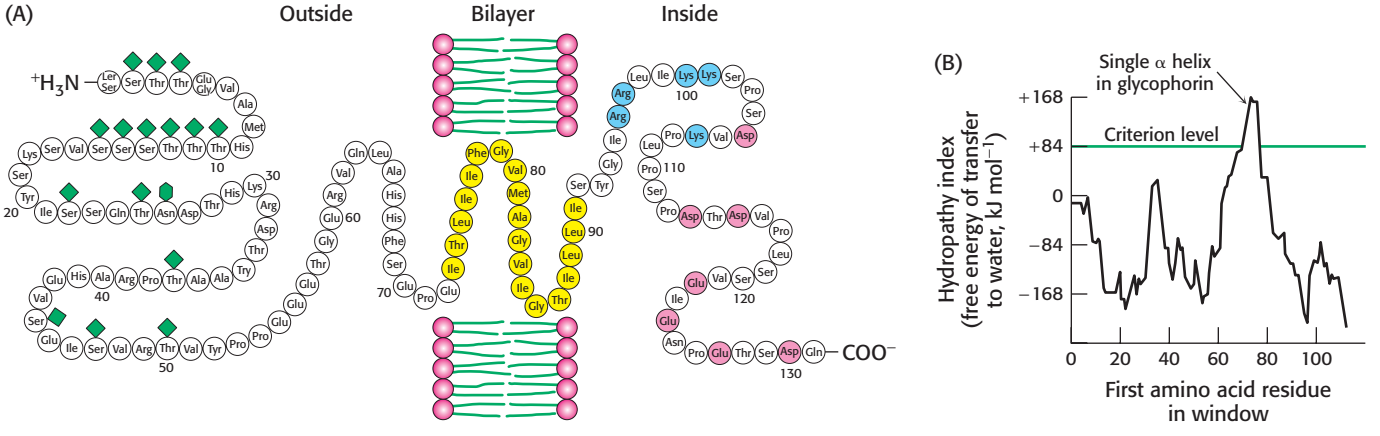


Figure 12.27 Locating the membrane-spanning helix of glycophorin. (A) Amino acid sequence and transmembrane disposition of glycophorin A from the red-blood-cell membrane. Fifteen *O*-linked carbohydrate units are shown as diamond shapes, and an *N*-linked unit is shown as a lozenge shape. The hydrophobic residues (yellow) buried in the bilayer form a transmembrane α helix. The carboxyl-terminal part of the molecule, located on the cytoplasmic side of the membrane, is rich in negatively charged (red) and positively charged (blue) residues.
(B) Hydropathy plot for glycophorin. The free energy for transferring a helix of 20 residues from the membrane to water is plotted as a function of the position of the first residue of the helix in the sequence of the protein. Peaks of greater than +84 kJ mol^{-1} (+20 kcal mol^{-1}) in hydropathy plots are indicative of potential transmembrane helices. [(A) Courtesy of Dr. Vincent Marchesi; (B) after D. M. Engelman, T. A. Steitz, and A. Goldman. *Annu. Rev. Biophys. Biophys. Chem. 15*: 321–353, 1986. Copyright © 1986 by Annual Reviews, Inc. All rights reserved.]

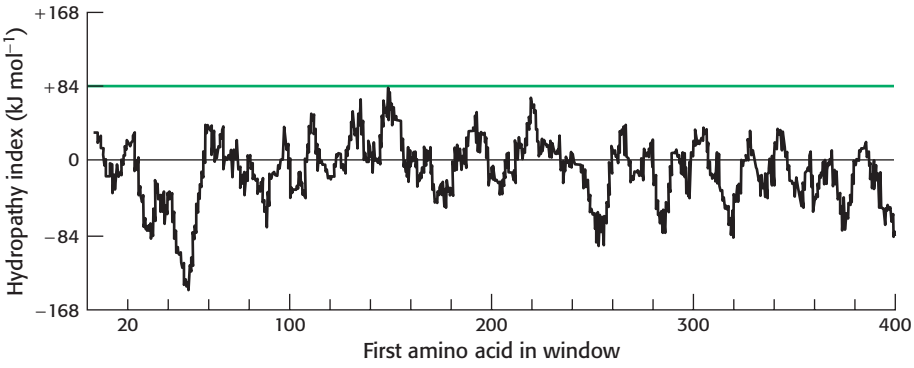


Figure 12.28 Hydropathy plot for porin. No strong peaks are observed for this intrinsic membrane protein, because it is constructed from membrane-spanning β strands rather than α helices.

12.5 Lipids and Many Membrane Proteins Diffuse Rapidly in the Plane of the Membrane

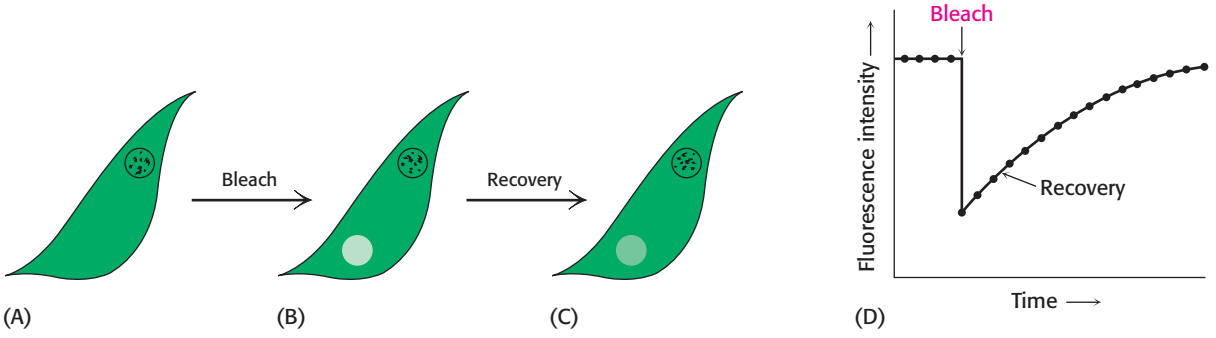
Biological membranes are not rigid, static structures. On the contrary, lipids and many membrane proteins are constantly in lateral motion, a process called *lateral diffusion*. The rapid lateral movement of membrane proteins has been visualized by means of fluorescence microscopy using the technique of *fluorescence recovery after photobleaching* (FRAP; Figure 12.29). First, a cell-surface component is specifically labeled with a fluorescent chromophore. A small region of the cell surface ($\sim 3\ \mu m^2$) is viewed through a fluorescence microscope. The fluorescent molecules in this region are then destroyed (bleached) by a very intense light pulse from a laser. The fluorescence of this region is subsequently monitored as a function of time by using a light level sufficiently low to prevent further bleaching. If the labeled component is mobile, bleached molecules leave and unbleached molecules enter the illuminated region, resulting in an increase in the fluorescence intensity. The rate of recovery of fluorescence depends on the lateral mobility of the fluorescence-labeled component, which can be expressed in terms of a diffusion coefficient, D. The average distance S traversed in time t depends on D according to the expression

$$S = (4Dt)^{1/2}$$

The diffusion coefficient of lipids in a variety of membranes is about $1\ \mu m^2\ s^{-1}$. Thus, a phospholipid molecule diffuses an average distance of 2 μm in 1 s. This rate means that *a lipid molecule can travel from one end of a bacterium to the other in a second*. The magnitude of the observed diffusion coefficient indicates that the viscosity of the membrane is about 100 times that of water, rather like that of olive oil.

In contrast, proteins vary markedly in their lateral mobility. *Some proteins are nearly as mobile as lipids, whereas others are virtually immobile*. For example, the photoreceptor protein rhodopsin (Section 33.3), a very mobile

Figure 12.29 Fluorescence recovery after photobleaching (FRAP) technique. (A) The cell surface fluoresces because of a labeled surface component. (B) The fluorescent molecules of a small part of the surface are bleached by an intense light pulse. (C) The fluorescence intensity recovers as bleached molecules diffuse out of the region and unbleached molecules diffuse into it. (D) The rate of recovery depends on the diffusion coefficient.



protein, has a diffusion coefficient of 0.4 $\mu m^2\ s^{-1}$. The rapid movement of rhodopsin is essential for fast signaling. At the other extreme is fibronectin, a peripheral glycoprotein that interacts with the extracellular matrix. For fibronectin, D is less than $10^{-4}\ \mu m^2\ s^{-1}$. Fibronectin has a very low mobility because it is anchored to actin filaments on the inside of the plasma membrane through *integrin*, a transmembrane protein that links the extracellular matrix to the cytoskeleton.

The fluid mosaic model allows lateral movement but not rotation through the membrane

On the basis of the mobility of proteins in membranes, in 1972 S. Jonathan Singer and Garth Nicolson proposed a *fluid mosaic model* to describe the overall organization of biological membranes. The essence of their model is that *membranes are two-dimensional solutions of oriented lipids and globular proteins.* The lipid bilayer has a dual role: it is both a *solvent* for integral membrane proteins and a *permeability barrier.* Membrane proteins are free to diffuse laterally in the lipid matrix unless restricted by special interactions.

Although the lateral diffusion of membrane components can be rapid, the spontaneous rotation of lipids from one face of a membrane to the other is a very slow process. The transition of a molecule from one membrane surface to the other is called *transverse diffusion* or *flip-flop* (Figure 12.30) The flip-flop of phospholipid molecules in phosphatidylcholine vesicles has been directly measured by electron spin resonance techniques, which show that *a phospholipid molecule flip-flops once in several hours.* Thus, a phospholipid molecule takes about 10^9 times as long to flip-flop across a membrane as it takes to diffuse a distance of 50 Å in the lateral direction. The free-energy barriers to flip-flopping are even larger for protein molecules than for lipids because proteins have more-extensive polar regions. In fact, the flip-flop of a protein molecule has not been observed. Hence, *membrane asymmetry can be preserved for long periods.*

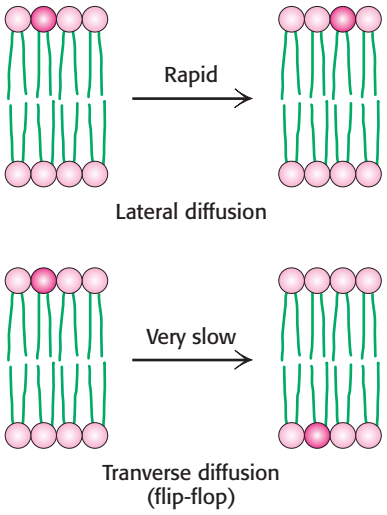


Figure 12.30 Lipid movement in membranes. Lateral diffusion of lipids is much more rapid than transverse diffusion (flip-flop).

Membrane fluidity is controlled by fatty acid composition and cholesterol content

Many membrane processes, such as transport or signal transduction, depend on the fluidity of the membrane lipids, which in turn depends on the properties of fatty acid chains. Fatty acid chains in membrane bilayers can exist in an ordered, rigid state or in a relatively disordered, fluid state. The transition from the rigid to the fluid state takes place abruptly as the temperature is raised above T_m, the melting temperature (Figure 12.31). *This transition temperature depends on the length of the fatty acid chains and on their degree of unsaturation* (Table 12.3). The presence of saturated fatty

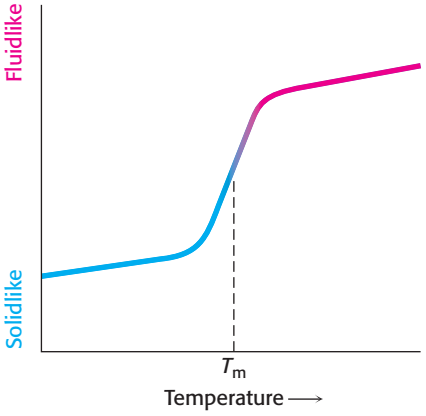


Figure 12.31 The phase-transition, or melting, temperature (T_m) for a phospholipid membrane. As the temperature is raised, the phospholipid membrane changes from a packed, ordered state to a more random one.

Table 12.3 The melting temperature of phosphatidylcholine containing different pairs of identical fatty acid chains

		Fatty acid		
Number of carbons	Number of double bonds	Common name	Systematic name	T_m (°C)
22	0	Behenate	*n*-Docosanote	75
18	0	Stearate	*n*-Octadecanoate	58
16	0	Palmitate	*n*-Hexadecanoate	41
14	0	Myristate	*n*-Tetradecanoate	24
18	1	Oleate	*cis*-Δ^9-Octadecenoate	−22

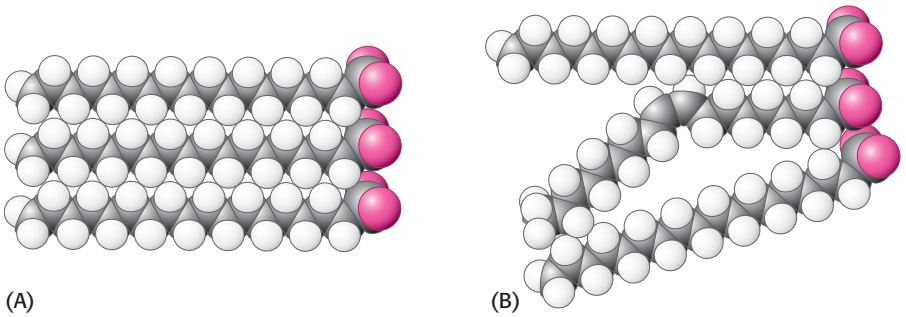


Figure 12.32 Packing of fatty acid chains in a membrane. The highly ordered packing of fatty acid chains is disrupted by the presence of cis double bonds. The space-filling models show the packing of (A) three molecules of stearate (C_{18}, saturated) and (B) a molecule of oleate (C_{18}, unsaturated) between two molecules of stearate.

acid residues favors the rigid state because their straight hydrocarbon chains interact very favorably with one another. On the other hand, *a cis double bond produces a bend in the hydrocarbon chain. This bend interferes with a highly ordered packing of fatty acid chains, and so* T_m *is lowered* (Figure 12.32). The length of the fatty acid chain also affects the transition temperature. Long hydrocarbon chains interact more strongly than do short ones. Specifically, each additional —CH_2— group makes a favorable contribution of about -2 kJ mol^{-1} (-0.5 kcal mol^{-1}) to the free energy of interaction of two adjacent hydrocarbon chains.

Bacteria regulate the fluidity of their membranes by varying the number of double bonds and the length of their fatty acid chains. For example, the ratio of saturated to unsaturated fatty acid chains in the *E. coli* membrane decreases from 1.6 to 1.0 as the growth temperature is lowered from 42°C to 27°C. This decrease in the proportion of saturated residues prevents the membrane from becoming too rigid at the lower temperature.

In animals, cholesterol is the key regulator of membrane fluidity. Cholesterol contains a bulky steroid nucleus with a hydroxyl group at one end and a flexible hydrocarbon tail at the other end. Cholesterol inserts into bilayers with its long axis perpendicular to the plane of the membrane. The hydroxyl group of cholesterol forms a hydrogen bond with a carbonyl oxygen atom of a phospholipid head group, whereas the hydrocarbon tail of cholesterol is located in the nonpolar core of the bilayer. The different shape of cholesterol compared with that of phospholipids disrupts the regular interactions between fatty acid chains (Figure 12.33)

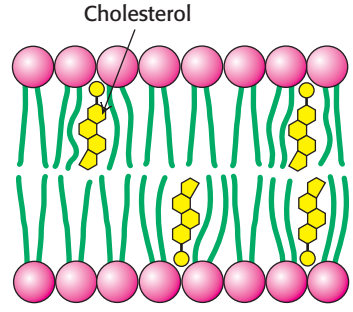


Figure 12.33 Cholesterol disrupts the tight packing of the fatty acid chains. [After S. L. Wolfe, *Molecular and Cellular Biology* (Wadsworth, 1993).]

Lipid rafts are highly dynamic complexes formed between cholesterol and specific lipids

In addition to its nonspecific effects on membrane fluidity, cholesterol can form specific complexes with lipids that contain the sphingosine backbone, including sphingomyelin and certain glycolipids, and with GPI-anchored proteins. These complexes concentrate within small (10–200 nm) and highly dynamic regions within membranes. The resulting structures are often referred to as *lipid rafts*. One result of these interactions is the *moderation of membrane fluidity,* making membranes less fluid but at the same time less subject to phase transitions. The presence of lipid rafts thus represents a modification of the original fluid mosaic model for biological membranes. Although their small size and dynamic nature have made them very difficult to study, it appears that lipid rafts may play a role in concentrating proteins that participate in signal transduction pathways and may also serve to regulate membrane curvature and budding.

All biological membranes are asymmetric

Membranes are structurally and functionally asymmetric. The outer and inner surfaces of *all known biological membranes have different components*

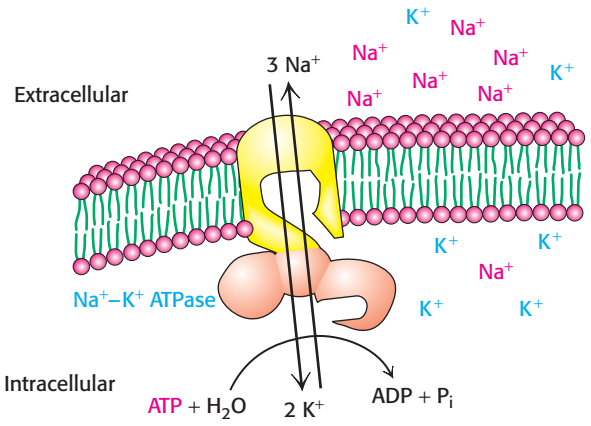


Figure 12.34 Asymmetry of the Na^+–K^+ transport system in plasma membranes. The Na^+–K^+ transport system pumps Na^+ out of the cell and K^+ into the cell by hydrolyzing ATP on the intracellular side of the membrane.

and different enzymatic activities. A clear-cut example is the pump that regulates the concentration of Na^+ and K^+ ions in cells (Figure 12.34). This transport protein is located in the plasma membrane of nearly all cells in higher organisms. The Na^+–K^+ pump is oriented so that it pumps Na^+ out of the cell and K^+ into it. Furthermore, ATP must be on the inside of the cell to drive the pump. Ouabain, a specific inhibitor of the pump, is effective only if it is located outside. We shall consider the mechanism of this important and fascinating pump and others in its family in Chapter 13.

Membrane proteins have a unique orientation because, after synthesis, they are inserted into the membrane in an asymmetric manner. This absolute asymmetry is preserved because membrane proteins do not rotate from one side of the membrane to the other and because *membranes are always synthesized by the growth of preexisting membranes.* Lipids, too, are asymmetrically distributed as a consequence of their mode of biosynthesis, but this asymmetry is usually not absolute, except for glycolipids. In the red-blood-cell membrane, sphingomyelin and phosphatidylcholine are preferentially located in the outer leaflet of the bilayer, whereas phosphatidylethanolamine and phosphatidylserine are located mainly in the inner leaflet. Large amounts of cholesterol are present in both leaflets.

12.6 Eukaryotic Cells Contain Compartments Bounded by Internal Membranes

Thus far, we have considered only the plasma membrane of cells. Some bacteria and archaea have only this single membrane, surrounded by a cell wall. Other bacteria, such as *E. coli,* have two membranes separated by a cell wall (made of proteins, peptides, and carbohydrates) lying between them (Figure 12.35). The inner membrane acts as the permeability barrier, and the outer membrane and the cell wall provide additional protection. The

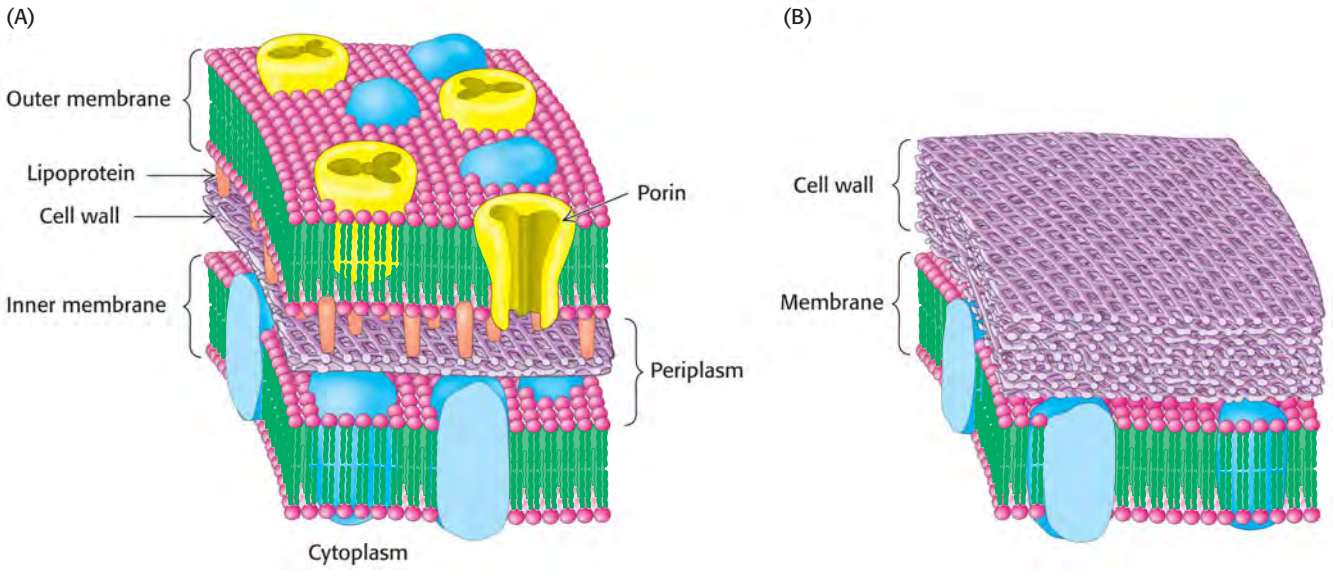


Figure 12.35 Cell membranes of prokaryotes. A schematic view of the membrane of bacterial cells surrounded by (A) two membranes or (B) one membrane.

outer membrane is quite permeable to small molecules, owing to the presence of porins. The region between the two membranes containing the cell wall is called the *periplasm*.

Eukaryotic cells, with the exception of plant cells, do not have cell walls, and their cell membranes consist of a single lipid bilayer. In plant cells, the cell wall is on the outside of the plasma membrane. Eukaryotic cells are distinguished from prokaryotic cells by the presence of membranes inside the cell that form internal compartments. For example, peroxisomes, organelles that play a major role in the oxidation of fatty acids for energy conversion, are defined by a single membrane. Mitochondria, the organelles in which ATP is synthesized, are surrounded by two membranes. As in the case for a bacterium, the outer membrane is quite permeable to small molecules, whereas the inner membrane is not. Indeed, considerable evidence now indicates that mitochondria evolved from bacteria by *endosymbiosis* (Section 18.1). The nucleus is also surrounded by a double membrane, the *nuclear envelope*, that consists of a set of closed membranes that come together at structures called *nuclear pores* (Figure 12.36). These pores regulate transport into and out of the nucleus. The nuclear envelope is linked to another membrane-defined structure, the *endoplasmic reticulum*, which plays a host of cellular roles, including drug detoxification and the modification of proteins for secretion. Thus, a eukaryotic cell contains interacting compartments, and transport into and out of these compartments is essential to many biochemical processes.

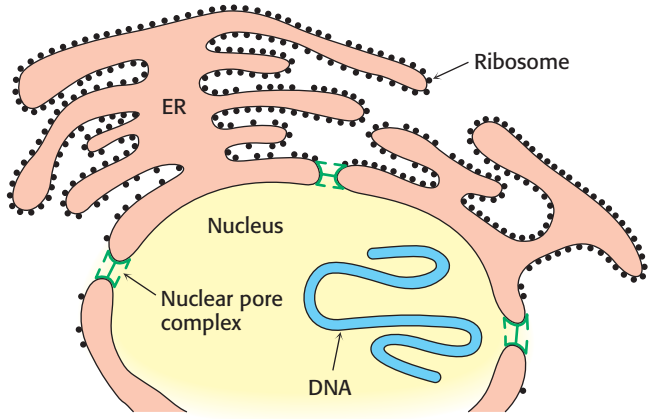


Figure 12.36 Nuclear envelope. The nuclear envelope is a double membrane connected to another membrane system of eukaryotes, the endoplasmic reticulum. [After E. C. Schirmer and L. Gerace. *Genome Biol.* 3(4):1008.1–1008.4, 2002, reviews, Fig.1.]

Membranes must be able to separate or join together so that cells and compartments may take up, transport, and release molecules. Many cells take up molecules through the process of *receptor-mediated endocytosis*. Here, a protein or larger complex initially binds to a receptor on the cell surface. After the receptor is bound, specialized proteins act to cause the membrane in this region to invaginate. One of these specialized proteins is *clathrin*, which polymerizes into a lattice network around the growing membrane bud, often referred to as a *clathrin-coated pit* (Figure 12.37). The invaginated membrane eventually breaks off and fuses to form a *vesicle*. Various hormones, transport proteins, and antibodies employ receptor-mediated endocytosis to gain entry into a cell. A less-advantageous consequence is that this pathway is available to viruses and toxins as a means

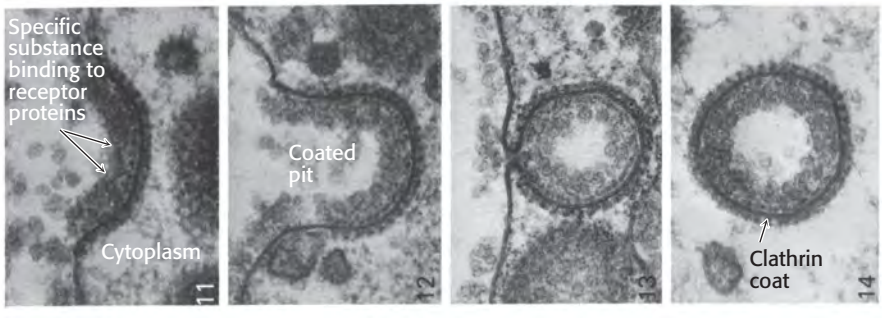


Figure 12.37 Vesicle formation by receptor-mediated endocytosis. Receptor binding on the surface of the cell induces the membrane to invaginate, with the assistance of specialized intracellular proteins such as clathrin. The process results in the formation of a vesicle within the cell. [M. M. Perry and A. B. Gilbert. *J. Cell Sci.* 39:266, 1979.]

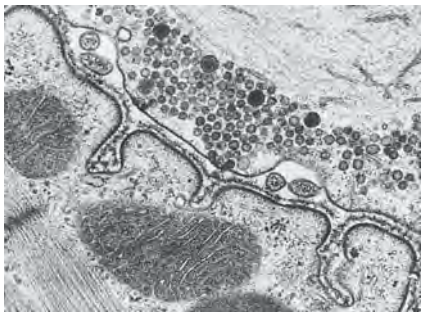


Figure 12.38 Neurotransmitter release. Neurotransmitter-containing synaptic vesicles are arrayed near the plasma membrane of a nerve cell. Synaptic vesicles fuse with the plasma membrane, releasing the neurotransmitter into the synaptic cleft. [T. Reese/Don Fawcett/Photo Researchers.]

of entry into cells. The reverse process—the fusion of a vesicle to a membrane—is a key step in the release of neurotransmitters from a neuron into the synaptic cleft (Figure 12.38).

Let us consider one example of receptor-mediated endocytosis. Iron is a critical element for the function and structure of many proteins, including hemoglobin and myoglobin (Chapter 7). However, free iron ions are highly toxic to cells, owing to their ability to catalyze the formation of free radicals. Hence, the transport of iron atoms from the digestive tract to the cells where they are most needed must be tightly controlled. In the bloodstream, iron is bound very tightly by the protein *transferrin*, which can bind two Fe^{3+} ions with a dissociation constant of 10^{-23} M at neutral pH. Cells requiring iron express the *transferrin receptor* in their plasma membranes (Section 32.4). Formation of a complex between the transferrin receptor and iron-bound transferrin initiates receptor-mediated endocytosis, internalizing these complexes within vesicles called *endosomes* (Figure 12.39). As the endosomes mature, proton pumps within the vesicle membrane lower the lumenal pH to about 5.5. Under these conditions, the affinity of iron ions for transferrin is reduced; these ions are released and are free to pass through channels in the endosomal membranes into the cytoplasm. The iron-free transferrin complex is recycled to the plasma membrane, where transferrin is released back into the bloodstream and the transferrin receptor can participate in another uptake cycle.

Although budding and fusion appear deceptively simple, the structures of the intermediates in these processes and the detailed mechanisms remain on-going areas of investigation. Key membrane components called *SNARE* (soluble *N*-ethylmaleimide-sensitive-factor attachment protein receptor) *proteins* help draw appropriate membranes together to initiate the fusion process. These proteins, encoded by gene families in all eukaryotic cells, largely determine the compartment with which a vesicle will fuse. The specificity of membrane fusion ensures the orderly trafficking of membrane vesicles and their cargos through eukaryotic cells.

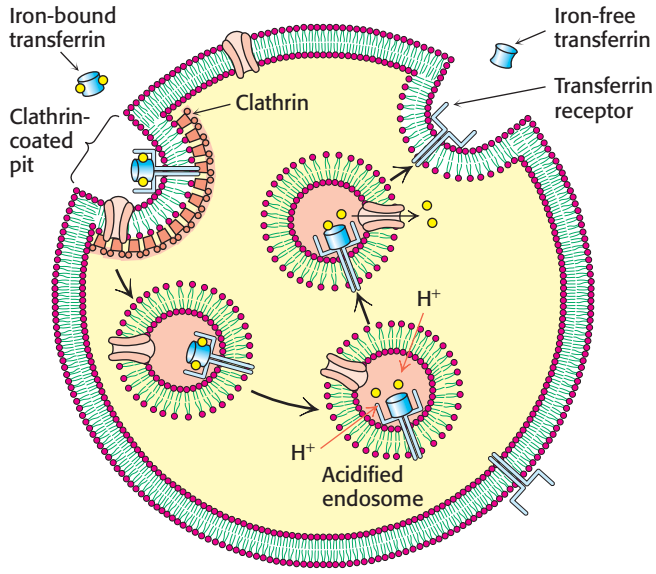


Figure 12.39 The transferrin receptor cycle. Iron-bound transferrin binds to the transferrin receptor (TfR) on the surface of cells. Receptor-mediated endocytosis occurs, leading to the formation of a vesicle called an endosome. As the lumen of the endosome is acidified by the action of proton pumps, iron is released from transferrin, passes through channels in the membrane, and is utilized by the cell. The complex between iron-free transferrin and the transferrin receptor is returned to the plasma membrane for another cycle. [After L. Zecca et al. *Nat. Rev. Neurosci.* 5:863–873, 2004, Fig.1.]

Summary

Biological membranes are sheetlike structures, typically from 60 to 100 Å thick, that are composed of protein and lipid molecules held together by noncovalent interactions. Membranes are highly selective permeability barriers. They create closed compartments, which may be entire cells or organelles within a cell. Proteins in membranes regulate the molecular and ionic compositions of these compartments. Membranes also control the flow of information between cells.

12.1 Fatty Acids Are Key Constituents of Lipids

Fatty acids are hydrocarbon chains of various lengths and degrees of unsaturation that terminate with a carboxylic acid group. The fatty acid chains in membranes usually contain between 14 and 24 carbon atoms; they may be saturated or unsaturated. Short chain length and unsaturation enhance the fluidity of fatty acids and their derivatives by lowering the melting temperature.

12.2 There Are Three Common Types of Membrane Lipids

The major types of membrane lipids are phospholipids, glycolipids, and cholesterol. Phosphoglycerides, a type of phospholipid, consist of a glycerol backbone, two fatty acid chains, and a phosphorylated alcohol. Phosphatidylcholine, phosphatidylserine, and phosphatidylethanolamine are major phosphoglycerides. Sphingomyelin, a different type of phospholipid, contains a sphingosine backbone instead of glycerol. Glycolipids are sugar-containing lipids derived from sphingosine. Cholesterol, which modulates membrane fluidity, is constructed from a steroid nucleus. A common feature of these membrane lipids is that they are amphipathic molecules, having one hydrophobic and one hydrophilic end.

12.3 Phospholipids and Glycolipids Readily Form Bimolecular Sheets in Aqueous Media

Membrane lipids spontaneously form extensive bimolecular sheets in aqueous solutions. The driving force for membrane formation is the hydrophobic interactions among the fatty acid tails of membrane lipids. The hydrophilic head groups interact with the aqueous medium. Lipid bilayers are cooperative structures, held together by many weak bonds. These lipid bilayers are highly impermeable to ions and most polar molecules, yet they are quite fluid, which enables them to act as a solvent for membrane proteins.

12.4 Proteins Carry Out Most Membrane Processes

Specific proteins mediate distinctive membrane functions such as transport, communication, and energy transduction. Many integral membrane proteins span the lipid bilayer, whereas others are only partly embedded in the membrane. Peripheral membrane proteins are bound to membrane surfaces by electrostatic and hydrogen-bond interactions. Membrane-spanning proteins have regular structures, including β strands, although the α helix is the most common membrane-spanning structure. Sequences of 20 consecutive nonpolar amino acids can be diagnostic of a membrane-spanning α-helical region of a protein.

12.5 Lipids and Many Membrane Proteins Diffuse Rapidly in the Plane of the Membrane

Membranes are structurally and functionally asymmetric, as exemplified by the restriction of sugar residues to the external surface of

mammalian plasma membranes. Membranes are dynamic structures in which proteins and lipids diffuse rapidly in the plane of the membrane (lateral diffusion), unless restricted by special interactions. In contrast, the rotation of lipids from one face of a membrane to the other (transverse diffusion, or flip-flop) is usually very slow. Proteins do not rotate across bilayers; hence, membrane asymmetry can be preserved. The degree of fluidity of a membrane depends on the chain length of its lipids and on the extent to which their constituent fatty acids are unsaturated. In animals, cholesterol content also regulates membrane fluidity.

12.6 Eukaryotic Cells Contain Compartments Bounded by Internal Membranes

An extensive array of internal membranes in eukaryotes creates compartments within a cell for distinct biochemical functions. For instance, a double membrane surrounds the nucleus, the location of most of the cell's genetic material, and the mitochondria, the location of most ATP synthesis. A single membrane defines the other internal compartments, such as the endoplasmic reticulum. Receptor-mediated endocytosis enables the formation of intracellular vesicles when ligands bind to their corresponding receptor proteins in the plasma membrane. The reverse process—the fusion of a vesicle to a membrane—is a key step in the release of signaling molecules outside the cell.

Key Terms

fatty acid (p. 346)
phospholipid (p. 348)
sphingosine (p. 348)
phosphoglyceride (p. 348)
sphingomyelin (p. 349)
glycolipid (p. 349)
cerebroside (p. 350)
ganglioside (p. 350)
cholesterol (p. 350)
amphipathic (amphiphilic) molecule (p. 351)
lipid bilayer (p. 352)
liposome (p. 353)
integral membrane protein (p. 355)
peripheral membrane protein (p. 355)
hydropathy plot (p. 360)
lateral diffusion (p. 361)
fluid mosaic model (p. 362)
lipid raft (p. 363)
receptor-mediated endocytosis (p. 365)
clathrin (p. 365)
transferrin (p. 366)
transferrin receptor (p. 366)
endosome (p. 366)
SNARE (soluble *N*-ethylmaleimide-sensitive-factor attachment protein receptor) proteins (p. 366)

Problems

1. *Population density*. How many phospholipid molecules are there in a 1-μm^2 region of a phospholipid bilayer membrane? Assume that a phospholipid molecule occupies 70 $Å^2$ of the surface area.

2. *Through the looking-glass*. Phospholipids form lipid bilayers in water. What structure might form if phospholipids were placed in an organic solvent?

3. *Lipid diffusion*. What is the average distance traversed by a membrane lipid in 1 μs, 1 ms, and 1 s? Assume a diffusion coefficient of 10^{-8} cm^2 s^{-1}.

4. *Protein diffusion*. The diffusion coefficient, D, of a rigid spherical molecule is given by

$$D = kT/6\pi\eta r$$

in which η is the viscosity of the solvent, r is the radius of the sphere, k is the Boltzman constant (1.38×10^{-16} erg degree^{-1}), and T is the absolute temperature. What is the diffusion coefficient at 37°C of a 100-kd protein in a membrane that has an effective viscosity of 1 poise (1 poise = 1 erg s^{-1} cm^{-3})? What is the average distance traversed by this protein in 1 μs, 1 ms, and 1 s? Assume that this protein is an unhydrated, rigid sphere of density 1.35 g cm^{-3}.

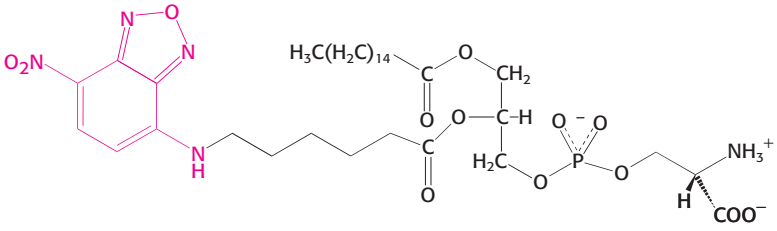
5. *Cold sensitivity*. Some antibiotics act as carriers that bind an ion on one side of a membrane, diffuse through the membrane, and release the ion on the other side. The conductance of a lipid-bilayer membrane containing a carrier antibiotic decreased abruptly when the temperature was lowered from 40°C to 36°C. In contrast, there was little change in conductance of the same bilayer membrane when it contained a channel-forming antibiotic. Why?

6. *Melting point 1*. Explain why oleic acid (18 carbons, one cis bond) has a lower melting point than stearic acid, which has the same number of carbon atoms but is saturated. How would you expect the melting point of *trans*-oleic acid to compare with that of *cis*-oleic acid? Why might most unsaturated fatty acids in phospholipids be in the cis rather than the trans conformation?

7. *Melting point 2*. Explain why the melting point of palmitic acid (C_{16}) is 6.5 degrees lower than that of stearic acid (C_{18}).

8. *A sound diet*. Small mammalian hibernators can withstand body temperatures of 0° to 5°C without injury. However, the body fats of most mammals have melting temperatures of approximately 25°C. Predict how the composition of the body fat of hibernators might differ from that of their nonhibernating cousins.

9. *Flip-flop 1*. The transverse diffusion of phospholipids in a bilayer membrane was investigated by using a fluorescently labeled analog of phosphatidylserine called NBD-PS.



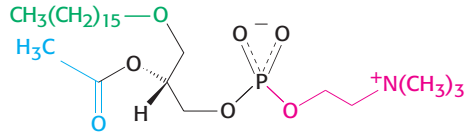
NBD-phosphatidylserine (NBD-PS)

The fluorescence signal of NBD-PS is quenched when exposed to sodium dithionite, a reducing agent that is not membrane permeable.

Lipid vesicles containing phosphatidylserine (98%) and NBD-PS (2%) were prepared by sonication and purified. Within a few minutes of the addition of sodium dithionite, the fluorescence signal of these vesicles decreased to ~45% of its initial value. Immediately adding a second addition of sodium dithionite yielded no change in the fluorescence signal. However, if the vesicles were allowed to incubate for 6.5 hours, a third addition of sodium dithionite decreased the remaining fluorescence signal by 50%. How would you interpret the fluorescence changes at each addition of sodium dithionite?

10. *Flip-flop 2*. Although proteins rarely if ever flip-flop across a membrane, the distribution of membrane lipids between the membrane leaflets is not absolute except for glycolipids. Why are glycosylated lipids less likely to flip-flop?

11. *Linkages*. Platelet-activating factor (PAF) is a phospholipid that plays a role in allergic and inflammatory responses, as well as in toxic shock syndrome. The structure of PAF is shown here. How does it differ from the structures of the phospholipids discussed in this chapter?



Platelet-activating factor (PAF)

12. *A question of competition*. Would a homopolymer of alanine be more likely to form an α helix in water or in a hydrophobic medium? Explain.

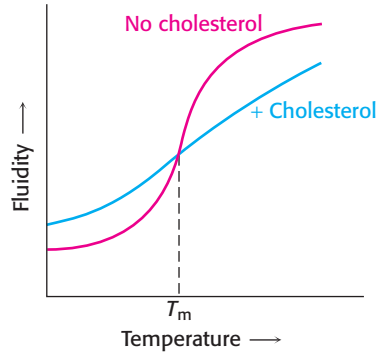
13. *A false positive*. Hydropathy plot analysis of your protein of interest reveals a single, prominent hydrophobic peak. However, you later discover that this protein is soluble and not membrane associated. Explain how the hydropathy plot may have been misleading.

14. *Maintaining fluidity*. A culture of bacteria growing at 37°C was shifted to 25°C. How would you expect this shift to alter the fatty acid composition of the membrane phospholipids? Explain.

15. *Let me count the ways*. Each intracellular fusion of a vesicle with a membrane requires a SNARE protein on the vesicle (called the v-SNARE) and a SNARE protein on the target membrane (called the t-SNARE). Assume that a genome encodes 21 members of the v-SNARE family and 7 members of the t-SNARE family. With the assumption of no specificity, how many potential v-SNARE–t-SNARE interactions could take place?

Data Interpretation Problems

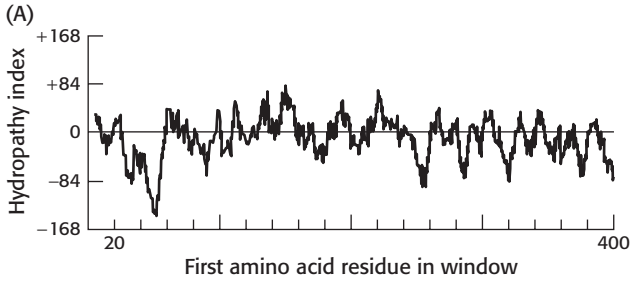
16. *Cholesterol effects*. The red curve on the following graph shows the fluidity of the fatty acids of a phospholipid bilayer as a function of temperature. The blue curve shows the fluidity in the presence of cholesterol.

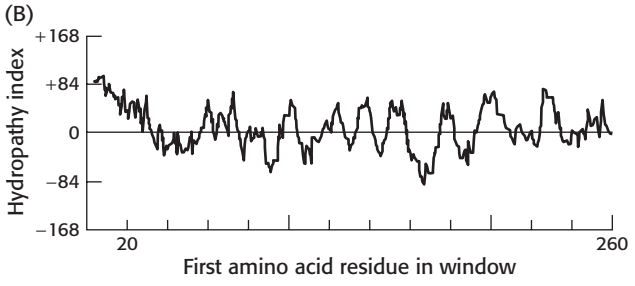


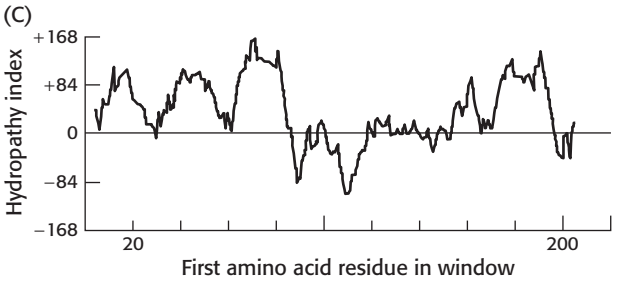
(a) What is the effect of cholesterol?

(b) Why might this effect be biologically important?

17. *Hydropathy plots*. On the basis of the following hydropathy plots for three proteins (A–C), predict which would be membrane proteins. What are the ambiguities with respect to using such plots to determine if a protein is a membrane protein?





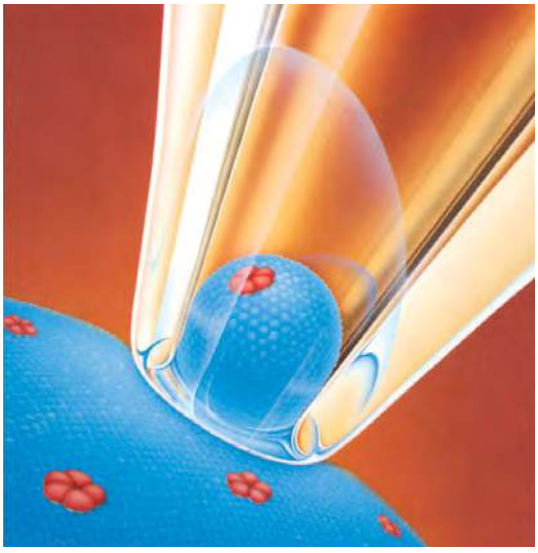


Chapter Integration Problem

18. *The proper environment*. An understanding of the structure and function of membrane proteins has lagged behind that of other proteins. The primary reason is that membrane proteins are more difficult to purify and crystallize. Why might this be the case?

CHAPTER 13

Membrane Channels and Pumps



The flow of ions through a single membrane channel (channels are shown in red in the illustration at the left) can be detected by the patch-clamp technique, which records current changes as the channel transits between open and closed states. [(Left) After E. Neher and B. Sakmann. The patch clamp technique. Copyright © 1992 by Scientific American, Inc. All rights reserved. (Right) Courtesy of Dr. Mauricio Montal.]

The lipid bilayer of biological membranes is intrinsically impermeable to ions and polar molecules, yet certain such species must be able to cross these membranes for normal cell function. Permeability is conferred by three classes of membrane proteins, *pumps, carriers,* and *channels.* Pumps use a source of free energy such as ATP hydrolysis or light absorption to drive the thermodynamically uphill transport of ions or molecules. Pump action is an example of *active transport*. Carriers mediate the transport of ions and small molecules across the membrane without consumption of ATP. Channels provide a membrane pore through which ions can flow very rapidly in a thermodynamically downhill direction. The action of channels illustrates *passive transport,* or *facilitated diffusion.*

Pumps are energy transducers in that they convert one form of free energy into another. Two types of *ATP-driven pumps*, P-type ATPases and the ATP-binding cassette (ABC) transporters, undergo conformational changes on ATP binding and hydrolysis that cause a bound ion to be transported across the membrane. The free energy of ATP hydrolysis is used to drive the movement of ions against their concentration gradients, a process referred to as *primary active transport*. In contrast, carriers utilize the gradient of one ion to drive the transport of another against its gradient. An example of this process, termed *secondary active transport,* is mediated by the *E. coli* lactose transporter, a well-studied protein responsible for the uptake of a specific sugar from the environment of a bacterium. Many transporters of this class are present in the membranes of our cells. The expression of these transporters determines which metabolites a cell can import from the environment. Hence, adjusting the level of transporter expression is a primary means of controlling metabolism.

OUTLINE

Pumps can establish persistent gradients of particular ions across membranes. Specific *ion channels* can allow these ions to flow rapidly across membranes down these gradients. These channels are among the most fascinating molecules in biochemistry in their ability to allow some ions to flow freely through a membrane while blocking the flow of even closely related species. The opening, or gating, of these channels can be controlled by the presence of certain ligands or a particular membrane voltage. Gated ion channels are central to the functioning of our nervous systems, acting as elaborately switched wires that allow the rapid flow of current.

Finally, a different class of channel, the cell-to-cell channel, or *gap junction*, allows the flow of metabolites or ions *between cells*. For example, gap junctions are responsible for synchronizing muscle-cell contraction in the beating heart.

The expression of transporters largely defines the metabolic activities of a given cell type

Each cell type expresses a specific set of transporters in its plasma membrane. This collection of expressed transporters is important because it largely determines the ionic composition inside cells and the compounds that can be taken up from the cell's environment. In some senses, the cell-specific array of transporters defines the cell's characteristics because a cell can execute only those biochemical reactions for which it has taken up the necessary substrates.

An example from glucose metabolism illustrates this point. As we will see in the discussion of glucose metabolism in Chapter 16, tissues differ in their ability to employ different molecules as energy sources. Which tissues can utilize glucose is largely governed by the expression of different members of a family of homologous glucose transporters called GLUT1, GLUT2, GLUT3, GLUT4, and GLUT5. For example, GLUT3 is expressed only on neurons and a few other cell types. This transporter binds glucose relatively tightly so that these cells have first call on glucose when it is present at relatively low concentrations. These are just the first of many examples that we will encounter that demonstrate the critical role that transporter expression plays in the control and integration of metabolism.

13.1 The Transport of Molecules Across a Membrane May Be Active or Passive

We first consider some general principles of membrane transport. Two factors determine whether a molecule will cross a membrane: (1) the permeability of the molecule in a lipid bilayer and (2) the availability of an energy source.

Many molecules require protein transporters to cross membranes

As stated in Chapter 12, some molecules can pass through cell membranes because they dissolve in the lipid bilayer. Such molecules are called *lipophilic molecules*. The steroid hormones provide a physiological example. These cholesterol relatives can pass through a membrane, but what determines the direction in which they will move? Such molecules will pass through a membrane down their concentration gradient in a process called *simple diffusion*. In accord with the Second Law of Thermodynamics, molecules spontaneously move from a region of higher concentration to one of lower concentration.

Matters become more complicated when the molecule is highly polar. For example, sodium ions are present at 143 mM outside a typical cell and at 14 mM inside the cell. However, sodium does not freely enter the cell, because the charged ion cannot pass through the hydrophobic membrane interior. In some circumstances, as during a nerve impulse, sodium ions must enter the cell. How are they able to do so? Sodium ions pass through specific channels in the hydrophobic barrier formed by membrane proteins. This means of crossing the membrane is called *facilitated diffusion* because the diffusion across the membrane is facilitated by the channel. It is also called *passive transport* because the energy driving the ion movement originates from the ion gradient itself, without any contribution by the transport system. Channels, like enzymes, display substrate specificity in that they facilitate the transport of some ions, but not other, even closely related, ions.

How is the sodium gradient established in the first place? In this case, sodium must move, or be pumped, *against* a concentration gradient. Because moving the ion from a low concentration to a higher concentration results in a decrease in entropy, it requires an input of free energy. Protein transporters embedded in the membrane are capable of using an energy source to move the molecule up a concentration gradient. Because an input of energy from another source is required, this means of crossing the membrane is called *active transport*.

Free energy stored in concentration gradients can be quantified

An unequal distribution of molecules is an energy-rich condition because free energy is minimized when all concentrations are equal. Consequently, to attain such an unequal distribution of molecules requires an input of free energy. How can we quantify the amount of energy required to generate a concentration gradient (Figure 13.1)? Consider an uncharged solute molecule. The free-energy change in transporting this species from side 1, where it is present at a concentration of c_1, to side 2, where it is present at concentration c_2, is

$$\Delta G = RT \ln (c_2/c_1)$$

where R is the gas constant (8.315×10^{-3} kJ mol^{-1} deg^{-1}, or 1.987×10^{-3} kcal mol^{-1} deg^{-1}) and T is the temperature in kelvins. For a charged species, the unequal distribution across the membrane generates an electrical potential that also must be considered because the ions will be repelled by the like charges. The sum of the concentration and electrical terms is called the *electrochemical potential* or *membrane potential*. The free-energy change is then given by

$$\Delta G = RT \ln (c_2/c_1) + ZF\Delta V$$

in which Z is the electrical charge of the transported species, ΔV is the potential in volts across the membrane, and F is the Faraday constant (96.5 kJ V^{-1} mol^{-1}, or 23.1 kcal V^{-1} mol^{-1}).

A transport process must be active when ΔG is positive, whereas it can be passive when ΔG is negative. For example, consider the transport of an uncharged molecule from $c_1 = 10^{-3}$ M to $c_2 = 10^{-1}$ M.

$$\begin{aligned}\Delta G &= RT \ln (10^{-1}/10^{-3}) \\ &= (8.315 \times 10^{-3}) \times 298 \times \ln (10^2) \\ &= +11.4 \text{ kJ mol}^{-1} (+2.7 \text{ kcal mol}^{-1})\end{aligned}$$

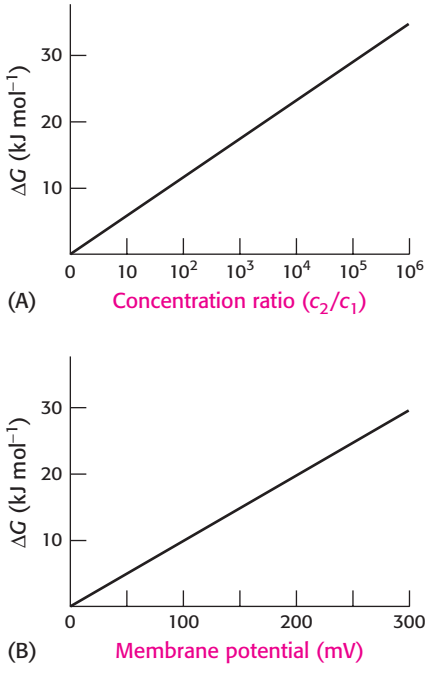


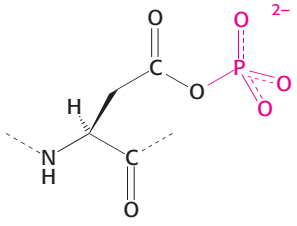
Figure 13.1 Free energy and transport. The free-energy change in transporting (A) an uncharged solute from a compartment at concentration c_1 to one at c_2 and (B) a singly charged species across a membrane to the side having the same charge as that of the transported ion. Note that the free-energy change imposed by a membrane potential of 59 mV is equivalent to that imposed by a concentration ratio of 10 for a singly charged ion at 25°C.

At 25°C (298 K), ΔG is +11.4 kJ mol^{-1} (+2.7 kcal mol^{-1}), indicating that this transport process requires an input of free energy.

13.2 Two Families of Membrane Proteins Use ATP Hydrolysis to Pump Ions and Molecules Across Membranes

The extracellular fluid of animal cells has a salt concentration similar to that of seawater. However, cells must control their intracellular salt concentrations to facilitate specific processes, such as signal transduction and action potential propagation, and prevent unfavorable interactions with high concentrations of ions such as Ca^{2+}. For instance, most animal cells contain a high concentration of K^+ and a low concentration of Na^+ relative to the external medium. These ionic gradients are generated by a specific transport system, an enzyme that is called the *Na^+–K^+ pump* or the *Na^+–K^+ ATPase.* The hydrolysis of ATP by the pump provides the energy needed for the active transport of Na^+ out of the cell and K^+ into the cell, generating the gradients. The pump is called the Na^+–K^+ ATPase because the hydrolysis of ATP takes place only when Na^+ and K^+ are present. This ATPase, like all such enzymes, requires Mg^{2+}.

The change in free energy accompanying the transport of Na^+ and K^+ can be calculated. Suppose that the concentrations of Na^+ outside and inside the cell are 143 and 14 mM, respectively, and the corresponding values for K^+ are 4 and 157 mM. At a membrane potential of −50 mV and a temperature of 37°C, we can use the equation on page 373 to determine that the free-energy change in transporting 3 mol of Na^+ out of the cell and 2 mol of K^+ into the cell is $3(5.99) + 2(9.46) = +36.9$ kJ mol^{-1} (+8.8 kcal mol^{-1}). Under typical cellular conditions, the hydrolysis of a single ATP molecule per transport cycle provides sufficient free energy, about −50 kJ mol^{-1} (–12 kcal mol^{-1}) to drive the uphill transport of these ions. The active transport of Na^+ and K^+ is of great physiological significance. Indeed, more than a third of the ATP consumed by a resting animal is used to pump these ions. The Na^+–K^+ gradient in animal cells controls cell volume, renders neurons and muscle cells electrically excitable, and drives the active transport of sugars and amino acids.



Phosphorylaspartate

The purification of other ion pumps has revealed a large family of evolutionarily related ion pumps including proteins from bacteria, archaea, and all eukaryotes. Each of these pumps is specific for a particular ion or set of ions. Two are of particular interest: the *sarcoplasmic reticulum Ca^{2+} ATPase* (or *SERCA*) transports Ca^{2+} out of the cytoplasm and into the sarcoplasmic reticulum of muscle cells, and the *gastric H^+–K^+ ATPase* is the enzyme responsible for pumping sufficient protons into the stomach to lower the pH to 1.0. These enzymes and the hundreds of known homologs, including the Na^+–K^+ ATPase, are referred to as *P-type ATPases* because they form a key phosphorylated intermediate. In the formation of this intermediate, a phosphoryl group from ATP is linked to the side chain of a specific conserved aspartate residue in the ATPase to form phosphorylaspartate.

P-type ATPases couple phosphorylation and conformational changes to pump calcium ions across membranes

Membrane pumps function by mechanisms that are simple in principle but often complex in detail. Fundamentally, each pump protein can exist in two principal conformational states, one with ion-binding sites open to one side of the membrane and the other with ion-binding sites open to the other

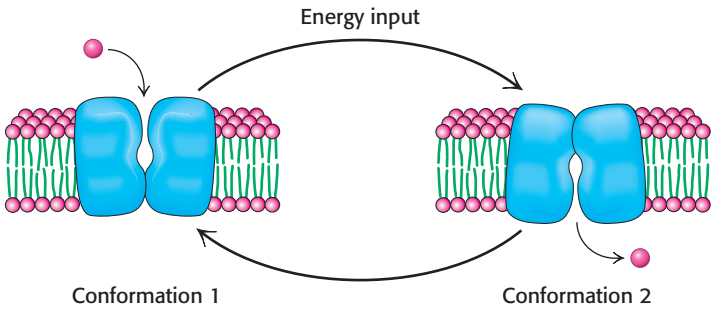


Figure 13.2 Pump action. A simple scheme for the pumping of a molecule across a membrane. The pump interconverts to two conformational states, each with a binding site accessible to a different side of the membrane.

side (Figure 13.2). To pump ions in a single direction across a membrane, the free energy of ATP hydrolysis must be coupled to the interconversion between these conformational states.

We will consider the structural and mechanistic features of P-type ATPases by examining SERCA. The properties of this P-type ATPase have been established in great detail by relying on crystal structures of the pump in five different states. This enzyme, which constitutes 80% of the protein in the sarcoplasmic reticulum membrane, plays an important role in relaxation of contracted muscle. Muscle contraction is triggered by an abrupt rise in the cytoplasmic calcium ion level. Subsequent muscle relaxation depends on the rapid removal of Ca^{2+} from the cytoplasm into the sarcoplasmic reticulum, a specialized compartment for Ca^{2+} storage, by SERCA. This pump maintains a Ca^{2+} concentration of approximately 0.1 μM in the cytoplasm compared with 1.5 mM in the sarcoplasmic reticulum.

The first structure of SERCA to be determined had Ca^{2+} bound, but no nucleotides present (Figure 13.3). SERCA is a single 110-kd polypeptide with a transmembrane domain consisting of 10 α helices. The transmembrane domain includes sites for binding two calcium ions. Each calcium ion is coordinated to seven oxygen atoms coming from a combination of side-chain glutamate, aspartate, threonine, and asparagine residues, backbone carbonyl groups, and water molecules. A large cytoplasmic headpiece constitutes nearly half the molecular weight of the protein and consists of three distinct domains, each with a distinct function. One domain (N) binds the ATP *n*ucleotide, another (P) accepts the *p*hosphoryl group on a conserved

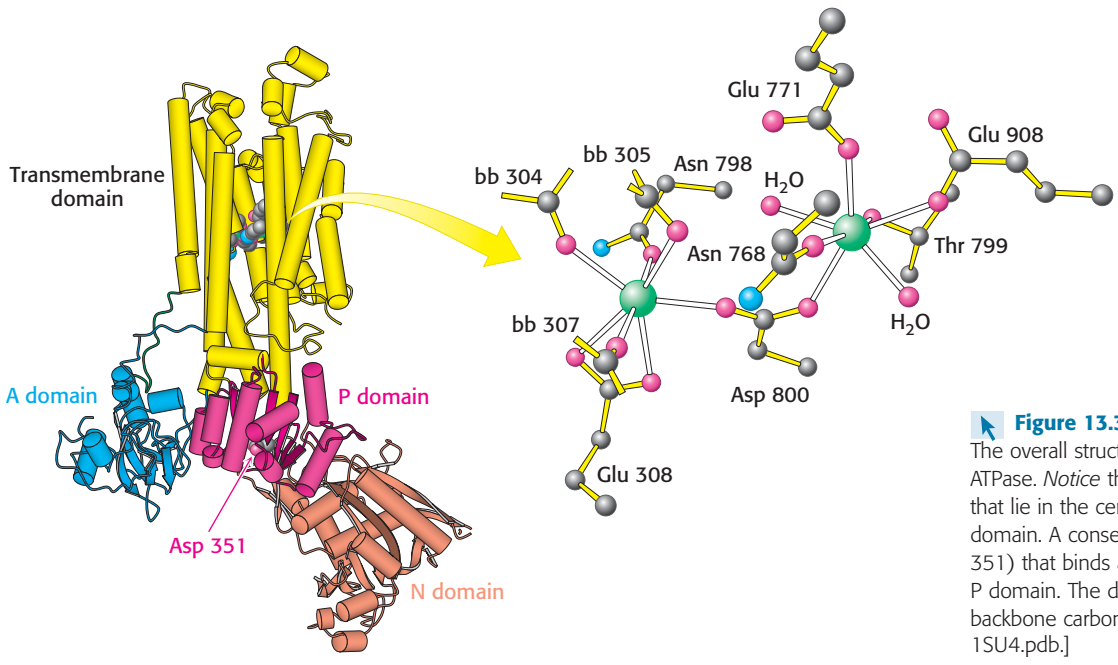


Figure 13.3 Calcium-pump structure. The overall structure of the SERCA P-type ATPase. *Notice* the two calcium ions (green) that lie in the center of the transmembrane domain. A conserved aspartate residue (Asp 351) that binds a phosphoryl group lies in the P domain. The designation bb refers to backbone carbonyl groups. [Drawn from 1SU4.pdb.]

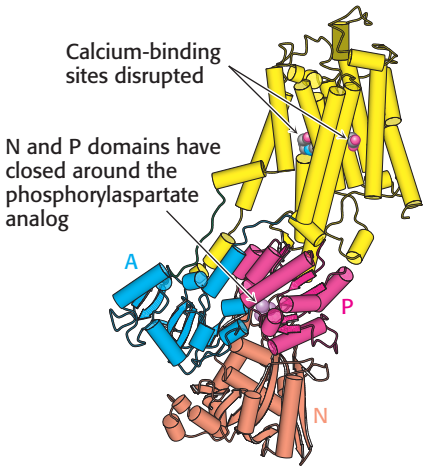


Figure 13.4 Conformational changes associated with calcium pumping. This structure was determined in the absence of bound calcium and with a phosphorylaspartate analog present in the P domain. *Notice* how different this structure is from the calcium-bound form shown in Figure 13.3: both the transmembrane part (yellow) and the A, P, and N domains have substantially rearranged. [Drawn from 1WPG.pdb.]

aspartate residue, and the third (A) serves as an *a*ctuator, linking changes in the N and P domains to the transmembrane part of the enzyme.

SERCA is remarkably structurally dynamic. For example, the structure of SERCA without bound Ca^{2+} and with a phosphorylaspartate analog present in the P domain is shown in Figure 13.4. The N and P domains are now closed around the phosphorylaspartate analog, and the A domain has rotated substantially relative to its position in SERCA with Ca^{2+} bound and without the phosphoryl analog. Furthermore, the transmembrane part of the enzyme has rearranged substantially and the well-organized Ca^{2+}-binding sites are disrupted. These sites are now accessible from the side of the membrane opposite the N, P, and A domains.

The structural results can be combined with other studies to construct a detailed mechanism for Ca^{2+} pumping by SERCA (Figure 13.5).

1. The catalytic cycle begins with the enzyme in its unphosphorylated state with two calcium ions bound. We will refer to the overall enzyme conformation in this state as E_1; with Ca^{2+} bound, it is E_1-$(Ca^{2+})_2$. In this conformation, SERCA can exchange calcium ions but only with calcium ions from the cytoplasmic side of the membrane. This conformation is shown in Figure 13.3.

2. In the E_1 conformation, the enzyme can bind ATP. The N, P, and A domains undergo substantial rearrangement as they close around the bound ATP, but there is no substantial conformational change in the transmembrane domain. The calcium ions are now trapped inside the enzyme.

3. The phosphoryl group is then transferred from ATP to Asp 351.

4. Upon ADP release, the enzyme again changes its overall conformation, including the membrane domain this time. This new conformation is referred to as E_2 or E_2-P in its phosphorylated form. The process of interconverting the E_1 and E_2 conformations is sometimes referred to as *eversion*.

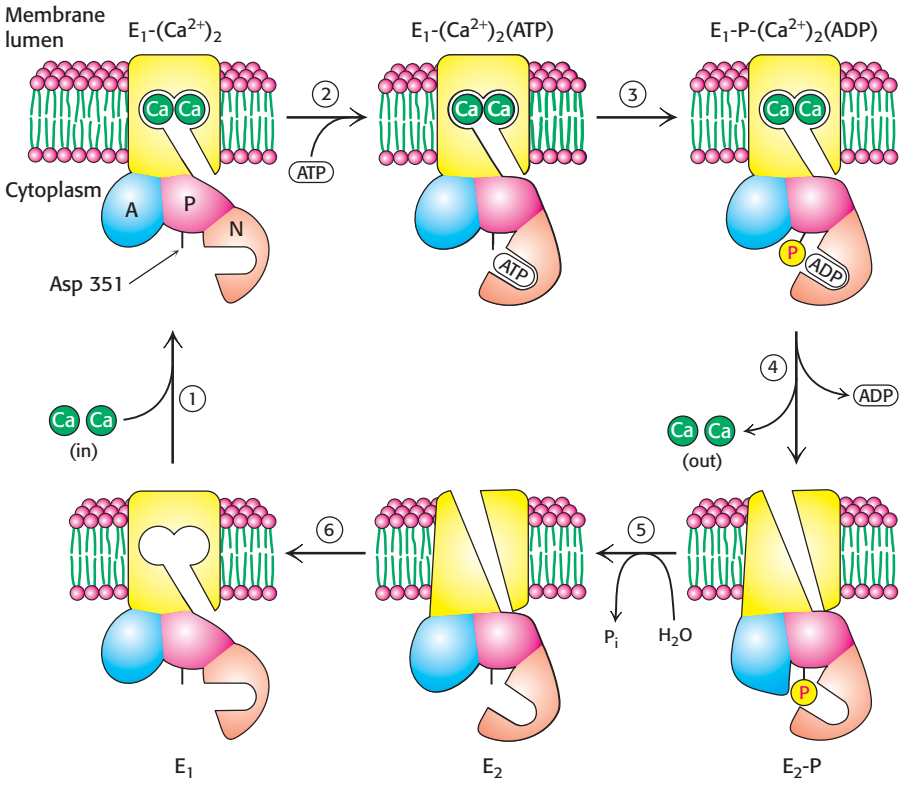


Figure 13.5 Pumping calcium. Ca^{2+}ATPase transports Ca^{2+} through the membrane by a mechanism that includes (1) Ca^{2+} binding from the cytoplasm, (2) ATP binding, (3) ATP cleavage with the transfer of a phosphoryl group to Asp 351 on the enzyme, (4) ADP release and eversion of the enzyme to release Ca^{2+} on the opposite side of the membrane, (5) hydrolysis of the phosphorylaspartate residue, and (6) eversion to prepare for the binding of Ca^{2+} from cytoplasm.

In the E_2-P conformation, the Ca^{2+}-binding sites become disrupted and the calcium ions are released to the side of the membrane opposite that at which they entered; ion transport has been achieved. This conformation is shown in Figure 13.4.

5. The phosphorylaspartate residue is hydrolyzed to release inorganic phosphate.

6. With the release of phosphate, the interactions stabilizing the E_2 conformation are lost, and the enzyme everts to the E_1 conformation.

The binding of two calcium ions from the cytoplasmic side of the membrane completes the cycle.

This mechanism likely applies to other P-type ATPases. For example, Na^+–K^+ ATPase is an $\alpha_2\beta_2$ tetramer. Its α subunit is homologous to SERCA and includes a key aspartate residue analogous to Asp 351. The β subunit does not directly take part in ion transport. A mechanism analogous to that shown in Figure 13.5 applies, with three Na^+ ions binding from the inside of the cell to the E_1 conformation and two K^+ ions binding from outside the cell to the E_2 conformation.

Digitalis specifically inhibits the Na^+–K^+ pump by blocking its dephosphorylation

Certain steroids derived from plants are potent inhibitors ($K_i \approx 10$ nM) of the Na^+–K^+ pump. Digitoxigenin and ouabain are members of this class of inhibitors, which are known as *cardiotonic steroids* because of their strong effects on the heart (Figure 13.6). These compounds inhibit the dephosphorylation of the E_2-P form of the ATPase when applied on the *extracellular* face of the membrane.

Digitalis is a mixture of cardiotonic steroids derived from the dried leaf of the foxglove plant (*Digitalis purpurea*). The compound increases the force of contraction of heart muscle and is consequently a choice drug in the treatment of congestive heart failure. Inhibition of the Na^+–K^+ pump by digitalis leads to a higher level of Na^+ inside the cell. The diminished Na^+ gradient results in slower extrusion of Ca^{2+} by the sodium–calcium exchanger. The subsequent increase in the intracellular level of Ca^{2+} enhances the ability of cardiac muscle to contract. It is interesting to note that digitalis was used effectively long before the discovery of the Na^+–K^+ ATPase. In 1785, William Withering, a British physician, heard tales of an elderly woman, known as "the old woman of Shropshire," who cured people of "dropsy" (which today would be recognized as congestive heart failure) with an extract of foxglove. Withering conducted the first scientific study of the effects of foxglove on congestive heart failure and documented its effectiveness.

Foxglove (*Digitalis purpurea*) is the source of digitalis, one of the most widely used drugs. [Inga Spence/Visuals Unlimited.]

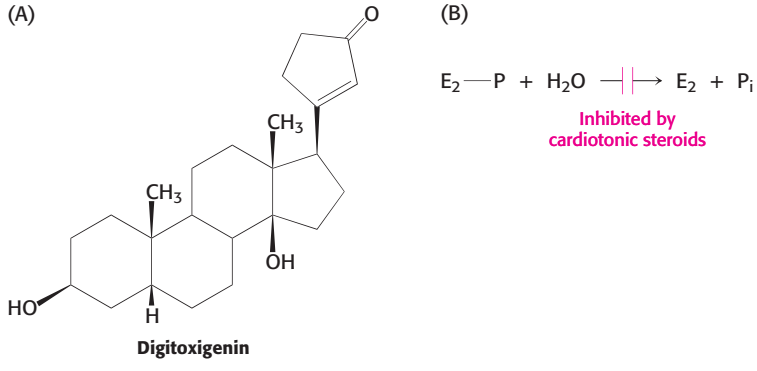


Figure 13.6 Digitoxigenin. Cardiotonic steroids such as digitoxigenin inhibit the Na^+–K^+ pump by blocking the dephosphorylation of E_2-P.

P-type ATPases are evolutionarily conserved and play a wide range of roles

Analysis of the complete yeast genome revealed the presence of 16 proteins that clearly belong to the P-type ATPase family. More-detailed sequence analysis suggests that 2 of these proteins transport H^+ ions, 2 transport Ca^{2+}, 3 transport Na^+, and 2 transport metals such as Cu^{2+}. In addition, 5 members of this family appear to participate in the transport of phospholipids with amino acid head groups. These 5 proteins help maintain membrane asymmetry by transporting lipids such as phosphatidylserine from the inner to the outer leaflet of the bilayer membrane. Such enzymes have been termed "flippases." Remarkably, the human genome encodes 70 P-type ATPases. All members of this protein family employ the same fundamental mechanism: the free energy of ATP hydrolysis drives membrane transport by means of conformational changes, which are induced by the addition and removal of a phosphoryl group at an analogous aspartate site in each protein.

Multidrug resistance highlights a family of membrane pumps with ATP-binding cassette domains

Studies of human disease revealed another large and important family of active-transport proteins, with structures and mechanisms quite different from those of the P-type ATPase family. These pumps were identified from studies on tumor cells in culture that developed resistance to drugs that had been initially quite toxic to the cells. Remarkably, the development of resistance to one drug had made the cells less sensitive to a range of other compounds. This phenomenon is known as *multidrug resistance*. In a significant discovery, the onset of multidrug resistance was found to correlate with the expression and activity of a membrane protein with an apparent molecular mass of 170 kd. This protein acts as an ATP-dependent pump that extrudes a wide range of small molecules from cells that express it. The protein is called the *multidrug-resistance* (MDR) *protein* or *P-glycoprotein* ("glyco" because it includes a carbohydrate moiety). Thus, when cells are exposed to a drug, the MDR pumps the drug out of the cell before the drug can exert its effects.

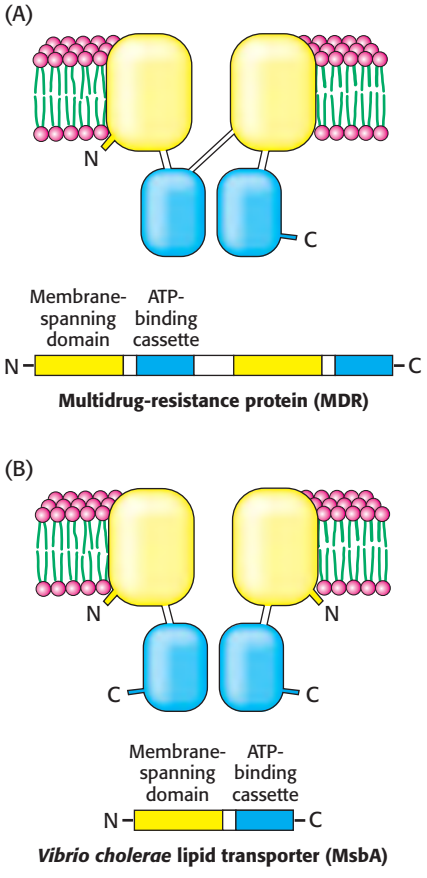


Figure 13.7 Domain arrangement of ABC transporters. ABC transporters are a large family of homologous proteins composed of two transmembrane domains and two ATP-binding domains called ATP-binding cassettes (ABCs). (A) The multidrug-resistance protein is a single polypeptide chain containing all four domains, whereas (B) the *Vibrio cholerae* lipid transporter MsbA consists of a dimer of two identical chains, containing one of each domain.

Analysis of the amino acid sequences of MDR and homologous proteins revealed a common architecture (Figure 13.7A). Each protein comprises four domains: two membrane-spanning domains and two ATP-binding domains. The ATP-binding domains of these proteins are called *ATP-binding cassettes* (ABCs) and are homologous to domains in a large family of transport proteins of bacteria and archaea. Transporters that include these domains are called *ABC transporters*. With 79 members, the ABC transporters are the largest single family identified in the *E. coli* genome. The human genome includes more than 150 ABC transporter genes.

The ABC proteins are members of the P-loop NTPase superfamily (Section 9.4). The three-dimensional structures of several members of the ABC transporter family have now been determined, including that of the lipid transporter MsbA from *Vibrio cholerae*. In contrast with the eukaryotic MDR protein, this protein is a dimer of 62-kd chains: the amino-terminal half of each protein contains the membrane-spanning domain, and the carboxyl-terminal half contains the ATP-binding cassette (Figure 13.7B). Prokaryotic ABC proteins are often made up of multiple subunits, such as a dimer of identical chains, as above, or as a heterotetramer of two membrane-spanning domain subunits and two ATP-binding-cassette subunits. The

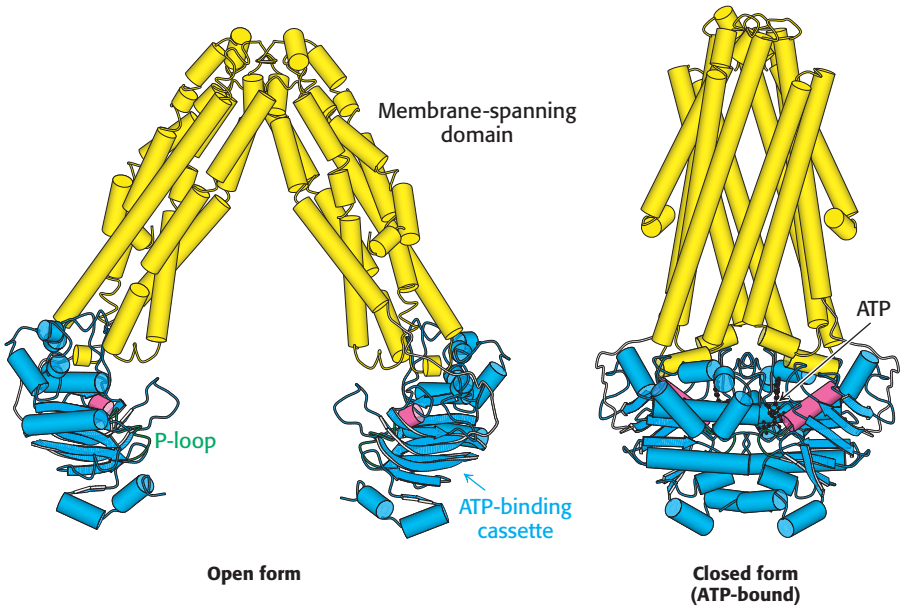


Figure 13.8 ABC transporter structure. Two structures of the lipid transporter MsbA from *Vibrio cholerae,* a representive ABC transporter. The open form is on the left and the closed, ATP-bound form is on the right. The two ATP-binding casettes (blue) are related to the P-loop NTPases and, like them, contain P-loops (green). The α helix adjacent to the P-loop is shown in red. [Drawn from 3B5W and 3B60.pdb.]

consolidation of the enzymatic activities of several polypeptide chains in prokaryotes to a single chain in eukaryotes is a theme that we will see again. The two ATP-binding casettes are in contact, but they do not interact strongly in the absence of bound ATP (Figure 13.8). On the basis of this structure and others, as well as on other experiments, a mechanism for active transport by these proteins has been developed (Figure 13.9).

1. The catalytic cycle begins with the transporter free of both ATP and substrate. The transporter can interconvert between closed and open forms.

2. Substrate enters the central cavity of the open form of the transporter from inside the cell. Substrate binding induces conformational changes in the ATP-binding cassettes that increase their affinity for ATP.

3. ATP binds to the ATP-binding cassettes, changing their conformations so that the two domains interact strongly with one another.

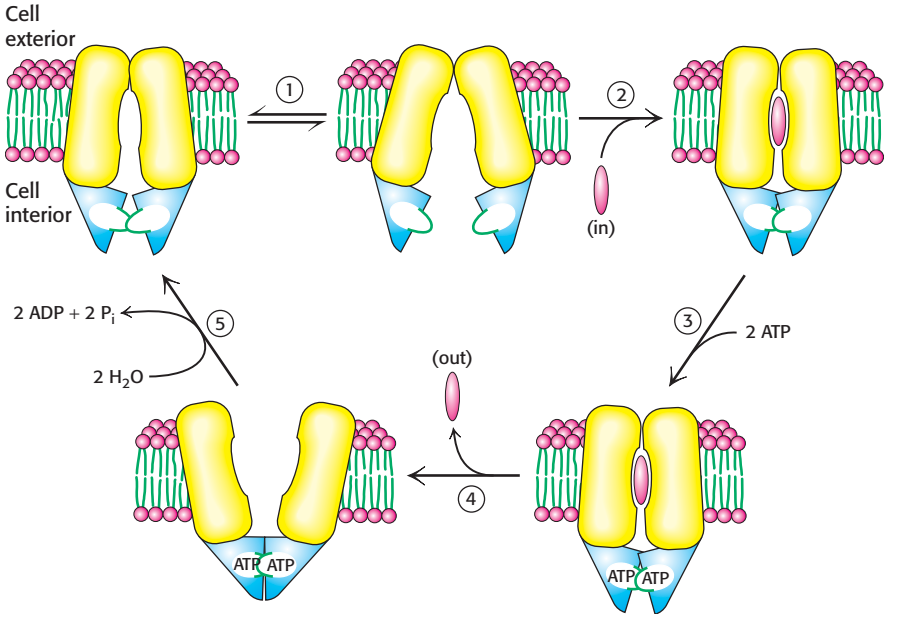


Figure 13.9 ABC transporter mechanism. The mechanism includes the following steps: (1) opening of the channel toward the inside of the cell; (2) substrate binding and conformational changes in the ATP-binding cassettes; (3) ATP binding and further conformational changes; (4) separation of the membrane-binding domains and release of the substrate to the other side of the membrane; and (5) ATP hydrolysis to reset the transporter to its initial state.

4. The strong interaction between the ATP-binding cassettes induces a change in the relation between the two membrane-spanning domains, releasing the substrate to the outside of the cell.

5. The hydrolysis of ATP and the release of ADP and inorganic phosphate reset the transporter for another cycle.

Whereas eukaryotic ABC transporters generally act to export molecules from inside the cell, prokaryotic ABC transporters often act to import specific molecules from *outside* the cell. A specific binding protein acts in concert with the bacterial ABC transporter, delivering the substrate to the transporter and stimulating ATP hydrolysis inside the cell. These binding proteins are present in the periplasm, the compartment between the two membranes that surround some bacterial cells (see Figure 12.35A).

Thus, ABC transporters use a substantially different mechanism from the P-type ATPases to couple the ATP hydrolysis reaction to conformational changes. Nonetheless, the net result is the same: the transporters are converted from one conformation capable of binding substrate from one side of the membrane to another that releases the substrate on the other side.

13.3 Lactose Permease Is an Archetype of Secondary Transporters That Use One Concentration Gradient to Power the Formation of Another

Carriers are proteins that transport ions or molecules across the membrane without hydrolysis of ATP. The mechanism of carriers involves both large conformational changes and the interaction of the protein with only a few molecules per transport cycle, limiting the maximum rate at which transport can occur. Although carriers cannot mediate primary active transport, owing to their inability to hydrolyze ATP, they can couple the thermodynamically unfavorable flow of one species of ion or molecule *up* a concentration gradient to the favorable flow of a different species *down* a concentration gradient, a process referred to as secondary active transport. Carriers that move ions or molecules "uphill" by this means are termed *secondary transporters* or *cotransporters*. These proteins can be classified as either *antiporters* or *symporters*. Antiporters couple the downhill flow of one species to the uphill flow of another in the *opposite direction* across the membrane; symporters use the flow of one species to drive the flow of a different species in the *same direction* across the membrane. *Uniporters*, another class of carriers, are able to transport a specific species in either direction governed only by concentrations of that species on either side of the membrane (Figure 13.10).

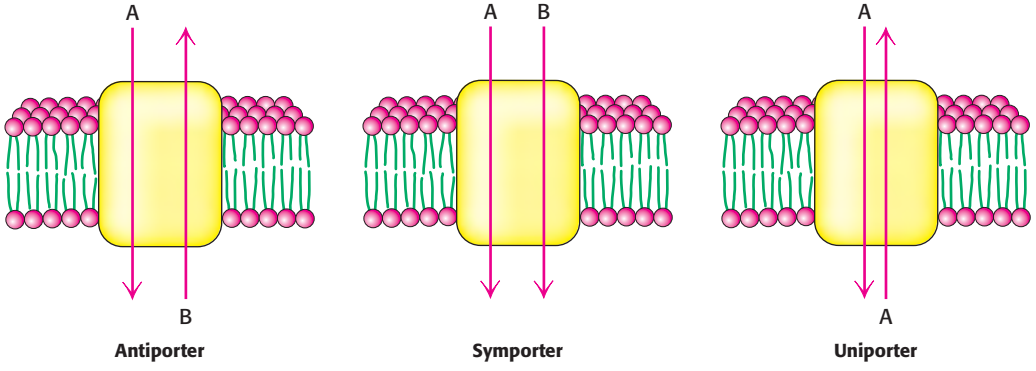


Figure 13.10 Antiporters, symporters, and uniporters. Secondary transporters can transport two substrates in opposite directions (antiporters), two substrates in the same direction (symporters), or one substrate in either direction (uniporter).

Secondary transporters are ancient molecular machines, common today in bacteria and archaea as well as in eukaryotes. For example, approximately 160 (of approximately 4000) proteins encoded by the *E. coli* genome appear to be secondary transporters. Sequence comparison and hydropathy analysis suggest that members of the largest family have 12 transmembrane helices that appear to have arisen by duplication and fusion of a membrane protein with 6 transmembrane helices. Included in this family is the *lactose permease* of *E. coli*. This symporter uses the H^+ gradient across the *E. coli* membrane (outside has higher H^+ concentration) generated by the oxidation of fuel molecules to drive the uptake of lactose and other sugars against a concentration gradient. This transporter has been extensively studied for many decades and is a useful archetype for this family.

The structure of lactose permease has been determined (Figure 13.11). As expected from the sequence analysis, this structure consists of two halves, each of which comprises six membrane-spanning α helices. Some of these helices are somewhat irregular. The two halves are well separated and are joined by a single stretch of polypeptide. In this structure, the sugar lies in a pocket in the center of the protein and is accessible from a path that leads from the interior of the cell. On the basis of these structures and a wide range of other experiments, a mechanism for symporter action has been developed. This mechanism (Figure 13.12) has many features similar to those for P-type ATPases and ABC transporters.

1. The cycle begins with the two halves oriented so that the opening to the binding pocket faces outside the cell, in a conformation different from that observed in the structures solved to date. A proton from outside the cell binds to a residue in the permease, quite possibly Glu 269.
2. In the protonated form, the permease binds lactose from outside the cell.
3. The structure everts to the form observed in the crystal structure (see Figure 13.11).
4. The permease releases lactose to the inside of the cell.
5. The permease releases a proton to the inside of the cell.
6. The permease everts to complete the cycle.

The site of protonation likely changes in the course of this cycle.

The same eversion mechanism very likely applies to all classes of secondary transporters, which appear to resemble the lactose permease in overall architecture.

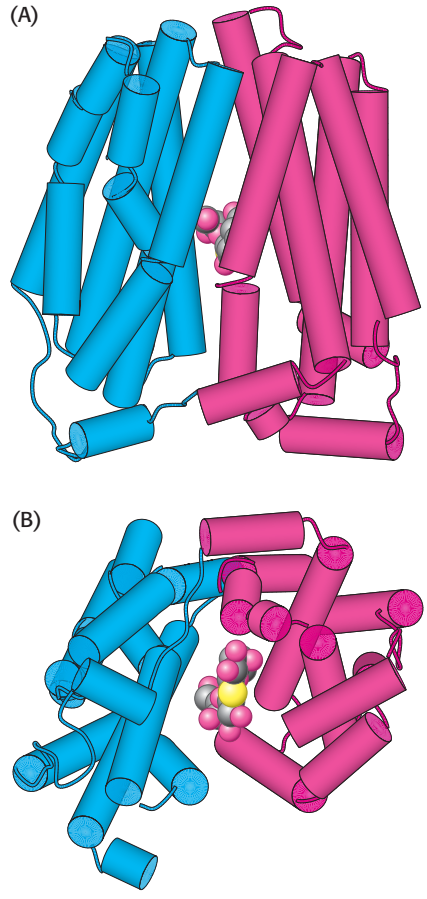


Figure 13.11 Structure of lactose permease with a bound lactose analog. The amino-terminal half of the protein is shown in blue and the carboxyl-terminal half in red. (A) Side view. (B) Bottom view (from inside the cell). *Notice* that the structure consists of two halves that surround the sugar and are linked to one another by only a single stretch of polypeptide. [Drawn from 1PV7.pdb.]

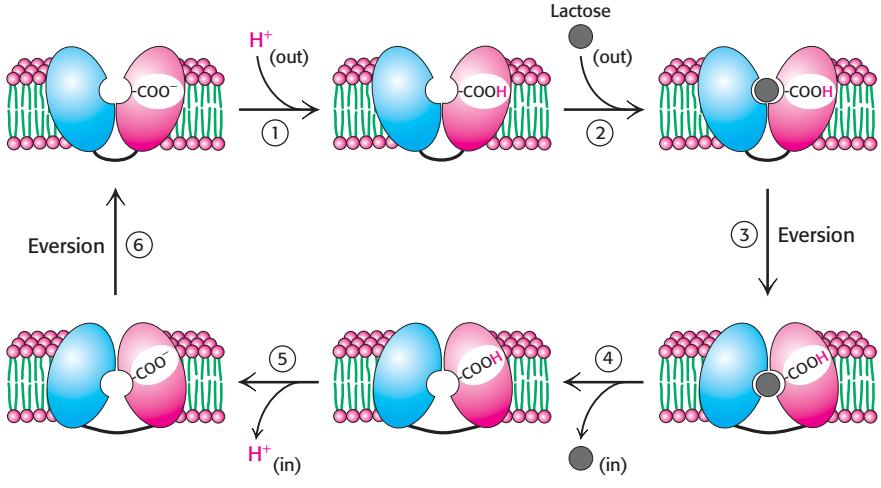


Figure 13.12 Lactose permease mechanism. The mechanism begins with the permease open to the outside of the cell (upper left). The permease binds a proton from the outside of the cell (1) and then binds its substrate (2). The permease everts (3) and then releases its substrate (4) and a proton (5) to the inside of the cell. It then everts (6) to complete the cycle.

13.4 Specific Channels Can Rapidly Transport Ions Across Membranes

Pumps and carriers can move ions across the membrane at rates approaching several thousand ions per second. Other membrane proteins, the passive-transport systems called *ion channels*, are capable of ion-transport rates that are more than 1000 times as fast. These rates of transport through ion channels are close to rates expected for ions diffusing freely through aqueous solution. Yet ion channels are not simply tubes that span membranes through which ions can rapidly flow. Instead, they are highly sophisticated molecular machines that respond to chemical and physical changes in their environments and undergo precisely timed conformational changes.

Action potentials are mediated by transient changes in Na^+ and K^+ permeability

One of the most important manifestations of ion-channel action is the nerve impulse, which is the fundamental means of communication in the nervous system. A *nerve impulse* is an electrical signal produced by the flow of ions across the plasma membrane of a neuron. The interior of a neuron, like that of most other cells, contains a high concentration of K^+ and a low concentration of Na^+. These ionic gradients are generated by the Na^+–K^+ ATPase. The cell membrane has an electrical potential determined by the ratio of the internal to the external concentration of ions. In the resting state, the membrane potential is typically −60 mV. A nerve impulse, or *action potential*, is generated when the membrane potential is depolarized beyond a critical threshold value (e.g., from −60 to −40 mV). The membrane potential becomes positive within about a millisecond and attains a value of about +30 mV before turning negative again (repolarization). This amplified depolarization is propagated along the nerve terminal (Figure 13.13).

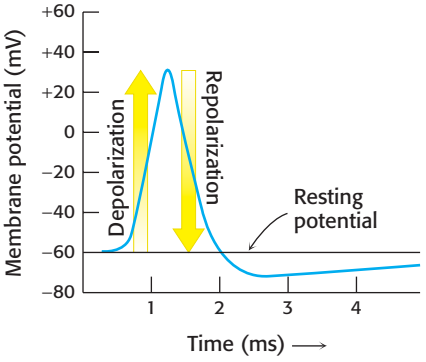


Figure 13.13 Action potential. Signals are sent along neurons by the transient depolarization and repolarization of the membrane.

Ingenious experiments carried out by Alan Hodgkin and Andrew Huxley revealed that action potentials arise from large, transient changes in the permeability of the axon membrane to Na^+ and K^+ ions. Depolarization of the membrane beyond the threshold level leads to an increase in permeability to Na^+. Sodium ions begin to flow into the cell because of the large electrochemical gradient across the plasma membrane. The entry of Na^+ further depolarizes the membrane, leading to a further increase in Na^+ permeability. This positive feedback leads to a very rapid and large change in membrane potential, from about −60 mV to +30 mV in a millisecond.

The membrane spontaneously becomes less permeable to Na^+ and more permeable to K^+. Consequently, K^+ flows outward, and so the membrane potential returns to a negative value. The resting level of −60 mV is restored in a few milliseconds as the K^+ conductance decreases to the value characteristic of the unstimulated state. The wave of depolarization followed by repolarization moves rapidly along a nerve cell. The propagation of these waves allows a touch at the tip of your toe to be detected in your brain in a few milliseconds.

This model for the action potential postulated the existence of ion channels specific for Na^+ and K^+. These channels must open in response to changes in membrane potential and then close after having remained open for a brief period of time. This bold hypothesis predicted the existence of molecules with a well-defined set of properties long before tools existed for their direct detection and characterization.

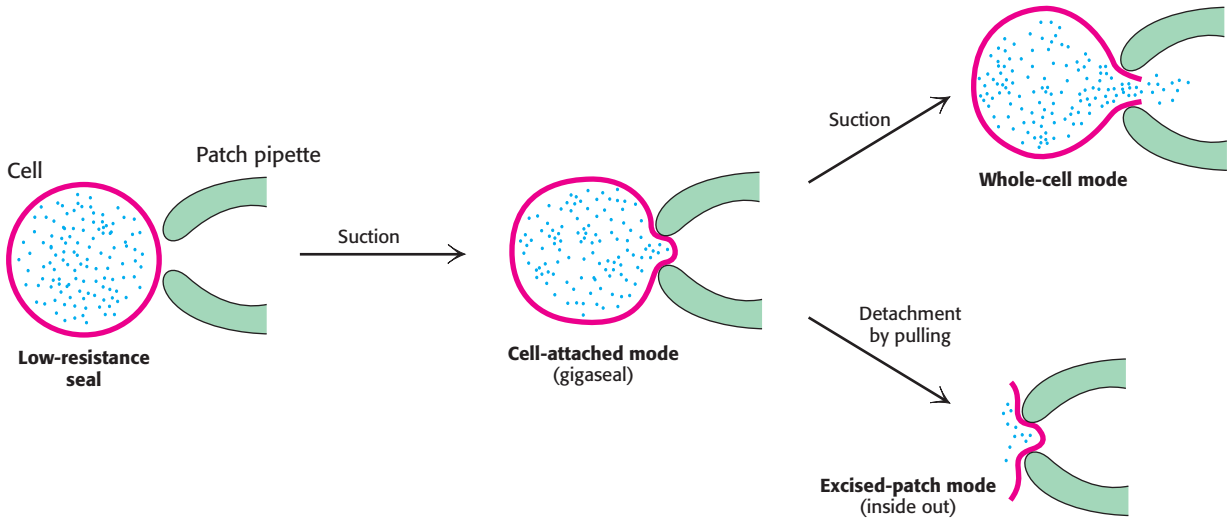


Figure 13.14 Patch-clamp modes. The patch-clamp technique for monitoring channel activity is highly versatile. A high-resistance seal (gigaseal) is formed between the pipette and a small patch of plasma membrane. This configuration is called *cell-attached mode*. The breaking of the membrane patch by increased suction produces a low-resistance pathway between the pipette and the interior of the cell. The activity of the channels in the entire plasma membrane can be monitored in this *whole-cell mode*. To prepare a membrane in the *excised-patch mode,* the pipette is pulled away from the cell. A piece of plasma membrane with its cytoplasmic side now facing the medium is monitored by the patch pipette.

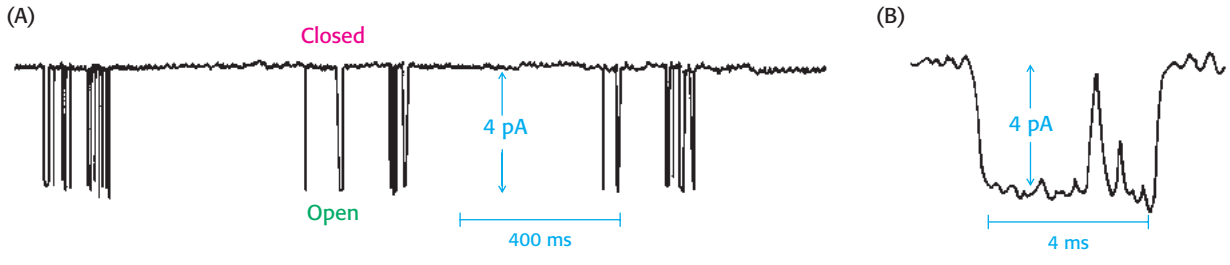
Patch-clamp conductance measurements reveal the activities of single channels

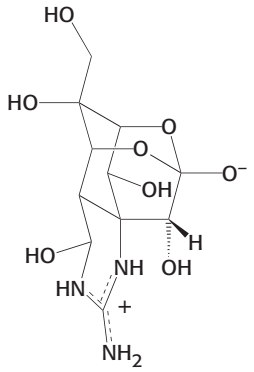
Direct evidence for the existence of these channels was provided by the *patch-clamp technique,* which was introduced by Erwin Neher and Bert Sakmann in 1976. This powerful technique enables the measurement of the ion conductance through a small patch of cell membrane. In this technique, a clean glass pipette with a tip diameter of about 1 μm is pressed against an intact cell to form a seal (Figure 13.14). Slight suction leads to the formation of a very tight seal so that the resistance between the inside of the pipette and the bathing solution is many gigaohms (1 gigaohm is equal to 10^9 ohms). Thus, a gigaohm seal (called a *gigaseal*) ensures that an electric current flowing through the pipette is identical with the current flowing through the membrane covered by the pipette. The gigaseal makes possible high-resolution current measurements while a known voltage is applied across the membrane. Remarkably, the flow of ions through a single channel and transitions between the open and the closed states of a channel can be monitored with a time resolution of microseconds (Figure 13.15). Furthermore, the activity of a channel in its native membrane environment, even in an intact cell, can be directly observed. Patch-clamp methods provided one of the first views of single biomolecules in action. Subsequently, other methods for observing single molecules were invented, opening new vistas on biochemistry at its most fundamental level.

The structure of a potassium ion channel is an archetype for many ion-channel structures

With the existence of ion channels firmly established by patch-clamp methods, scientists sought to identify the molecules that form ion channels. The Na^+ channel was first purified from the electric organ of electric eel,

Figure 13.15 Observing single channels. (A) The results of a patch-clamp experiment revealing a single ion channel undergoing transitions between closed and open states. (B) Closer inspection of the trace in (A) reveals the length of time the channel is in the open state.





Tetrodotoxin

which is a rich source of the protein forming this channel. The channel was purified on the basis of its ability to bind tetrodotoxin, a neurotoxin from the puffer fish that binds to Na^+ channels very tightly ($K_i \approx 1$ nM). The lethal dose of this poison for an adult human being is about 10 ng.

The isolated Na^+ channel is a single 260-kd chain. Cloning and sequencing of cDNAs encoding Na^+ channels revealed that the channel contains four internal repeats, each having a similar amino acid sequence, suggesting that gene duplication and divergence have produced the gene for this channel. Hydrophobicity profiles indicate that each repeat contains five hydrophobic segments (S1, S2, S3, S5, and S6). Each repeat also contains a highly positively charged S4 segment; positively charged arginine or lysine residues are present at nearly every third residue. It was proposed that segments S1 through S6 are membrane-spanning α helices. The positively charged residues in S4 were proposed to act as the voltage sensors of the channel.

The purification of K^+ channels proved to be much more difficult because of their low abundance and the lack of known high-affinity ligands comparable to tetrodotoxin. The breakthrough came in studies of mutant fruit flies that shake violently when anesthetized with ether. The mapping and cloning of the gene, termed *shaker,* responsible for this defect revealed the amino acid sequence encoded by a K^+-channel gene. The *shaker* gene encodes a 70-kd protein that contains sequences corresponding to segments S1 through S6 in one of the repeated units of the Na^+ channel. Thus, a K^+-channel subunit is homologous to one of the repeated units of Na^+ channels. Consistent with this homology, four Shaker polypeptides come together to form a functional channel. More recently, bacterial K^+ channels were discovered that contain only the two membrane-spanning regions corresponding to segments S5 and S6. This and other information suggested that S5 and S6, including the region between them, form the actual pore in the K^+ channel. Segments S1 through S4 contain the apparatus that opens the pore. The sequence relations between these ion channels are summarized in Figure 13.16.

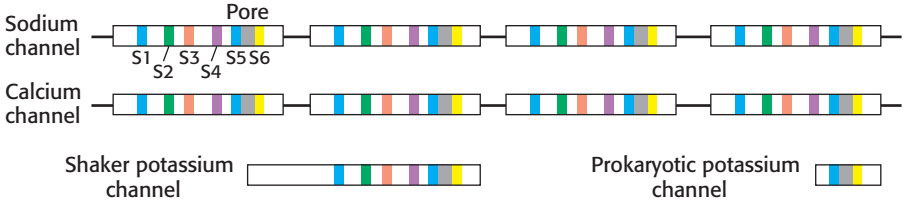


Figure 13.16 Sequence relations of ion channels. Like colors indicate structurally similar regions of the sodium, calcium, and potassium channels. Each of these channels exhibits approximate fourfold symmetry, either within one chain (sodium, calcium channels) or by forming tetramers (potassium channels).

In 1998, Roderick MacKinnon and coworkers determined the structure of a K^+ channel from the bacterium *Streptomyces lividans* by x-ray crystallography. This channel contains only the pore-forming segments S5 and S6. As expected, the K^+ channel is a tetramer of identical subunits, each of which includes two membrane-spanning α helices (Figure 13.17). The four subunits come together to form a pore in the shape of a cone that runs through the center of the structure.

The structure of the potassium ion channel reveals the basis of ion specificity

The structure presented in Figure 13.17 probably represents the K^+ channel in a closed form. Nonetheless, it suggests how the channel is able to exclude all but K^+ ions. Beginning from the inside of the cell, the pore starts with a diameter of approximately 10 Å and then constricts to a smaller

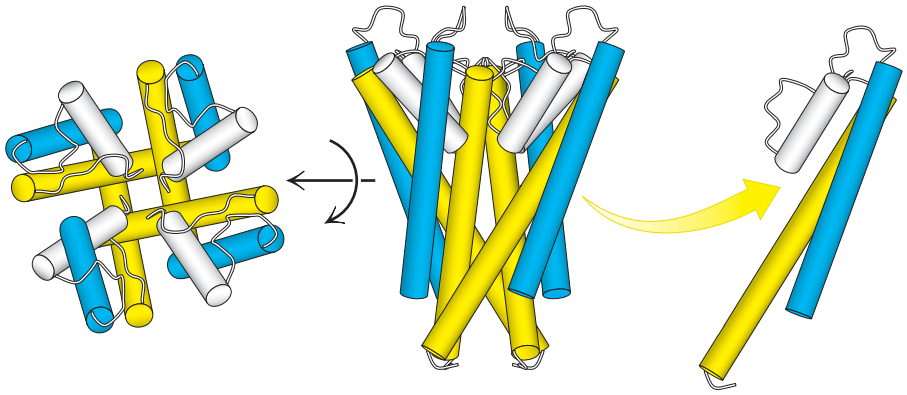


Figure 13.17 Structure of the potassium ion channel. The K^+ channel, composed of four identical subunits, is cone shaped, with the larger opening facing the inside of the cell (center). A view down the pore, looking toward the outside of the cell, shows the relations of the individual subunits (left). One of the four identical subunits of the pore is illustrated at the right, with the pore-forming region shown in gray. [Drawn from 1K4C.pdb.]

cavity with a diameter of 8 Å. Both the opening to the outside and the central cavity of the pore are filled with water, and a K^+ ion can fit in the pore without losing its shell of bound water molecules. Approximately two-thirds of the way through the membrane, the pore becomes more constricted (3-Å diameter). At that point, any K^+ ions must give up their water molecules and interact directly with groups from the protein. The channel structure effectively reduces the thickness of the membrane from 34 Å to 12 Å by allowing the solvated ions to penetrate into the membrane before the ions must directly interact with the channel (Figure 13.18).

For K^+ ions to relinquish their water molecules, other polar interactions must replace those with water. The restricted part of the pore is built from residues contributed by the two transmembrane α helices. In particular, a five-amino-acid stretch within this region functions as the *selectivity filter* that determines the preference for K^+ over other ions (Figure 13.19). The stretch has the sequence Thr-Val-Gly-Tyr-Gly (TVGYG), and is nearly completely conserved in all K^+ channels. The region of the strand containing the conserved sequence lies in an extended conformation and is oriented such that the peptide carbonyl

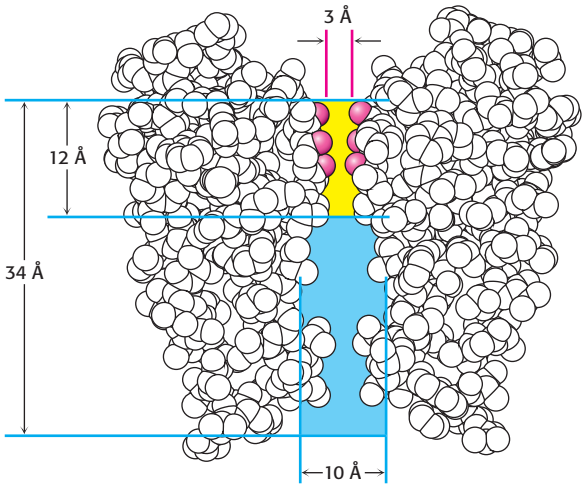


Figure 13.18 Path through a channel. A potassium ion entering the K^+ channel can pass a distance of 22 Å into the membrane while remaining solvated with water (blue). At this point, the pore diameter narrows to 3 Å (yellow), and potassium ions must shed their water and interact with carbonyl groups (red) of the pore amino acids.

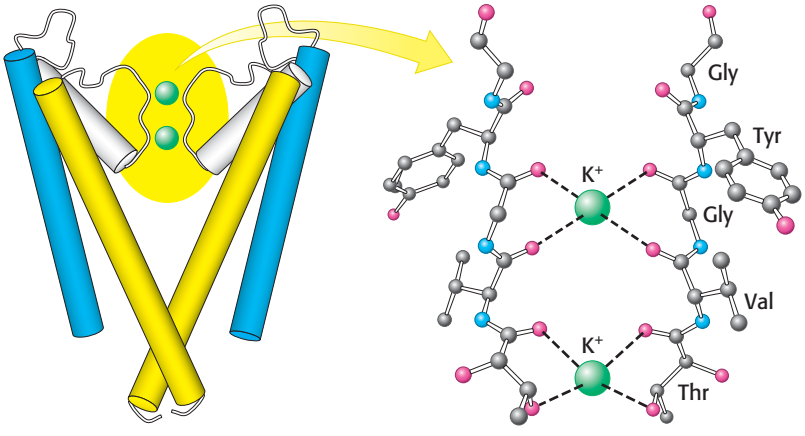


Figure 13.19 Selectivity filter of the potassium ion channel. Potassium ions interact with the carbonyl groups of the TVGYG sequence of the selectivity filter, located at the 3-Å-diameter pore of the K^+ channel. Only two of the four channel subunits are shown.

groups are directed into the channel, in good position to interact with the potassium ions.

Potassium ion channels are 100-fold more permeable to K^+ than to Na^+. How is this high degree of selectivity achieved? Ions having a radius larger than 1.5 Å cannot pass into the narrow diameter (3 Å) of the selectivity filter of the K^+ channel. However, a bare Na^+ is small enough (Table 13.1) to pass through the pore. Indeed, the ionic radius of Na^+ is substantially smaller than that of K^+. How then is Na^+ rejected?

Table 13.1 Properties of alkali cations

Ion	Ionic radius (Å)	Hydration free energy in kJ mol^{-1} (kcal mol^{-1})
Li^+	0.60	−410 (−98)
Na^+	0.95	−301 (−72)
K^+	1.33	−230 (−55)
Rb^+	1.48	−213 (−51)
Cs^+	1.69	−197 (−47)

The key point is that the free-energy costs of dehydrating these ions are considerable [Na^+, 301 kJ mol^{-1} (72 kcal mol^{-1}), and K^+, 230 kJ mol^{-1} (55 kcal mol^{-1})]. *The channel pays the cost of dehydrating K^+ by providing compensating interactions with the carbonyl oxygen atoms lining the selectivity filter.* However, these oxygen atoms are positioned such that they do not interact favorably with Na^+, because the ion is too small (Figure 13.20). Sodium ions are rejected because the higher cost of dehydrating them would be unrecovered. The potassium ion channel avoids closely embracing sodium ions, which must stay hydrated and hence cannot pass through the channel.

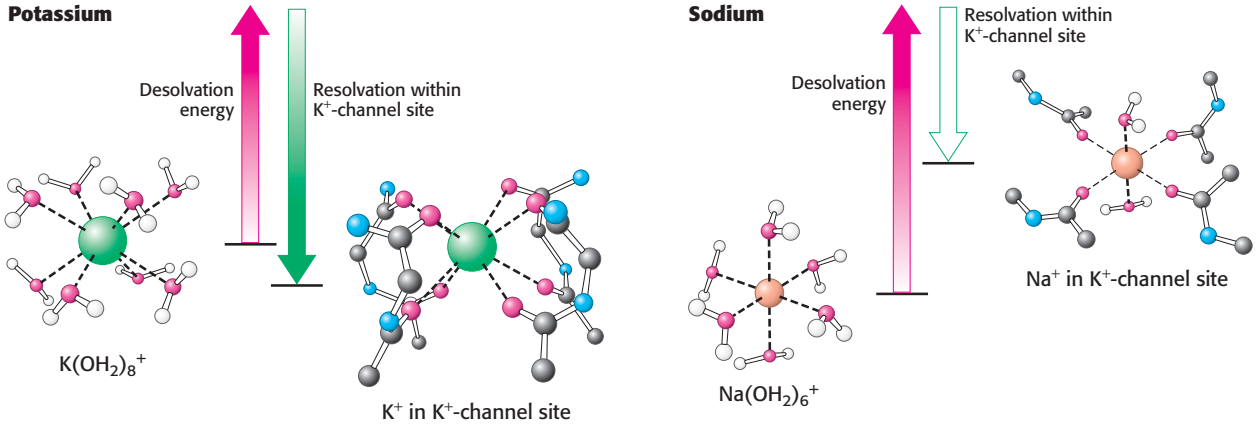


Figure 13.20 Energetic basis of ion selectivity. The energy cost of dehydrating a potassium ion is compensated by favorable interactions with the selectivity filter. Because a sodium ion is too small to interact favorably with the selectivity filter, the free energy of desolvation cannot be compensated and the sodium ion does not pass through the channel.

The K^+ channel structure enables a clearer understanding of the structure and function of Na^+ and Ca^{2+} channels because of their homology to K^+ channels. Sequence comparisons and the results of mutagenesis experiments have implicated the region between segments S5 and S6 in ion selectivity in the Ca^{2+} channel. In Ca^{2+} channels, one glutamate residue of this region in each of the four repeated units plays a major role in determining ion selectivity. Residues in the positions corresponding to the glutamate residues in Ca^{2+} channels are major components of the selectivity filter of the Na^+ channel. These residues—aspartate, glutamate, lysine, and alanine—are located in each of the internal repeats of the Na^+ channel, forming a region termed the DEKA locus. Thus, the potential fourfold symmetry of the channel is clearly broken in this region, which explains why Na^+ channels consist of a single large polypeptide chain rather than a noncovalent assembly of four identical subunits. The preference of the Na^+ channel for Na^+ over K^+ depends on ionic radius; the diameter of the pore determined by these residues and others is sufficiently restricted that small ions such as Na^+ and Li^+ can pass through the channel, but larger ions such as K^+ are significantly hindered.

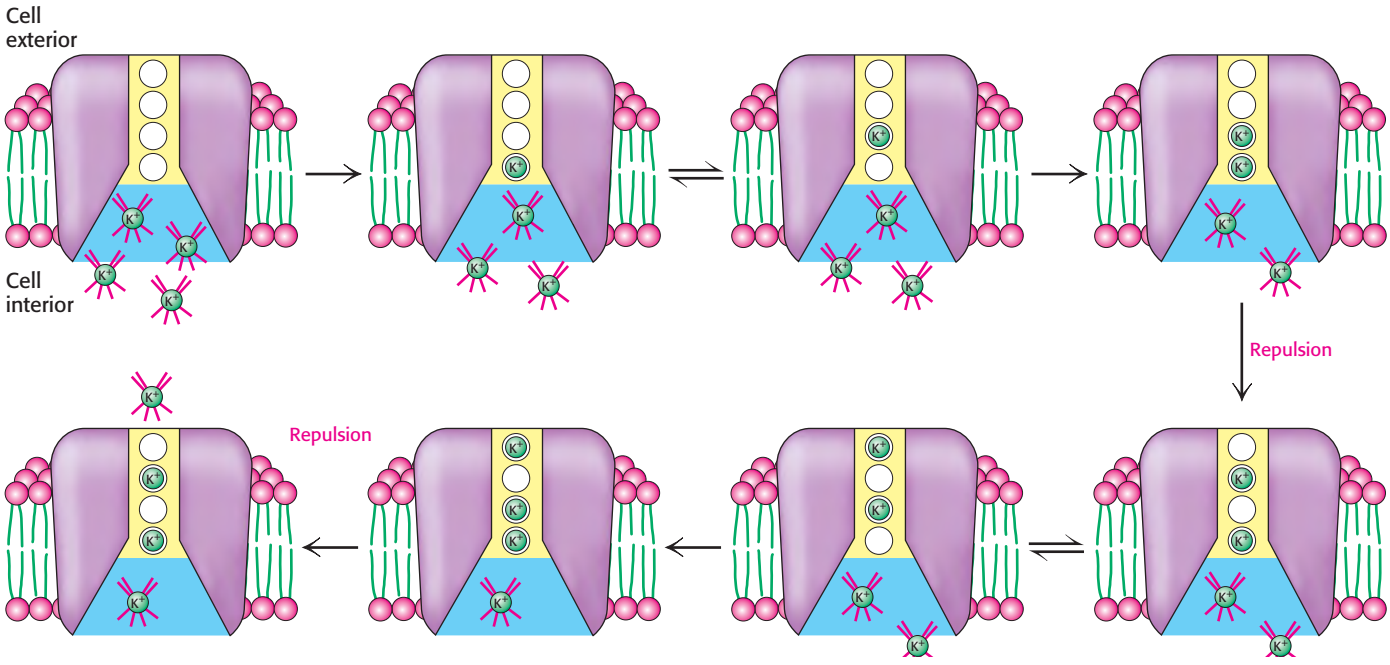


Figure 13.21 Model for K^+-channel ion transport. The selectivity filter has four binding sites. Hydrated potassium ions can enter these sites, one at a time, losing their hydration shells. When two ions occupy adjacent sites, electrostatic repulsion forces them apart. Thus, as ions enter the channel from one side, other ions are pushed out the other side.

The structure of the potassium ion channel explains its rapid rate of transport

The tight binding sites required for ion selectivity should slow the progress of ions through a channel, yet ion channels achieve rapid rates of ion transport. How is this paradox resolved? A structural analysis of the K^+ channel at high resolution provides an appealing explanation. Four K^+-binding sites crucial for rapid ion flow are present in the constricted region of the K^+ channel. Consider the process of ion conductance starting from inside the cell (Figure 13.21). A hydrated potassium ion proceeds into the channel and through the relatively unrestricted part of the channel. The ion then gives up its coordinated water molecules and binds to a site within the selectivity-filter region. The ion can move between the four sites within the selectivity filter because they have similar ion affinities. As each subsequent potassium ion moves into the selectivity filter, its positive charge will repel the potassium ion at the nearest site, causing it to shift to a site farther up the channel and in turn push upward any potassium ion already bound to a site farther up. Thus, each ion that binds anew favors the release of an ion from the other side of the channel. This multiple-binding-site mechanism solves the paradox of high ion selectivity and rapid flow.

Voltage gating requires substantial conformational changes in specific ion-channel domains

Some Na^+ and K^+ channels are gated by membrane potential; that is, they change conformation to a highly conducting form in response to changes in voltage across the membrane. As already noted, these *voltage-gated channels* include segments S1 through S4 in addition to the pore itself formed by S5 and S6. The structure of a voltage-gated K^+ channel from *Aeropyrum pernix* has been determined by x-ray crystallography (Figure 13.22). The segments S1 through S4 form domains, termed "paddles," that extend from the core of the channel. These paddles include the segment S4, the voltage sensor itself. Segment S4 forms an α helix lined with positively charged residues. In contrast with expectations, segments S1 through S4 are not enclosed within the protein but, instead, are positioned to lie in the membrane itself.

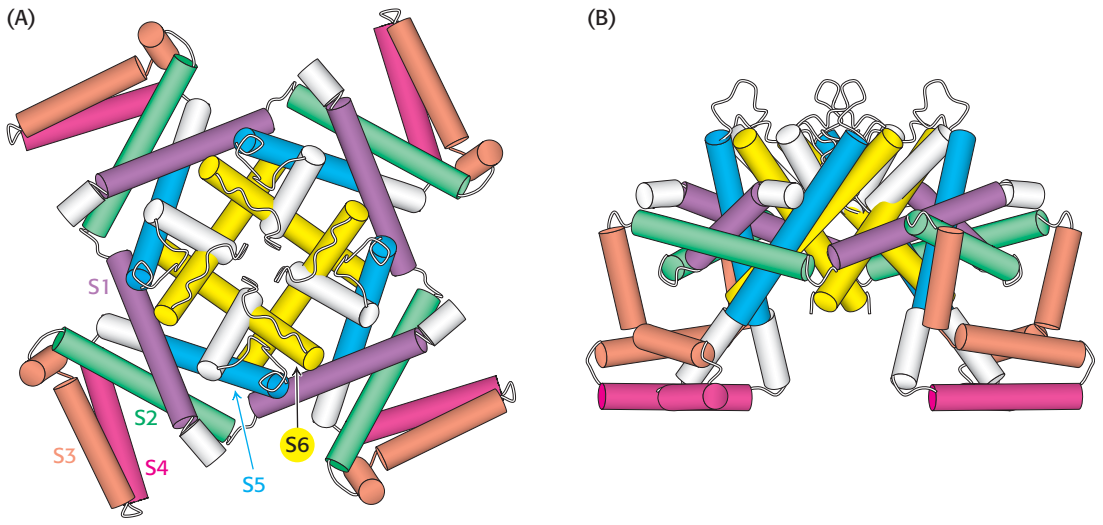


Figure 13.22 Structure of a voltage-gated potassium channel. (A) A view looking down through the pore. (B) A side view. *Notice* that the positively charged S4 region (red) lies on the outside of the structure at the bottom of the pore. [Drawn from 1ORQ.pdb.]

A model for voltage gating has been proposed by Roderick MacKinnon and coworkers on the basis of this structure and a range of other experiments (Figure 13.23). In the closed state, the paddles lie in a "down" position. On membrane depolarization, the cytoplasmic side of the membrane becomes more positively charged, and the paddles are pulled through the membrane into an "up" position. In this position, they pull the four sides of the base on the pore apart, increasing access to the selectivity filter and opening the channel.

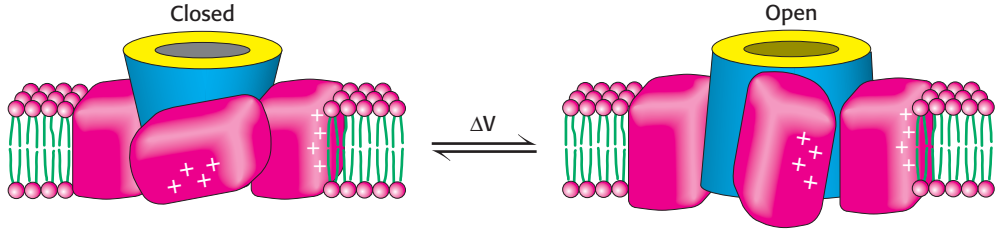


Figure 13.23 A model for voltage gating of ion channels. The voltage-sensing paddles lie in the "down" position below the closed channel (left). Membrane depolarization pulls these paddles through the membrane. The motion pulls the base of the channel apart, opening the channel (right).

A channel can be inactivated by occlusion of the pore: the ball-and-chain model

The K^+ channel and the Na^+ channel undergo inactivation within milliseconds of opening (Figure 13.24). A first clue to the mechanism of inactivation came from exposing the cytoplasmic side of either channel to trypsin; cleavage by trypsin produced trimmed channels that stayed persistently open after depolarization. Furthermore, a mutant Shaker channel lacking 42 amino acids near the amino terminus opened in response to depolarization but did not inactivate. Remarkably, inactivation was restored by adding a synthetic peptide corresponding to the first 20 residues of the native channel.

These experiments strongly support the *ball-and-chain model* for channel inactivation that had been proposed years earlier (Figure 13.25). According to this model, the first 20 residues of the K^+ channel form a cytoplasmic unit (the *ball*) that is attached to a flexible segment of the

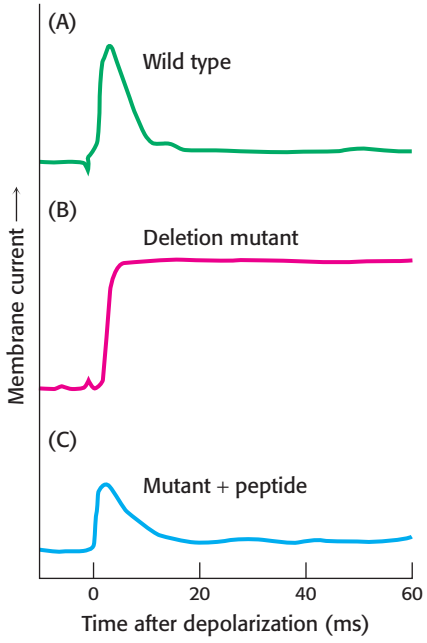
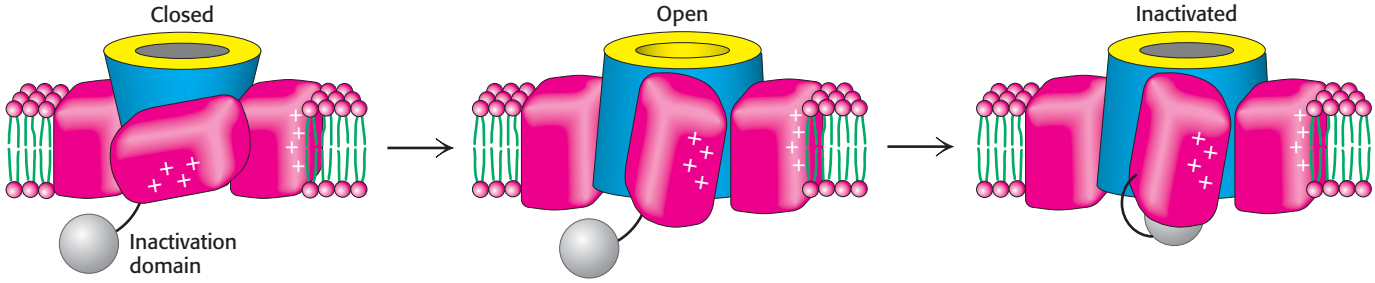


Figure 13.24 Inactivation of the potassium ion channel. The amino-terminal region of the K^+ chain is critical for inactivation. (A) The wild-type Shaker K^+ channel displays rapid inactivation after opening. (B) A mutant channel lacking residues 6 through 46 does not inactivate. (C) Inactivation can be restored by adding a peptide consisting of residues 1 through 20 at a concentration of 100 μM. [After W. N. Zagotta, T. Hoshi, and R. W. Aldrich. *Science* 250(1990):568–571.]

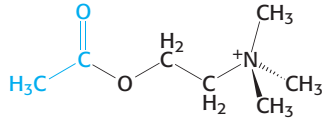
polypeptide (the *chain*). When the channel is closed, the ball rotates freely in the aqueous solution. When the channel opens, the ball quickly finds a complementary site in the open pore and occludes it. Hence, the channel opens for only a brief interval before it undergoes inactivation by occlusion. Shortening the chain speeds inactivation because the ball finds its target more quickly. Conversely, lengthening the chain slows inactivation. Thus, the duration of the open state can be controlled by the length and flexibility of the tether. In some senses, the "ball" domains, which include substantial regions of positive charge, can be thought of as large, tethered cations that are pulled into the open channel but get stuck and block further ion conductance.

Figure 13.25 Ball-and-chain model for channel inactivation. The inactivation domain, or "ball" (gray), is tethered to the channel by a flexible "chain." In the closed state, the ball is located in the cytoplasm. Depolarization opens the channel and creates a binding site for the positively charged ball in the mouth of the pore. Movement of the ball into this site inactivates the channel by occluding it.



The acetylcholine receptor is an archetype for ligand-gated ion channels

Nerve impulses are communicated across synapses by small, diffusible molecules called *neurotransmitters*. One neurotransmitter is *acetylcholine*. The presynaptic membrane of a synapse is separated from the postsynaptic membrane by a gap of about 50 nm called the *synaptic cleft*. The arrival of a nerve impulse at the end of an axon leads to the synchronous export of the contents of some 300 vesicles of acetylcholine into the cleft (Figure 13.26). The binding of acetylcholine to the postsynaptic membrane markedly changes its ionic permeability, triggering an action potential. Acetylcholine opens a single kind of cation channel, called the *acetylcholine receptor,* which is almost equally permeable to Na^+ and to K^+.



Acetylcholine

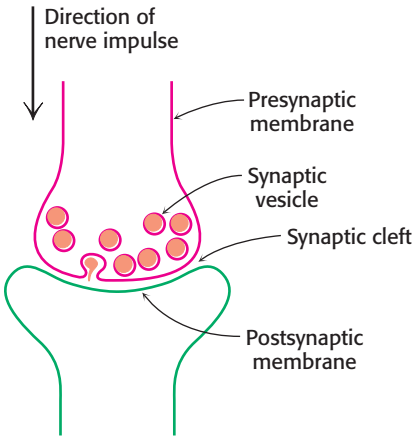


Figure 13.26 Schematic representation of a synapse.

The torpedo (*Torpedo marmorata,* also known as the electric ray) has an electric organ, rich in acetylcholine receptors, that can deliver a shock of as much as 200 V for approximately 1 s. [Yves Gladu/Jacana/Photo Researchers.]

The acetylcholine receptor is the best-understood *ligand-gated channel*. This type of channel is gated not by voltage but by the presence of specific ligands. The binding of acetylcholine to the channel is followed by its transient opening. The electric organ of *Torpedo marmorata*, an electric ray, is a choice source of acetylcholine receptors for study because its electroplaxes (voltage-generating cells) are very rich in postsynaptic membranes that respond to this neurotransmitter. The receptor is very densely packed in these membranes ($\sim$20,000 μm^{-2}). The acetylcholine receptor of the electric organ has been solubilized by adding a nonionic detergent to a postsynaptic membrane preparation and purified by affinity chromatography on a column bearing covalently attached cobratoxin, a small protein toxin from snakes that has a high affinity for acetylcholine receptors. With the use of techniques presented in Chapter 3, the 268-kd receptor was identified as a pentamer of four kinds of membrane-spanning subunits—α_2, β, γ, and δ—arranged in the form of a ring that creates a pore through the membrane.

The cloning and sequencing of the cDNAs for the four kinds of subunits (50–58 kd) showed that they have clearly similar sequences; the genes for the α, β, γ, and δ subunits arose by duplication and divergence of a common ancestral gene. Each subunit has a large extracellular domain, followed at the carboxyl end by four predominantly hydrophobic segments that span the bilayer membrane. Acetylcholine binds at the α–γ and α–δ interfaces. Electron microscopic studies of purified acetylcholine receptors demonstrated that the structure has approximate fivefold symmetry, in harmony with the similarity of its five constituent subunits (Figure 13.27).

What is the basis of channel opening? A comparison of the structures of the closed and open forms of the channel would be highly revealing, but such comparisons have been difficult to obtain. Cryoelectron micrographs indicate that the binding of acetylcholine to the extracellular domain causes a structural alteration that initiates rotations of the α-helical rods lining the membrane-spanning pore. The amino acid sequences of these helices point to the presence of alternating ridges of small polar or neutral residues

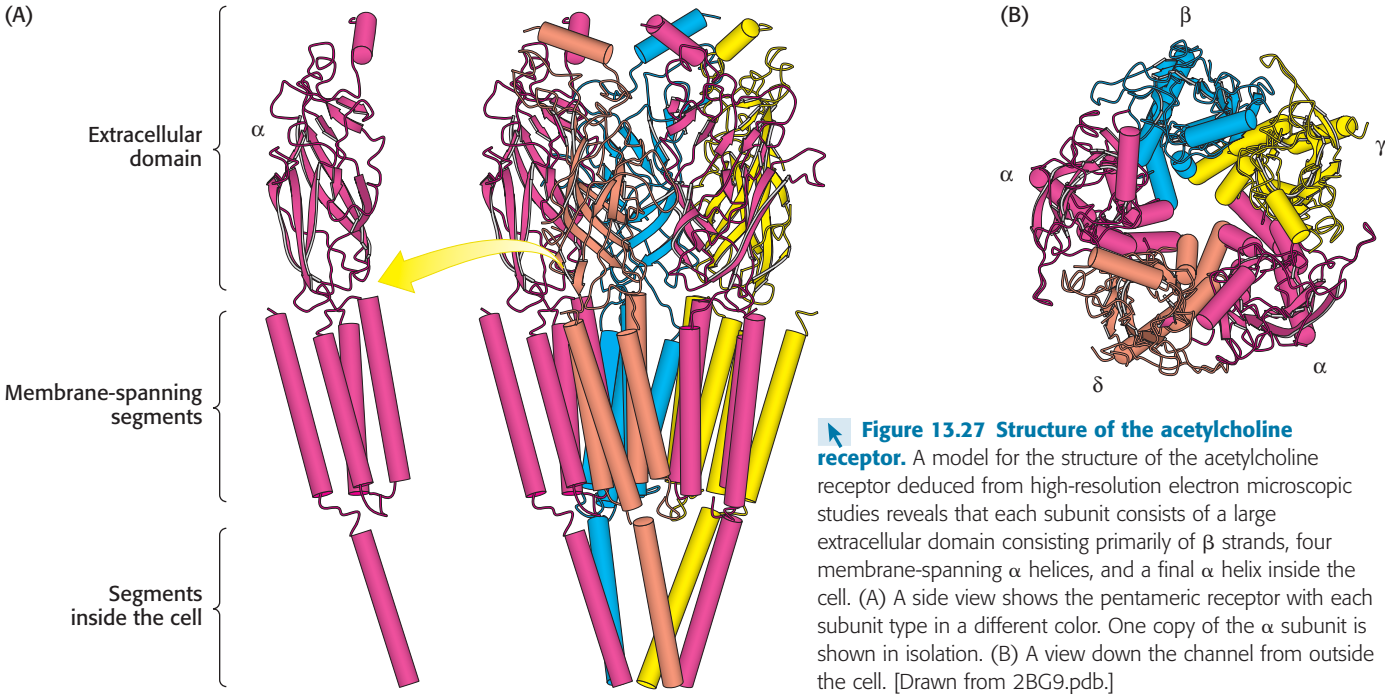


Figure 13.27 Structure of the acetylcholine receptor. A model for the structure of the acetylcholine receptor deduced from high-resolution electron microscopic studies reveals that each subunit consists of a large extracellular domain consisting primarily of β strands, four membrane-spanning α helices, and a final α helix inside the cell. (A) A side view shows the pentameric receptor with each subunit type in a different color. One copy of the α subunit is shown in isolation. (B) A view down the channel from outside the cell. [Drawn from 2BG9.pdb.]

(serine, threonine, glycine) and large nonpolar ones (isoleucine, leucine, phenylalanine). In the closed state, the large residues may occlude the channel by forming a tight hydrophobic ring (Figure 13.28). Indeed, each subunit has a bulky leucine residue at a critical position. The binding of acetylcholine could allosterically rotate the membrane-spanning helices so that the pore would be lined by small polar residues rather than by large hydrophobic ones. The wider, more polar pore would then be open to the passage of Na^+ and K^+ ions.

Action potentials integrate the activities of several ion channels working in concert

To see how ligand-gated and ion-gated channels work together to generate a sophisticated physiological response, we now revisit the action potential introduced at the beginning of this section. First, we need to introduce the concept of *equilibrium potential*. Suppose that a membrane separates two solutions that contain different concentrations of some cation X^+ (Figure 13.29). Let $[X^+]_{in}$ be the concentration of X^+ on one side of the membrane (corresponding to the inside of a cell) and $[X^+]_{out}$ be the concentration of X^+ on the other side (corresponding to the outside of a cell). Suppose that an ion channel opens that allows X^+ to move across the membrane. What will happen? It seems clear that X^+ will move through the channel from the side with the higher concentration to the side with the lower concentration. However, positive charges will start to accumulate on the side with the lower concentration, making it more difficult to move each additional positively charged ion. An equilibrium will be achieved when the driving force due to the concentration gradient is balanced by the electrostatic force resisting the motion of an additional charge. In these circumstances, the membrane potential is given by the *Nernst equation:*

$$V_{eq} = -(RT/zF) \ln ([X]_{in}/[X]_{out})$$

where R is the gas constant and F is the Faraday constant (96.5 kJ V^{-1} mol^{-1}, or 23.1 kcal V^{-1} mol^{-1}) and z is the charge on the ion X (e.g., +1 for X^+).

The membrane potential at equilibrium is called the equilibrium potential for a given ion at a given concentration ratio across a membrane. For sodium with $[Na^+]_{in}$ = 14 mM and $[Na^+]_{out}$ = 143 mM, the equilibrium potential is +62 mV at 37°C. Similarly, for potassium with $[K^+]_{in}$ = 157 mM and $[K^+]_{out}$ = 4 mM, the equilibrium potential is −98 mV. In the absence

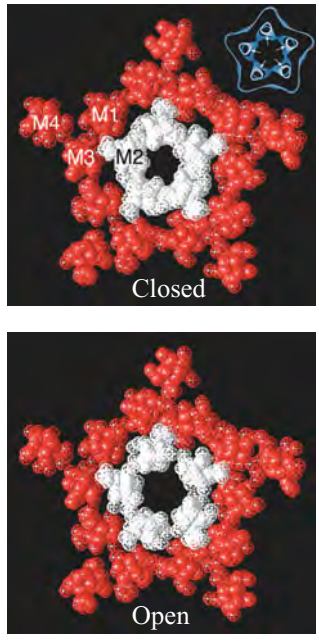


Figure 13.28 Opening the acetylcholine receptor. Cross sections from electron microscopic reconstructions of the acetylcholine receptor in (top) its closed form and (bottom) its open form. (The open form corresponds to the structure shown in Figure 13.27). The areas labeled M1, M2, M3, and M4 correspond to the four membrane-spanning α helices of one subunit. The cross section of the open channel was generated by treating the receptor with acetylcholine and freezing the sample within 20 ms. *Notice* that the hole in the center of the channel is substantially larger in the open structure. The enlargement of the hole is due to the rotation of the M2 helices by approximately 15 degrees along their long axes. [Courtesy of Nigel Unwin.]

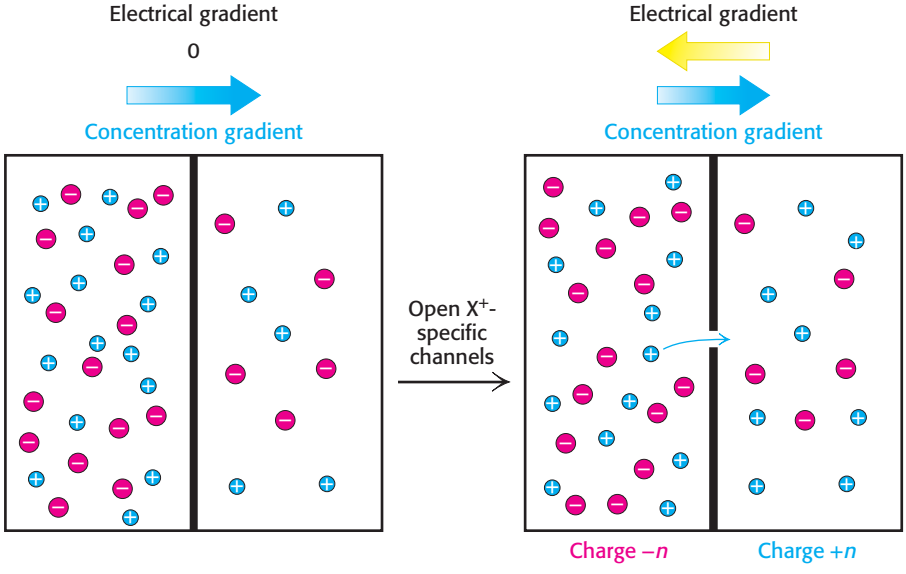


Figure 13.29 Equilibrium potential. The membrane potential reaches an equilibrium when the driving force due to the concentration gradient is exactly balanced by the opposing force due to the repulsion of like charges.

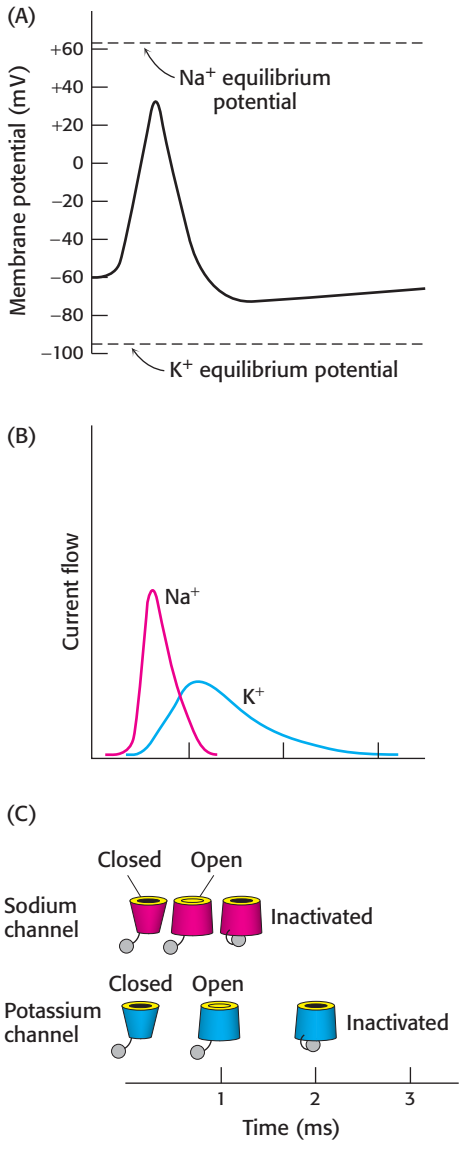


Figure 13.30 Action-potential mechanism. (A) On the initation of an action potential, the membrane potential moves from the resting potential upward toward the Na^+ equilibrium potential and then downward toward the K^+ equilibrium potential. (B) The currents through the Na^+ and K^+ channels underlying the action potential. (C) The states of the Na^+ and K^+ channels during the action potential.

of stimulation, the resting potential for a typical neuron is −60 mV. This value is close to the equilibrium potential for K^+ owing to the fact that a small number of K^+ channels are open.

We are now prepared to consider what happens in the generation of an action potential (Figure 13.30). Initially, a neurotransmitter such as acetylcholine is released from an adjacent cell. The released acetylcholine binds to the acetylcholine receptor, causing it to open within less than a millisecond. The acetylcholine receptor is a nonspecific cation channel. Sodium ions flow into the cell and potassium ions flow out of the cell. Without any further events, the membrane potential would move to a value corresponding to the average of the equilibrium potentials for Na^+ and K^+, approximately −20 mV. However, as the membrane potential approaches −40 mV, the voltage-sensing paddles of Na^+ channels are pulled into the membrane, opening the Na^+ channels. With these channels open, sodium ions flow rapidly into the cell and the membrane potential rises rapidly toward the Na^+ equilibrium potential. The voltage-sensing paddles of K^+ channels also are pulled into the membrane by the changed membrane potential, but more slowly than Na^+ channel paddles. Nonetheless, after approximately 1 ms, many K^+ channels start to open. At the same time, inactivation "ball" domains plug the open Na^+ channels, decreasing the Na^+ current. The acetylcholine receptors that initiated these events are also inactivated on this time scale. With the Na^+ channels inactivated and only the K^+ channels open, the membrane potential drops rapidly toward the K^+ equilibrium potential. The open K^+ channels are susceptible to inactivation by their "ball" domains, and these K^+ currents, too, are blocked. With the membrane potential returned to close to its initial value, the inactivation domains are released and the channels return to their original closed states. These events propagate along the neuron as the depolarization of the membrane opens channels in nearby patches of membrane.

How much current actually flows across the membrane over the course of an action potential? This question can be addressed from two complementary directions. First, a typical nerve cell contains 100 Na^+ channels per square micrometer. At a membrane potential of +20 mV, each channel conducts 10^7 ions per second. Thus, in a period of 1 millisecond, approximately 10^5 ions flow through each square micrometer of membrane surface. Assuming a cell volume of 10^4 μm^3 and a surface area of 10^4 μm^2, this rate of ion flow corresponds to an increase in the Na^+ concentration of less than 1%. How can this be? A robust action potential is generated because the membrane potential is very sensitive to even a slight change in the distribution of charge. This sensitivity makes the action potential a very efficient means of signaling over long distances and with rapid repetition rates.

Disruption of ion channels by mutations or chemicals can be potentially life threatening

The generation of an action potential requires the precise coordination of gating events of a collection of ion channels. Perturbation of this timing can have devastating effects. For example, the rhythmic generation of action potentials by the heart is absolutely essential to maintain delivery of oxygenated blood to the peripheral tissues. *Long QT syndrome* (LQTS) is a genetic disorder in which the recovery of the action potential from its peak potential to the resting equilibrium potential is delayed. The term "QT" refers to a specific feature of the cardiac electrical activity pattern as measured by electrocardiography. LQTS can lead to brief losses of consciousness (syncope), disruption of normal cardiac rhythm

(arrhythmia), and sudden death. The most common mutations identified in LQTS patients inactivate K^+ channels or prevent the proper trafficking of these channels to the plasma membrane. The resulting loss in potassium permeability slows the repolarization of the membrane and delays the induction of the subsequent cardiac contraction, rendering the cardiac tissue susceptible to arrhythmias.

Prolongation of the cardiac action potential in this manner can also be induced by a number of therapeutic drugs. In particular, the K^+ channel hERG (for human *ether-a-go-go*-related gene, named for its ortholog in *Drosophila melanogaster*) is highly susceptible to interactions with certain drugs. The hydrophobic regions of these drugs can block hERG by binding to two nonconserved aromatic residues on the internal surface of the channel cavity. In addition, this cavity is predicted to be wider than other K^+ channels because of the absence of a conserved Pro-X-Pro motif within the S6 hydrophobic segment. Inhibition of hERG by these drugs can lead to an increased risk of cardiac arrhythmias and sudden death. Accordingly, a number of these agents, such as the antihistamine terfenadine, have been withdrawn from the market. Screening for the inhibition of hERG is now a critical safety hurdle for the pharmaceutical advancement of a molecule to an approved drug.

13.5 Gap Junctions Allow Ions and Small Molecules to Flow Between Communicating Cells

The ion channels that we have considered thus far have narrow pores and are moderately to highly selective in the ions that they allow to pass through them. They are closed in the resting state and have short lifetimes in the open state, typically a millisecond, that enable them to transmit highly frequent neural signals. We turn now to a channel with a very different role. *Gap junctions*, also known as *cell-to-cell channels*, serve as passageways between the interiors of contiguous cells. Gap junctions are clustered in discrete regions of the plasma membranes of apposed cells. Electron micrographs of sheets of gap junctions show them tightly packed in a regular hexagonal array (Figure 13.31). A 20-Å central hole, the lumen of the channel, is prominent in each gap junction. These channels span the intervening space, or gap, between apposed cells (hence, the name "gap junction"). The width of the gap between the cytoplasms of the two cells is about 35 Å.

Small hydrophilic molecules as well as ions can pass through gap junctions. The pore size of the junctions was determined by microinjecting a series of fluorescent molecules into cells and observing their passage into adjoining cells. All polar molecules with a mass of less than about 1 kd can readily pass through these cell-to-cell channels. Thus, *inorganic ions and most metabolites (e.g., sugars, amino acids, and nucleotides) can flow between the interiors of cells joined by gap junctions*. In contrast, proteins, nucleic acids, and polysaccharides are too large to traverse these channels. *Gap junctions are important for intercellular communication*. Cells in some excitable tissues, such as heart muscle, are coupled by the rapid flow of ions through these junctions, which ensures a rapid and synchronous response to stimuli. Gap junctions are also essential for the nourishment of cells that are distant from blood vessels, as in lens and bone. Moreover, communicating channels are important in development and differentiation. For example, the quiescent uterus transforms to a forcefully contracting organ at the onset of labor; the formation of functional gap junctions at that time creates a syncytium of muscle cells that contract in synchrony.

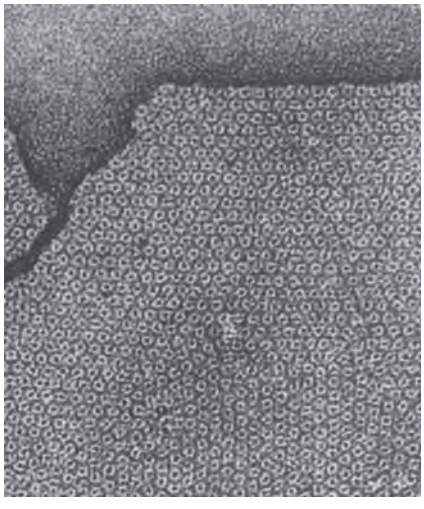


Figure 13.31 Gap junctions. This electron micrograph shows a sheet of isolated gap junctions. The cylindrical connexons form a hexagonal lattice having a unit-cell length of 85 Å. The densely stained central hole has a diameter of about 20 Å. [Don W. Fawcett/ Photo Researchers.]

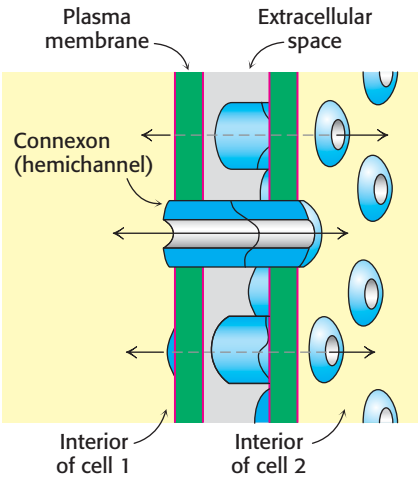


Figure 13.32 Schematic representation of a gap junction. [Courtesy of Dr. Werner Loewenstein.]

A cell-to-cell channel is made of 12 molecules of *connexin*, one of a family of transmembrane proteins with molecular masses ranging from 30 to 42 kd. Each connexin molecule appears to have four membrane-spanning helices. Six connexin molecules are hexagonally arrayed to form a half-channel, called a *connexon* or *hemichannel* (Figure 13.32). Two connexons join end to end in the intercellular space to form a functional channel between the communicating cells. Cell-to-cell channels differ from other membrane channels in three respects: (1) they traverse *two* membranes rather than one; (2) they connect cytoplasm to cytoplasm, rather than to the extracellular space or the lumen of an organelle; and (3) the connexons forming a channel are synthesized by different cells. Gap junctions form readily when cells are brought together. A cell-to-cell channel, once formed, tends to stay open for seconds to minutes. They are closed by high concentrations of calcium ion and by low pH. *The closing of gap junctions by Ca^{2+} and H^+ serves to seal normal cells from injured or dying neighbors.* Gap junctions are also controlled by membrane potential and by hormone-induced phosphorylation.

The human genome encodes 21 distinct connexins. Different members of this family are expressed in different tissues. For example, connexin 26 is expressed in key tissues in the ear. Mutations in this connexin are associated with hereditary deafness. The mechanistic basis for this deafness appears to be insufficient transport of ions or second-messenger molecules, such as inositol trisphosphate, between sensory cells.

13.6 Specific Channels Increase the Permeability of Some Membranes to Water

One more important class of channels does not take part in ion transport at all. Instead, these channels increase the rate at which water flows through membranes. As noted in Chapter 12, membranes are reasonably permeable to water. Why, then, are water-specific channels required? In certain tissues, in some circumstances, rapid water transport through membranes is necessary. In the kidney, for example, water must be rapidly reabsorbed into the bloodstream after filtration. Similarly, in the secretion of saliva and tears, water must flow quickly through membranes. These observations suggested the existence of specific water channels, but initially the channels could not be identified.

The channels (now called *aquaporins*) were discovered serendipitously. Peter Agre noticed a protein present at high levels in red-blood-cell membranes that had been missed because the protein does not stain well with Coomassie brilliant blue. This protein was found in large quantities in red blood cells as well as in tissues such as kidneys and corneas, precisely the tissues thought to contain water channels. On the basis of this observation, further studies were designed that revealed that this 24-kd membrane protein is, indeed, a water channel.

The structure of aquaporin has been determined (Figure 13.33). The protein consists of six membrane-spanning α helices. Two loops containing hydrophilic residues line the actual channel. Water molecules pass through in single file at a rate of 10^6 molecules per second. Importantly, specific positively charged residues toward the center of the channel prevent the transport of protons through aquaporin. Thus, aquaporin channels will not disrupt proton gradients, which play fundamental roles in energy transduction, as we will see in Chapter 18. The aquaporins reveal that channels can evolve that specifically do not conduct ions, as can those that do.

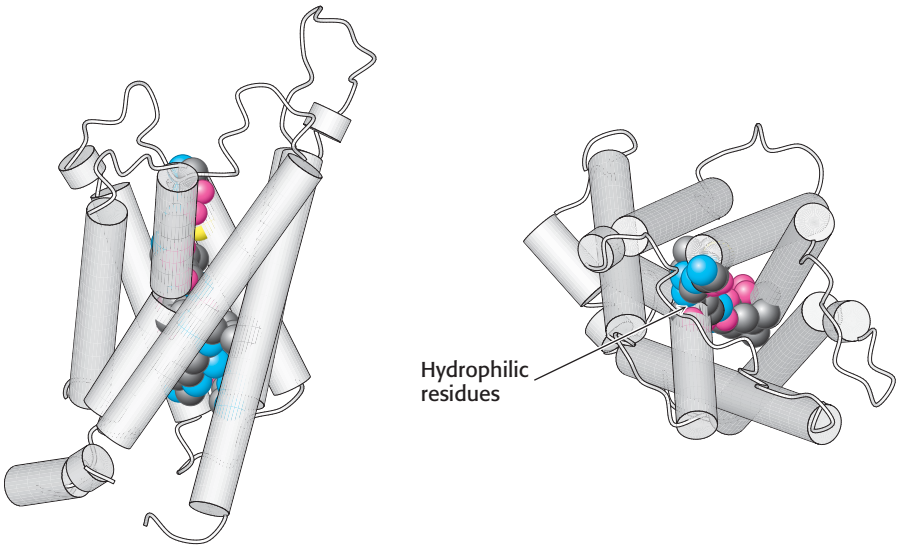


Figure 13.33 Structure of aquaporin. The structure of aquaporin viewed from the side (left) and from the top (right). *Notice* the hydrophilic residues (shown as space-filling models) that line the water channel. [Drawn from 1J4N.pdb.]

Summary

13.1 The Transport of Molecules Across a Membrane May Be Active or Passive

For a net movement of molecules across a membrane, two features are required: (1) the molecule must be able to cross a hydrophobic barrier and (2) an energy source must power the movement. Lipophilic molecules can pass through a membrane's hydrophobic interior by simple diffusion. These molecules will move down their concentration gradients. Polar or charged molecules require proteins to form passages through the hydrophobic barrier. Passive transport or facilitated diffusion takes place when an ion or polar molecule moves down its concentration gradient. If a molecule moves against a concentration gradient, an external energy source is required; this movement is referred to as active transport and results in the generation of concentration gradients. The electrochemical potential measures the combined ability of a concentration gradient and an uneven distribution of charge to drive species across a membrane.

13.2 Two Families of Membrane Proteins Use ATP Hydrolysis to Pump Ions Across Membranes

Active transport is often carried out at the expense of ATP hydrolysis. P-type ATPases pump ions against a concentration gradient and become transiently phosphorylated on an aspartic acid residue in the process of transport. P-type ATPases, which include the sarcoplasmic reticulum Ca^{2+} ATPase and the Na^+–K^+ ATPase, are integral membrane proteins with conserved structures and catalytic mechanisms. Membrane proteins containing ATP-binding cassette domains are another family of ATP-dependent pumps. Each pump includes four major domains: two domains span the membrane and two others contain ABC P-loop ATPase structures. These pumps are not phosphorylated during pumping; rather, they use the energy of ATP binding and hydrolysis to drive conformational changes that result in the transport of specific substrates across membranes. The multidrug-resistance proteins confer resistance on cancer cells by pumping chemotherapeutic drugs out of a cancer cell before the drugs can exert their effects.

13.3 Secondary Transporters Use One Concentration Gradient to Power the Formation of Another

Carriers are proteins that transport ions or molecules across the membrane without hydrolysis of ATP. They can be classified as uniporters, antiporters, and symporters. Uniporters transport a substrate in either direction, determined by the concentration differences. Antiporters and symporters can mediate secondary active transport by coupling the uphill flow of one ion or molecule to the downhill flow of another. Antiporters couple the downhill flow of one type of ion in one direction to the uphill flow of another in the opposite direction. Symporters move both ions in the same direction. Studies of the lactose permease from *E. coli* have been a source of insight into both the structures and the mechanisms of secondary transporters.

13.4 Specific Channels Can Rapidly Transport Ions Across Membranes

Ion channels allow the rapid movement of ions across the hydrophobic barrier of the membrane. The activity of individual ion-channel molecules can be observed by using patch-clamp techniques. Many ion channels have a common structural framework. In regard to K^+ channels, hydrated potassium ions must transiently lose their coordinated water molecules as they move to the narrowest part of the channel, termed the selectivity filter. In the selectivity filter, peptide carbonyl groups coordinate the ions. Rapid ion flow through the selectivity filter is facilitated by ion–ion repulsion, with one ion pushing the next ion through the channel. Some ion channels are voltage gated: changes in membrane potential induce conformational changes that open these channels. Many channels spontaneously inactivate after having been open for a short period of time. In some cases, inactivation is due to the binding of a domain of the channel termed the "ball" in the mouth of the channel to block it. Other channels, typified by the acetylcholine receptor, are opened or closed by the binding of ligands. Ligand-gated and voltage-gated channels work in concert to generate action potentials. Inherited mutations or drugs that interfere with the ion channels that produce the action potential can result in potentially life threatening conditions.

13.5 Gap Junctions Allow Ions and Small Molecules to Flow Between Communicating Cells

In contrast with many channels, which connect the cell interior with the environment, gap junctions, or cell-to-cell channels, serve to connect the interiors of contiguous groups of cells. A cell-to-cell channel is composed of 12 molecules of connexin, which associate to form two 6-membered connexons.

13.6 Specific Channels Increase the Permeability of Some Membranes to Water

Some tissues contain proteins that increase the permeability of membranes to water. Each water-channel-forming protein, termed an aquaporin, consists of six membrane-spanning α helices and a central channel lined with hydrophilic residues that allow water molecules to pass in single file. Aquaporins do not transport protons.

Key Terms

pump (p. 371)
carrier (p. 371)
channel (p. 371)
active transport (p. 371)
facilitated diffusion (passive transport) (p. 371)
ATP-driven pump (p. 371)
primary active transport (p. 371)
secondary active transport (p. 371)

simple diffusion (p. 372)
electrochemical potential (membrane potential) (p. 373)
Na^+–K^+ pump (Na^+–K^+ATPase) (p. 374)
sarcoplasmic reticulum Ca^{2+} ATPase (SERCA) (p. 374)
gastric H^+–K^+ ATPase (p. 374)
P-type ATPase (p. 374)
eversion (p. 376)
cardiotonic steroid (p. 377)
digitalis (p. 377)
multidrug resistance (p. 378)
multidrug-resistance (MDR) protein (P-glycoprotein) (p. 378)
ATP-binding cassette (ABC) domain (p. 378)
ABC transporter (p. 378)
secondary transporter (cotransporter) (p. 380)
antitransporter (symporter) (p. 380)
uniporter (p. 380)
lactose permease (p. 381)
ion channel (p. 382)
nerve impulse (p. 382)
action potential (p. 382)
patch-clamp technique (p. 383)
gigaseal (p. 383)
selectivity filter (p. 385)
voltage-gated channel (p. 387)
ball-and-chain model (p. 388)
neurotransmitter (p. 389)
acetylcholine (p. 389)
synaptic cleft (p. 389)
acetylcholine receptor (p. 389)
ligand-gated channel (p. 390)
equilibrium potential (p. 391)
Nernst equation (p. 391)
long QT syndrome (LQTS) (p. 392)
gap junction (cell-to-cell channels) (p. 393)
connexin (p. 394)
connexon (hemichannel) (p. 394)
aquaporin (p. 394)

Problems

1. *A helping hand.* Differentiate between simple diffusion and facilitated diffusion.

2. *Powering movement.* What are the two forms of energy that can power active transport?

3. *Carriers.* Name the three types of carrier proteins. Which of these can mediate secondary active transport?

4. *The price of extrusion.* What is the free-energy cost of pumping Ca^{2+} out of a cell when the cytoplasmic concentration is 0.4 μM, the extracellular concentration is 1.5 mM, and the membrane potential is −60 mV?

5. *Equilibium potentials.* For a typical mammalian cell, the intracellular and extracellular concentrations of the chloride ion (Cl^-) are 4 mM and 150 mM, respectively. For the calcium ion (Ca^{2+}), the intracellular and extracellular concentrations are 0.2 μM and 1.8 mM, respectively. Calculate the equilibrium potentials at 37°C for these two ions.

6. *How sweet it is.* Some animal cells take up glucose by a symporter powered by the simultaneous entry of Na^+. The entry of Na^+ provides a free-energy input of 10.8 kJ mol^{-1} (2.6 kcal mol^{-1}) under typical cellular conditions (external $[Na^+]$ = 143 mM, internal $[Na^+]$ = 14 mM, and membrane potential = −50 mV). How large a concentration of glucose can be generated by this free-energy input?

7. *Variations on a theme.* Write a detailed mechanism for transport by the Na^+–K^+ ATPase based on analogy with the mechanism of the Ca^{2+} ATPase shown in Figure 13.5.

8. *Pumping protons.* Design an experiment to show that the action of lactose permease can be reversed in vitro to pump protons.

9. *Opening channels.* Differentiate between ligand-gated and voltage-gated channels.

10. *Different directions.* The K^+ channel and the Na^+ channel have similar structures and are arranged in the same orientation in the cell membrane. Yet the Na^+ channel allows sodium ions to flow into the cell and the K^+ channel allows potassium ions to flow out of the cell. Explain.

11. *Differing mechanisms.* Distinguish the mechanisms by which uniporters and channels transport ions or molecules across the membrane.

12. *Short circuit.* Carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP) is a proton ionophore: it enables protons to pass freely through membranes. Treatment of *E. coli* with FCCP prevents the accumulation of lactose in these cells. Explain.

13. *Working together.* The human genome contains more than 20 connexin-encoding genes. Several of these genes are expressed in high levels in the heart. Why are connexins so highly expressed in cardiac tissue?

14. *Structure–activity relations.* On the basis of the structure of tetrodotoxin, propose a mechanism by which the toxin inhibits Na^+ flow through the Na^+ channel.

15. *Hot stuff.* When SERCA is incubated with [γ-^{32}P]ATP (a form of ATP in which the terminal phosphate is labeled with radioactive ^{32}P) and calcium at 0°C for 20 seconds and analyzed by gel electrophoresis, a radioactive band is observed at the molecular weight corresponding to full-length SERCA. Why is a labeled band observed? Would you expect a similar band if you were performing a similar assay, with a suitable substrate, for the MDR protein?

16. *A dangerous snail.* Cone snails are carnivores that inject a powerful set of toxins into their prey, leading to rapid paralysis. Many of these toxins are found to bind to specific ion-channel proteins. Why are such molecules so toxic? How might such toxins be useful for biochemical studies?

17. *Pause for effect.* Immediately after the repolarization phase of an action potential, the neuronal membrane is temporarily unable to respond to the stimulation of a second action potential, a phenomenon referred to as the *refractory period*. What is the mechanistic basis for the refractory period?

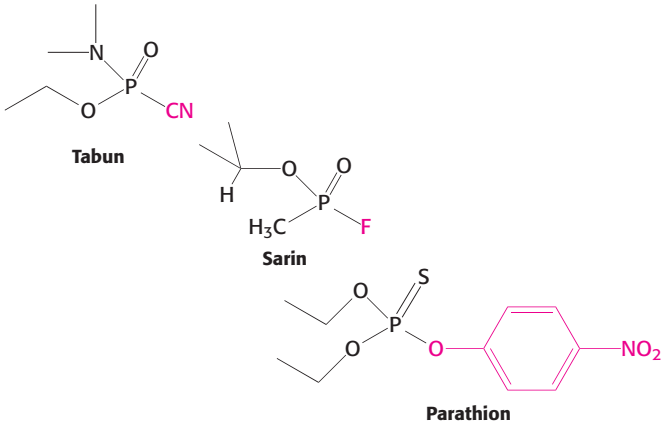
18. *Only a few.* Why do only a small number of sodium ions need to flow through the Na^+ channel to change the membrane potential significantly?

19. *More than one mechanism.* How might a mutation in a cardiac voltage-dependent sodium channel cause long QT syndrome?

20. *Mechanosensitive channels.* Many species contain ion channels that respond to mechanical stimuli. On the basis of the properties of other ion channels, would you expect the flow of ions through a single open mechanosensitive channel to increase in response to an appropriate stimulus? Why or why not?

21. *Concerted opening.* Suppose that a channel obeys the concerted allosteric model (MWC model, Section 7.2). The binding of ligand to the R state (the open form) is 20 times as tight as that to the T state (the closed form). In the absence of ligand, the ratio of closed to open channels is 10^5. If the channel is a tetramer, what is the fraction of open channels when 1, 2, 3, and 4 ligands are bound?

22. *Respiratory paralysis.* The neurotransmitter acetylcholine is degraded by a specific enzyme that is inactivated by Tabun, sarin, and parathion. On the basis of the structures below, propose a possible basis for their lethal actions.



23. *Ligand-induced channel opening.* The ratio of open to closed forms of the acetylcholine receptor channel containing zero, one, and two bound acetylcholine molecules is 5×10^{-6}, 1.2×10^{-3}, and 14, respectively.

(a) By what factor is the open-to-closed ratio increased by the binding of the first acetylcholine molecule? The second acetylcholine molecule?

(b) What are the corresponding free-energy contributions to channel opening at 25°C?

(c) Can the allosteric transition be accounted for by the MWC concerted model (Section 7.2)?

24. *Frog poison.* Batrachotoxin (BTX) is a steroidal alkaloid from the skin of *Phyllobates terribilis*, a poisonous Colombian frog (the source of the poison used on blowgun darts). In the presence of BTX, Na^+ channels in an excised patch stay persistently open when the membrane is depolarized. They close when the membrane is repolarized. Which transition is blocked by BTX?

25. *Valium target.* γ-Aminobutyric acid (GABA) opens channels that are specific for chloride ions. The $GABA_A$ receptor channel is pharmacologically important because it is the target of Valium, which is used to diminish anxiety.

(a) The extracellular concentration of Cl^- is 123 mM and the intracellular concentration is 4 mM. In which direction does Cl^- flow through an open channel when the membrane potential is in the −60 mV to +30 mV range?

(b) What is the effect of Cl^--channel opening on the excitability of a neuron?

(c) The hydropathy profile of the $GABA_A$ receptor resembles that of the acetylcholine receptor. Predict the number of subunits in this Cl^- channel.

26. *Understanding SERCA.* To study the mechanism of SERCA, you prepare membrane vesicles containing this protein oriented such that its ATP binding site is on the outer surface of the vesicle. To measure pump activity, you use an assay that detects the formation of inorganic phosphate in the medium. When you add calcium and ATP to the medium, you observe phosphate production for only a short period of time. Only after the addition of calcimycin, a molecule that makes membranes selectively permeable to calcium, do you observe sustained phosphate production. Explain.

Chapter Integration Problem

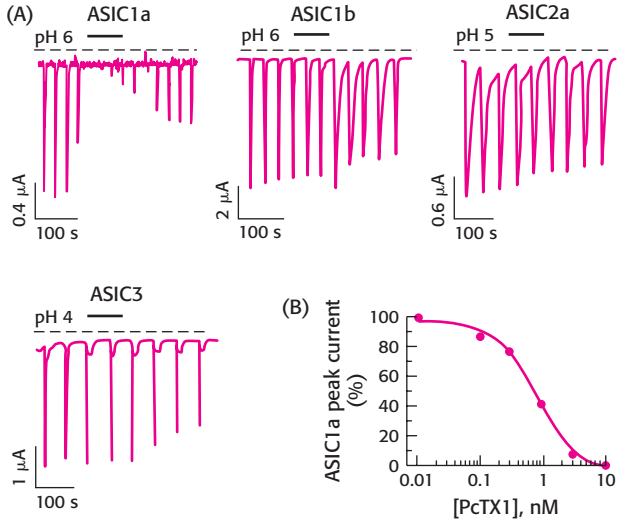
27. *Speed and efficiency matter.* Acetylcholine is rapidly destroyed by the enzyme acetylcholinesterase. This enzyme, which has a turnover number of 25,000 per second, has attained catalytic perfection with a k_{cat}/K_M of $2 \times 10^8\ M^{-1}s^{-1}$. Why is the efficiency of this enzyme physiologically crucial?

Mechanism Problem

28. *Remembrance of mechanisms past*. Acetylcholinesterase converts acetylcholine into acetate and choline. Like serine proteases, acetylcholinesterase is inhibited by DIPF. Propose a catalytic mechanism for acetylcholine digestion by acetylcholinesterase. Show the reaction as chemical structures.

Data Interpretation Problems

29. *Tarantula toxin*. Acid sensing is associated with pain, tasting, and other biological activities (Chapter 33). Acid sensing is carried out by a ligand-gated channel that permits Na^+ influx in response to H^+. This family of acid-sensitive ion channels (ASICs) includes a number of members. Psalmotoxin 1 (PcTX1), a venom from the tarantula, inhibits some members of this family. The following electrophysiological recordings of cells containing several members of the ASIC family were made in the presence of the toxin at a concentration of 10 nM. The channels were opened by changing the pH from 7.4 to the indicated values. The PcTX1 was present for a short time (indicated by the black bar above the recordings below), after which time it was rapidly washed from the system.



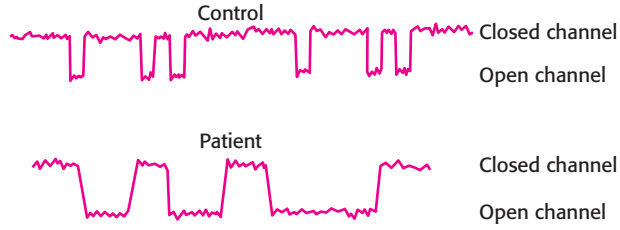
(A) Electrophysiological recordings of cells exposed to tarantula toxin. (B) Plot of peak current of a cell containing the ASIC1a protein versus the toxin concentration. [From P. Escoubas et al. *J. Biol. Chem.* 275(2000):25116–25121.]

(a) Which member of the ASIC family—ASIC1a, ASIC1b, ASIC2a, or ASIC3—is most sensitive to the toxin?

(b) Is the effect of the toxin reversible? Explain.

(c) What concentration of PcTX1 yields 50% inhibition of the sensitive channel?

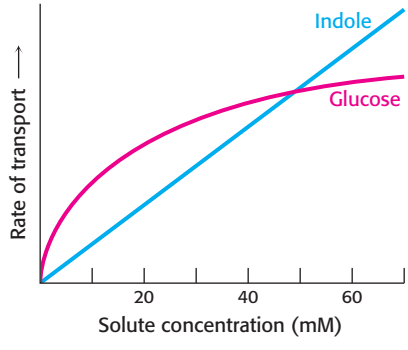
30. *Channel problems 1*. A number of pathological conditions result from mutations in the acetylcholine receptor channel. One such mutation in the β subunit, βV266M, causes muscle weakness and rapid fatigue. An investigation of the acetylcholine-generated currents through the acetylcholine receptor channel for both a control and a patient yielded the following results.



What is the effect of the mutation on channel function? Suggest some possible biochemical explanations for the effect.

31. *Channel problems 2*. The acetylcholine receptor channel can also undergo mutation leading to fast-channel syndrome (FCS), with clinical manifestations similar to those of slow-channel syndrome (SCS). What would the recordings of ion movement look like in this syndrome? Suggest a biochemical explanation.

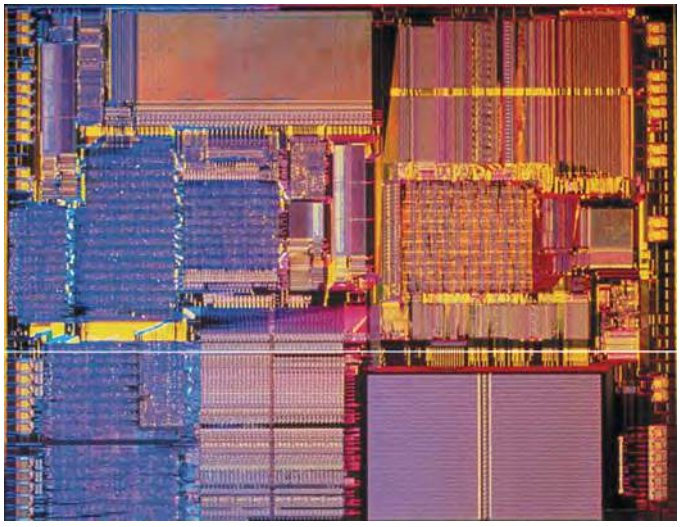
32. *Transport differences*. The rate of transport of two molecules, indole and glucose, across a cell membrane is shown below. What are the differences between the transport mechanisms of the two molecules? Suppose that ouabain inhibited the transport of glucose. What would this inhibition suggest about the mechanism of transport?

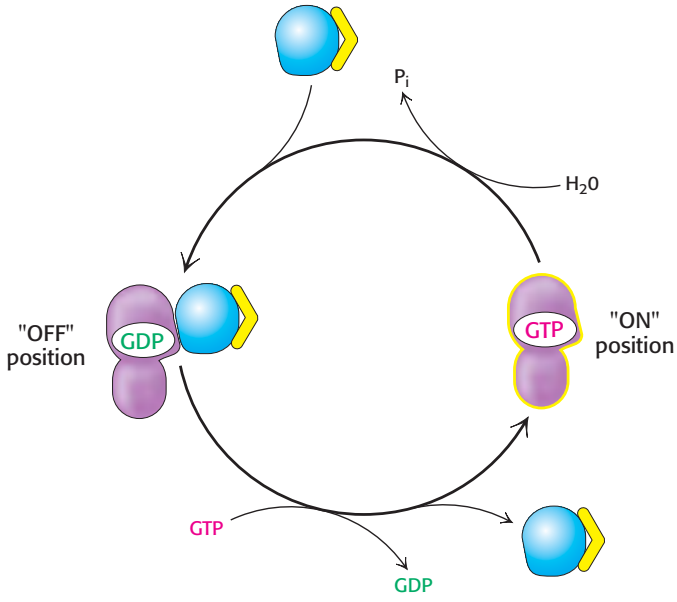


This page intentionally left blank

CHAPTER 14

Signal-Transduction Pathways





Signal-transduction circuits in biological systems have molecular on–off switches that, like those in a computer chip (above), transmit information when "on." Common among these circuits are those including G proteins (right), which transmit a signal when bound to GTP and are silent when bound to GDP. [(Left) Courtesy of Intel.]

A cell is highly responsive to specific chemicals in its environment: it may adjust its metabolism or alter gene-expression patterns on sensing their presence. In multicellular organisms, these chemical signals are crucial to coordinating physiological responses (Figure 14.1). Three examples of molecular signals that stimulate a physiological response are epinephrine (sometimes called adrenaline), insulin, and epidermal growth factor (EGF). When a mammal is threatened, its adrenal glands release the hormone epinephrine, which stimulates the mobilization of energy stores and leads to improved cardiac function. After a full meal, the β cells in the pancreas release insulin, which stimulates a host of physiological responses, including the uptake of glucose from the bloodstream and its storage as glycogen. The release of EGF in response to a wound stimulates specific cells to grow and divide. In all these cases, the cell receives information that a certain molecule within its environment is present above some threshold concentration. The chain of events that converts the message "this molecule is present" into the ultimate physiological response is called *signal transduction*.

Signal-transduction pathways often comprise many components and branches. They can thus be immensely complicated and confusing. However, the logic of signal transduction can be revealed by examining the

OUTLINE

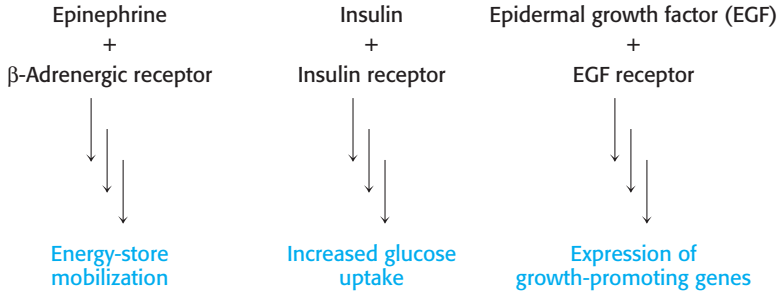


Figure 14.1 Three signal-transduction pathways. The binding of signaling molecules to their receptors initiates pathways that lead to important physiological responses.

common strategies and classes of molecules that recur in these pathways. These principles are introduced here because signal-transduction pathways affect essentially all of the metabolic pathways that we will be exploring throughout the rest of the book.

Signal transduction depends on molecular circuits

Signal-transduction pathways follow a broadly similar course that can be viewed as a molecular circuit (Figure 14.2). All such circuits contain certain key steps:

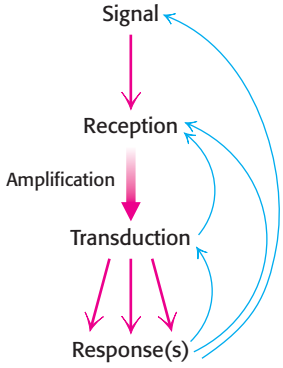


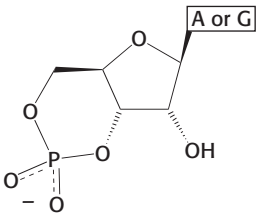
Figure 14.2 Principles of signal transduction. An environmental signal is first received by interaction with a cellular component, most often a cell-surface receptor. The information that the signal has arrived is then converted into other chemical forms, or *transduced*. The transduction process often comprises many steps. The signal is often amplified before evoking a response. Feedback pathways regulate the entire signaling process.

1. *Release of the Primary Messenger*. A stimulus such as a wound or digested meal triggers the release of the signal molecule, also called the *primary messenger*.

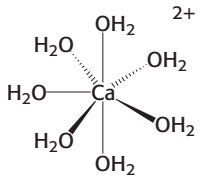
2. *Reception of the Primary Messenger*. Most signal molecules do not enter cells. Instead, proteins in the cell membrane act as *receptors* that bind the signal molecules and transfer the information that the molecule has bound from the environment to the cell's interior. Receptors span the cell membrane and thus have both extracellular and intracellular components. A binding site on the extracellular side specifically recognizes the signal molecule (often referred to as the *ligand*). Such binding sites are analogous to enzyme active sites except that no catalysis takes place within them. The interaction of the ligand and the receptor alters the tertiary or quaternary structure of the receptor so as to induce a structural change on the intracellular side.

3. *Delivery of the Message Inside the Cell by the Second Messenger*. Other small molecules, called *second messengers*, are used to relay information from receptor–ligand complexes. Second messengers are intracellular molecules that change in concentration in response to environmental signals and mediate the next step in the molecular information circuit. Some particularly important second messengers are cyclic AMP (cAMP) and cyclic GMP (cGMP), calcium ion, inositol 1,4,5-trisphosphate (IP_3), and diacylglycerol (DAG; Figure 14.3).

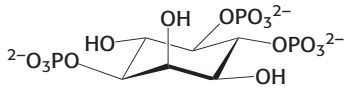
The use of second messengers has several consequences. First, the signal may be amplified significantly in the generation of second messengers. Only a small number of receptor molecules may be activated by the direct binding of signal molecules, but each activated receptor molecule can lead to the generation of many second messengers. Thus, *a low concentration of signal in the environment, even as little as a single molecule, can yield a large intracellular signal and response.* Second, second messengers are often free to diffuse to other cellular compartments where they can influence processes



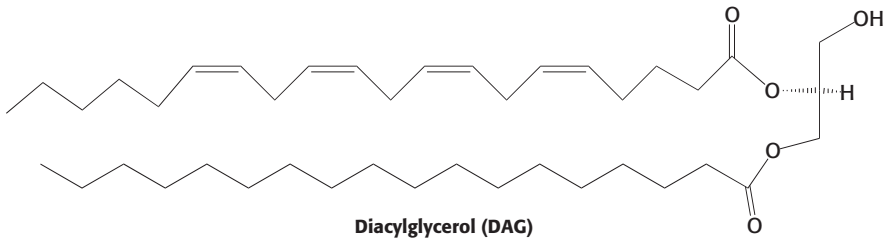
cAMP, cGMP



Calcium ion



Inositol 1,4,5-trisphosphate (IP_3)



Diacylglycerol (DAG)

Figure 14.3 Common second messengers. Second messengers are intracellular molecules that change in concentration in response to environmental signals. That change in concentration conveys information inside the cell.

throughout the cell. Third, the use of common second messengers in multiple signaling pathways creates both opportunities and potential problems. Input from several signaling pathways, often called *cross talk,* may alter the concentration of a common second messenger. Cross talk permits more finely tuned regulation of cell activity than would the action of individual independent pathways. However, inappropriate cross talk can result in the misinterpretation of changes in second-messenger concentration.

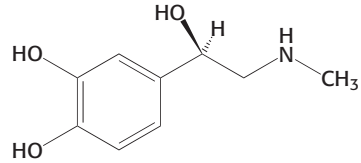
4. *Activation of Effectors That Directly Alter the Physiological Response*. The ultimate effect of the signal pathway is to activate (or inhibit) the pumps, channels, enzymes, and transcription factors that directly control metabolic pathways, gene expression, and the permeability of membranes to specific ions.

5. *Termination of the Signal.* After a cell has completed its response to a signal, the signaling process must be terminated or the cell loses its responsiveness to new signals. Moreover, signaling processes that fail to terminate properly can have highly undesirable consequences. As we will see, many cancers are associated with signal-transduction processes that are not properly terminated, especially processes that control cell growth.

In this chapter, we will examine components of the three signal-transduction pathways shown in Figure 14.1. In doing so, we will see several classes of adaptor domains present in signal-transduction proteins. These domains usually recognize specific classes of molecules and help transfer information from one protein to another. The components described in the context of these three pathways recur in many other signal-transduction pathways; so bear in mind that the specific examples are representative of many such pathways.

14.1 Heterotrimeric G Proteins Transmit Signals and Reset Themselves

Epinephrine is a hormone secreted by the adrenal glands of mammals in response to internal and external stressors. It exerts a wide range of effects—referred to as the "fight or flight" response—to help organisms anticipate the need for rapid muscular activity, including acceleration of heart rate, dilation of the smooth muscle of the airways, and initiation of



Epinephrine

Table 14.1 Biological functions mediated by 7TM receptors

Hormone action
Hormone secretion
Neurotransmission
Chemotaxis
Exocytosis
Control of blood pressure
Embryogenesis
Cell growth and differentiation
Development
Smell
Taste
Vision
Viral infection

Source: After J. S. Gutkind, *J. Biol. Chem.* 273:1839–1842, 1998.

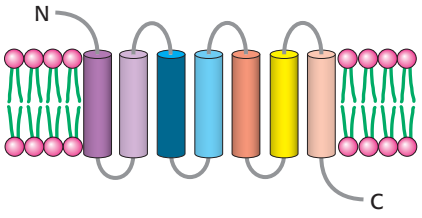


Figure 14.4 The 7TM receptor. Schematic representation of a 7TM receptor showing its passage through the membrane seven times.

the breakdown of glycogen (Section 21.3) and fatty acids (Section 22.2). Epinephrine signaling begins with ligand binding to a protein called the *β-adrenergic receptor* (β-AR). The β-AR is a member of the largest class of cell-surface receptors, called the *seven-transmembrane-helix* (7TM) *receptors*. Members of this family are responsible for transmitting information initiated by signals as diverse as hormones, neurotransmitters, odorants, tastants, and even photons (Table 14.1). More than 20,000 such receptors are now known. Furthermore, about one-third of the therapeutic drugs that we use target receptors of this class. As the name indicates, these receptors contain seven helices that span the membrane bilayer. The receptors are sometimes referred to as *serpentine receptors* because the single polypeptide chain "snakes" through the membrane seven times (Figure 14.4).

The first member of the 7TM receptor family to have its three-dimensional structure determined was *rhodopsin* (Figure 14.5A), a protein in the retina of the eye that senses the presence of photons and initiates the signaling cascade responsible for visual sensation. A single lysine residue within rhodopsin is covalently modified by a form of vitamin A, 11-*cis*-retinal. This modification is located near the extracellular side of the receptor, within the region surrounded by the seven transmembrane helices. As will be considered in greater detail in Section 33.3, exposure to light induces the isomerization of 11-*cis*-retinal to its all-*trans* form, producing a structural change in the receptor that results in the initiation of an action potential that is ultimately interpreted by the brain as visual stimulus.

In 2007, the three-dimensional structure of the β_2 subtype of the human adrenergic receptor (β_2-AR) bound to an inhibitor was solved by x-ray crystallography. This inhibitor, carazolol, competes with epinephrine for binding to the β_2-AR, much in the same way that competitive inhibitors act at enzyme active sites (Section 8.5). The structure of the β_2-AR revealed considerable similarities with that of rhodopsin, particularly with respect to the locations of 11-*cis*-retinal in rhodopsin and the binding site for carazolol (Figure 14.5B). Although the precise details of the conformational changes induced by ligand binding to the β-AR remain to be established, epinephrine likely binds to the β-AR in a similar region of the receptor as carazolol

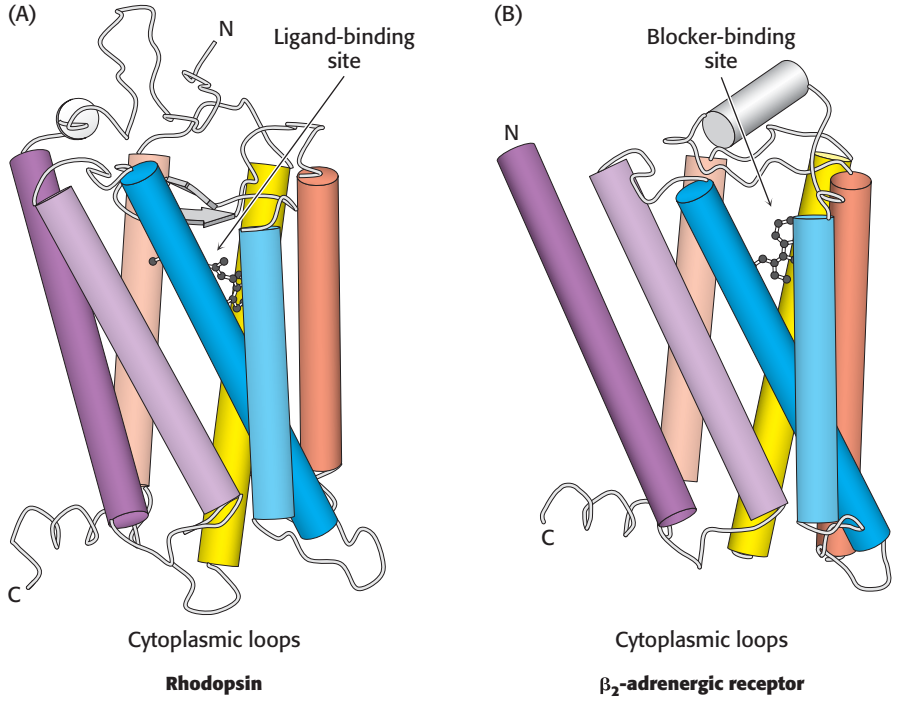


Figure 14.5 Structures of rhodopsin and the β_2-adrenergic receptor. Three-dimensional structure of rhodopsin (A) and the β_2-adrenergic receptor (B). *Notice* the resemblance in the overall architecture of both receptors and the similar locations of the rhodopsin ligand 11-*cis*-retinal and the β_2-AR blocker carazolol. [Drawn from 1F88.pdb and 2RH1.pdb.]

binds, triggering conformational changes in the cytoplasmic parts of the β-AR comparable to those induced by retinal isomerization in rhodopsin. Thus, the binding of a ligand from outside the cell induces a structural rearrangement in the part of the 7TM receptor that is positioned inside the cell.

Ligand binding to 7TM receptors leads to the activation of heterotrimeric G proteins

What is the next step in the pathway? The conformational change in the receptor's cytoplasmic domain activates a protein called a *G protein* because it binds guanyl nucleotides. The activated G protein stimulates the activity of adenylate cyclase, an enzyme that catalyzes the conversion of ATP into cAMP. The G protein and adenylate cyclase remain attached to the membrane, whereas cAMP can travel throughout the cell carrying the signal originally brought by the binding of epinephrine. Figure 14.6 provides a broad overview of these steps.

Let us consider the role of the G protein in this signaling pathway in greater detail. In its unactivated state, the G protein is bound to GDP. In this form, the G protein exists as a heterotrimer consisting of α, β, and γ subunits; the α subunit (referred to as G_α) binds the nucleotide (Figure 14.7). The α subunit is a member of the P-loop NTPase family (Section 9.4), and the P-loop participates in nucleotide binding. The α and γ subunits are usually anchored to the membrane by covalently attached fatty acids. *The role of the hormone-bound receptor is to catalyze the exchange of GTP for bound GDP.* The hormone–receptor complex interacts with the heterotrimeric G protein and opens the nucleotide-binding site such that GTP in the cell can displace GDP.

On GTP binding, the α subunit simultaneously dissociates from the βγ dimer ($G_{\beta\gamma}$), transmitting the signal that the receptor has bound its ligand. *A single hormone–receptor complex can stimulate nucleotide exchange in many G-protein heterotrimers.* Thus, hundreds of G_α molecules are

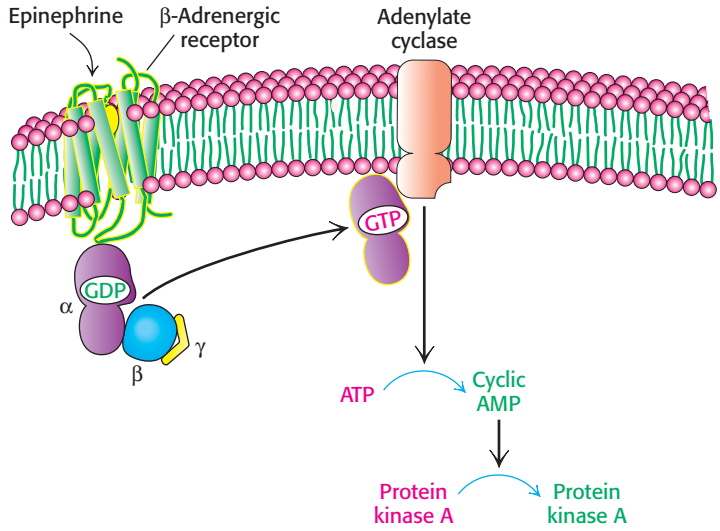


Figure 14.6 Activation of protein kinase A by a G-protein pathway. Hormone binding to a 7TM receptor initiates a signal-transduction pathway that acts through a G protein and cAMP to activate protein kinase A.

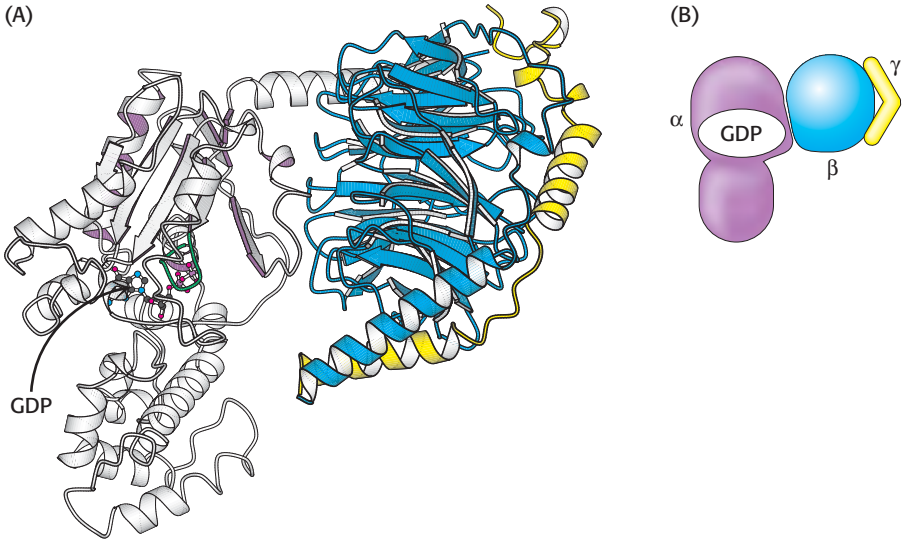


Figure 14.7 A heterotrimeric G protein. (A) A ribbon diagram shows the relation between the three subunits. In this complex, the α subunit (gray and purple) is bound to GDP. *Notice* that GDP is bound in a pocket close to the surface at which the α subunit interacts with the βγ dimer. (B) A schematic representation of the heterotrimeric G protein. [Drawn from 1GOT.pdb.]

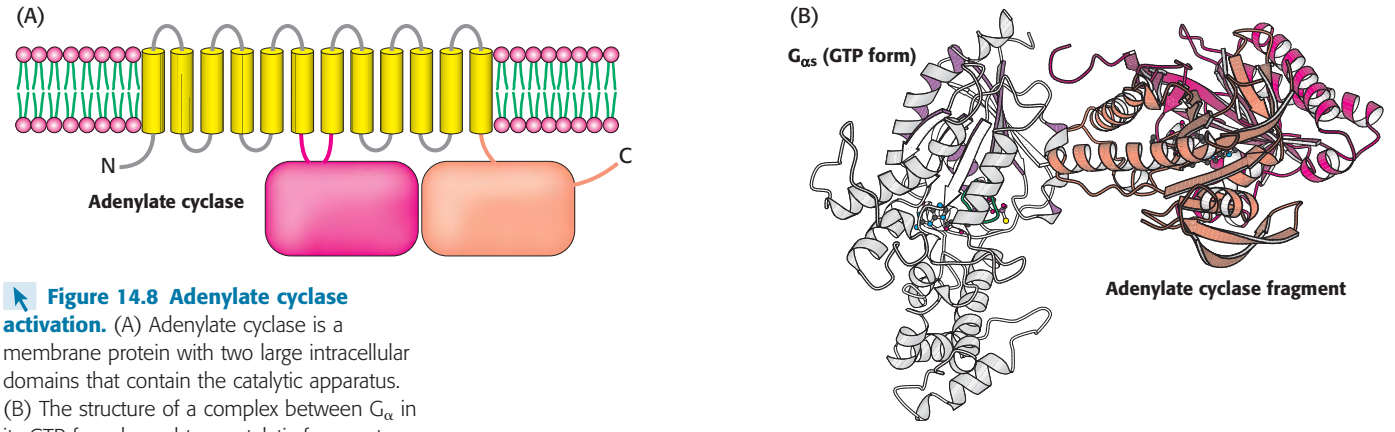


Figure 14.8 Adenylate cyclase activation. (A) Adenylate cyclase is a membrane protein with two large intracellular domains that contain the catalytic apparatus. (B) The structure of a complex between G_α in its GTP form bound to a catalytic fragment from adenylate cyclase. *Notice* that the surface of G_α that had been bound to the βγ dimer now binds adenylate cyclase. [Drawn from 1AZS.pdb.]

converted from their GDP form into their GTP form for each bound molecule of hormone, giving an amplified response. Because they signal through G proteins, 7TM receptors are often called *G-protein-coupled receptors* (GPCRs).

Activated G proteins transmit signals by binding to other proteins

In the GTP form, the surface of G_α that had been bound to $G_{\beta\gamma}$ has changed its conformation from the GDP form so that it no longer has a high affinity for $G_{\beta\gamma}$. This surface is now exposed for binding to other proteins. In the β-AR pathway, the new binding partner is *adenylate cyclase,* the enzyme that converts ATP into cAMP. This enzyme is a membrane protein that contains 12 membrane-spanning helices; two large cytoplasmic domains form the catalytic part of the enzyme (Figure 14.8). The interaction of G_α with adenylate cyclase favors a more catalytically active conformation of the enzyme, thus stimulating cAMP production. Indeed, the G_α subunit that participates in the β-AR pathway is called $G_{\alpha s}$ ("s" stands for stimulatory). *The net result is that the binding of epinephrine to the receptor on the cell surface increases the rate of cAMP production inside the cell.* The generation of cAMP by adenylate cyclase provides a second level of amplification because each activated adenylate cyclase can convert many molecules of ATP into cAMP.

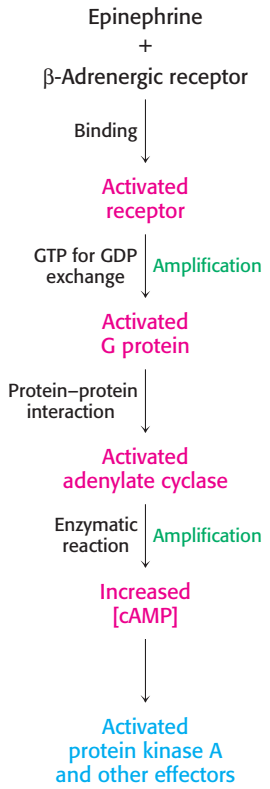


Figure 14.9 Epinephrine signaling pathway. The binding of epinephrine to the β-adrenergic receptor initiates the signal-transduction pathway. The process in each step is indicated (in black) at the left of each arrow. Steps that have the potential for signal amplification are indicated at the right in green.

Cyclic AMP stimulates the phosphorylation of many target proteins by activating protein kinase A

The increased concentration of cAMP can affect a wide range of cellular processes. In the muscle, cAMP stimulates the production of ATP for muscle contraction. In other cell types, cAMP enhances the degradation of storage fuels, increases the secretion of acid by the gastric mucosa, leads to the dispersion of melanin pigment granules, diminishes the aggregation of blood platelets, and induces the opening of chloride channels. How does cAMP influence so many cellular processes? *Most effects of cAMP in eukaryotic cells are mediated by the activation of a single protein kinase.* This key enzyme is *protein kinase A* (PKA).

As described earlier, PKA consists of two regulatory (R) chains and two catalytic (C) chains (see Figure 10.17). In the absence of cAMP, the R_2C_2 complex is catalytically inactive. The binding of cAMP to the regulatory chains releases the catalytic chains, which are catalytically active on their own. Activated PKA then phosphorylates specific serine and threonine residues in many targets to alter their activity. For instance, PKA phosphorylates two enzymes that lead to the breakdown of glycogen, the polymeric store of glucose, and the inhibition of further glycogen synthesis

(Section 21.3). *PKA stimulates the expression of specific genes* by phosphorylating a transcriptional activator called the cAMP response element binding (CREB) protein. This activity of PKA illustrates that signal-transduction pathways can extend into the nucleus to alter gene expression.

The signal-transduction pathway initiated by epinephrine is summarized in Figure 14.9.

G proteins spontaneously reset themselves through GTP hydrolysis

How is the signal initiated by epinephrine switched off? *G_α subunits have intrinsic GTPase activity,* which is used to hydrolyze bound GTP to GDP and P_i. This hydrolysis reaction is slow, however, requiring from seconds to minutes. Thus, the GTP form of G_α is able to activate downstream components of the signal-transduction pathway before it is deactivated by GTP hydrolysis. In essence, *the bound GTP acts as a built-in clock that spontaneously resets the G_α subunit after a short time period.* After GTP hydrolysis and the release of P_i, the GDP-bound form of G_α then reassociates with $G_{\beta\gamma}$ to re-form the inactive heterotrimeric protein (Figure 14.10).

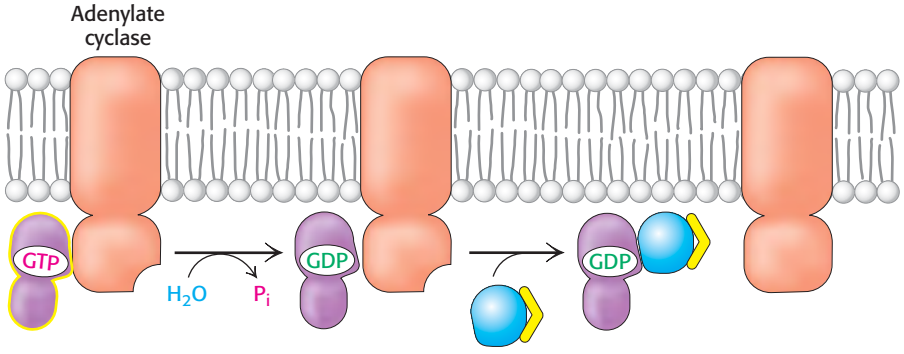


Figure 14.10 Resetting Gα. On hydrolysis of the bound GTP by the intrinsic GTPase activity of G_α, G_α reassociates with the βγ dimer to form the heterotrimeric G protein, thereby terminating the activation of adenylate cyclase.

The hormone-bound activated receptor must be reset as well to prevent the continuous activation of G proteins. This resetting is accomplished by two processes (Figure 14.11). First, the hormone dissociates, returning the receptor to its initial, unactivated state. The likelihood that the receptor

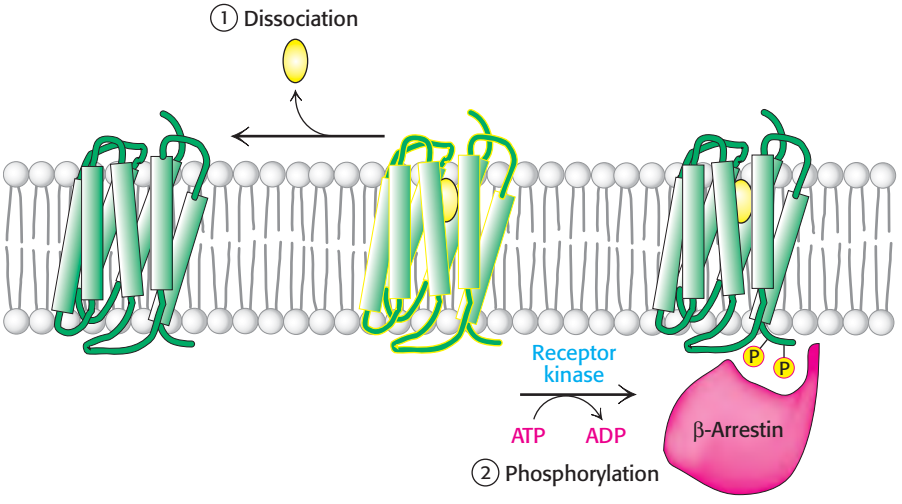


Figure 14.11 Signal termination. Signal transduction by the 7TM receptor is halted (1) by dissociation of the signal molecule from the receptor and (2) by phosphorylation of the cytoplasmic C-terminal tail of the receptor and the subsequent binding of β-arrestin.

remains in its unbound state depends on the extracellular concentration of hormone. Second, the signaling cascade initiated by the hormone–receptor complex activates a kinase that phosphorylates serine and threonine residues in the carboxyl-terminal tail of the receptor. These phosphorylation events result in the deactivation of the receptor. In the example under consideration, *β-adrenergic-receptor kinase* (also called G-protein receptor kinase 2, GRK2) phosphorylates the carboxyl-terminal tail of the hormone–receptor complex but not the unoccupied receptor. Finally, the molecule *β-arrestin* binds to the phosphorylated receptor and further diminishes its ability to activate G proteins.

Some 7TM receptors activate the phosphoinositide cascade

We now turn to another common second-messenger cascade, also employing a 7TM receptor, that is used by many hormones to evoke a variety of responses. The *phosphoinositide cascade,* like the cAMP cascade, converts extracellular signals into intracellular ones. The intracellular messengers formed by activation of this pathway arise from the cleavage of *phosphatidylinositol 4,5-bisphosphate* (PIP_2), a phospholipid present in cell membranes. An example of a signaling pathway based on the phosphoinositide cascade is the one triggered by the receptor for angiotensin II, a peptide hormone that controls blood pressure.

Each type of 7TM receptor signals through a distinct G protein. Whereas the β-adrenergic receptor activates the G protein $G_{\alpha s}$, the angiotensin II receptor activates a G protein called $G_{\alpha q}$. In its GTP-form, $G_{\alpha q}$ binds to and activates the β isoform of the enzyme *phospholipase C*. This enzyme catalyzes the cleavage of PIP_2 into the two second messengers inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG; Figure 14.12).

IP_3 is soluble and diffuses away from the membrane. This second messenger causes the rapid release of Ca^{2+} from the intracellular stores in the

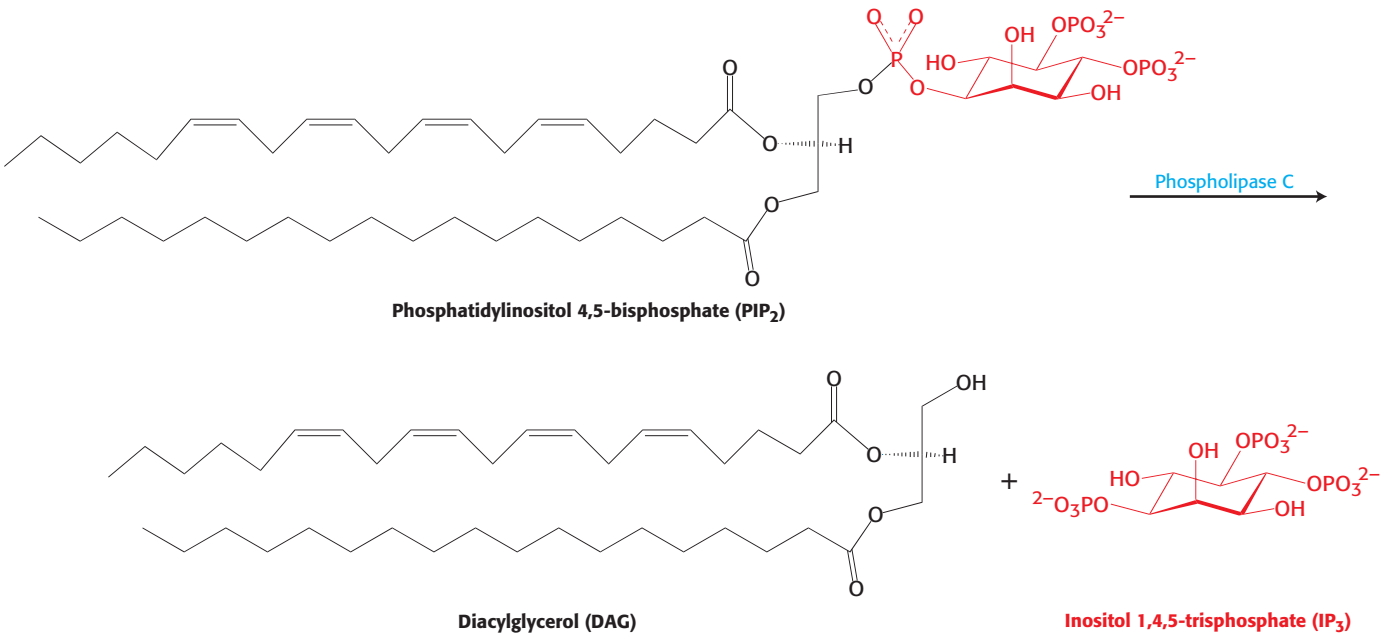


Figure 14.12 Phospholipase C reaction. Phospholipase C cleaves the membrane lipid phosphatidylinositol 4,5-bisphosphate into two second messengers: diacylglycerol, which remains in the membrane, and inositol 1,4,5-trisphosphate, which diffuses away from the membrane.

endoplasmic reticulum (ER), which accumulates a reservoir of Ca^{2+} through the action of transporters such as Ca^{2+} ATPase (Section 13.2). On binding IP_3, specific IP_3-gated Ca^{2+}-channel proteins in the ER membrane open to allow calcium ions to flow from the ER into the cytoplasm. Calcium ion is itself a signaling molecule: it can bind proteins, including a ubiquitous signaling protein called calmodulin and enzymes such as protein kinase C. By such means, the elevated level of cytoplasmic Ca^{2+} triggers processes such as smooth-muscle contraction, glycogen breakdown, and vesicle release.

DAG remains in the plasma membrane. There, it activates *protein kinase C* (PKC), a protein kinase that phosphorylates serine and threonine residues in many target proteins. To bind DAG, the specialized DAG-binding domains of this kinase require bound calcium. Note that diacylglycerol and IP_3 work in tandem: IP_3 increases the Ca^{2+} concentration, and Ca^{2+} facilitates the DAG-mediated activation of protein kinase C. The phosphoinositide cascade is summarized in Figure 14.13. Both IP_3 and DAG act transiently because they are converted into other species by phosphorylation or other processes.

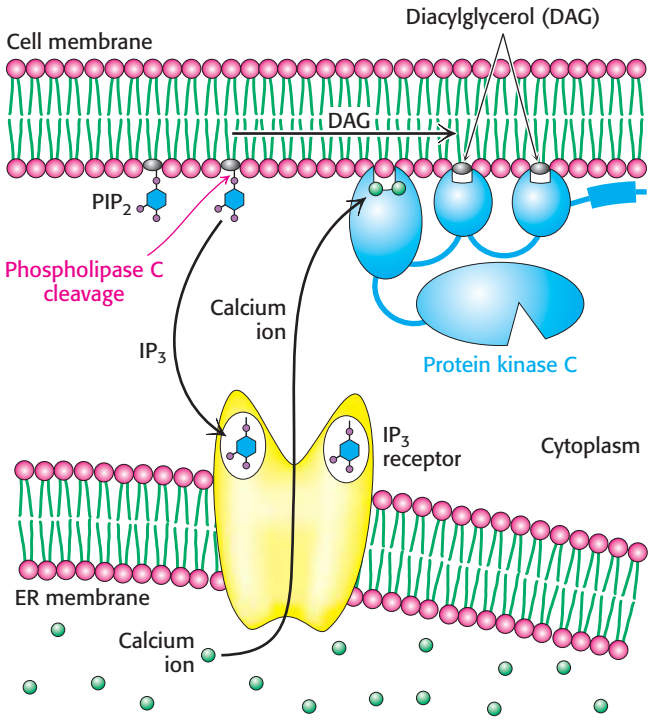


Figure 14.13 Phosphoinositide cascade. The cleavage of phosphatidylinositol 4,5-bisphosphate (PIP_2) into diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3) results in the release of calcium ions (owing to the opening of the IP_3 receptor ion channels) and the activation of protein kinase C (owing to the binding of protein kinase C to free DAG in the membrane). Calcium ions bind to protein kinase C and help facilitate its activation.

Calcium ion is a widely used second messenger

Calcium ion participates in many signaling processes in addition to the phosphoinositide cascade. Several properties of this ion account for its widespread use as an intracellular messenger. First, fleeting changes in Ca^{2+} concentration are readily detected. At steady state, intracellular levels of Ca^{2+} must be kept low to prevent the precipitation of carboxylated and phosphorylated compounds, which form poorly soluble salts with Ca^{2+}. Calcium ion levels are kept low by transport systems that extrude Ca^{2+} from the cytoplasm. Because of their action, the cytoplasmic concentration of Ca^{2+} is approximately 100 nM, several orders of magnitude lower than that of the extracellular medium. Given this low steady-state level, transient increases in Ca^{2+} concentration produced by signaling events can be readily sensed.

A second property of Ca^{2+} that makes it a highly suitable intracellular messenger is that it can bind tightly to proteins and induce substantial structural rearrangements. Calcium ions bind well to negatively charged oxygen atoms (from the side chains of glutamate and aspartate) and uncharged oxygen atoms (main-chain carbonyl groups and side-chain oxygen atoms from glutamine and asparagine; Figure 14.14). *The capacity of Ca^{2+} to be coordinated to multiple ligands—from six to eight oxygen atoms—enables it to cross-link different segments of a protein and induce significant conformational changes.*

Our understanding of the role of Ca^{2+} in cellular processes has been greatly enhanced by our ability to detect changes in Ca^{2+} concentrations inside cells and even monitor these changes in real time. This ability depends on the use of specially designed dyes such as Fura-2 that bind Ca^{2+} and change their fluorescent properties on Ca^{2+} binding. Fura-2 binds Ca^{2+} through appropriately positioned oxygen atoms (shown in red) within its structure.

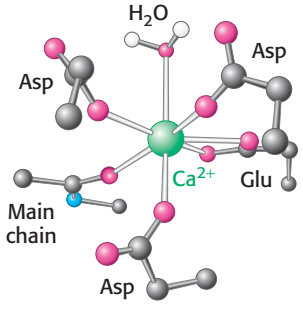
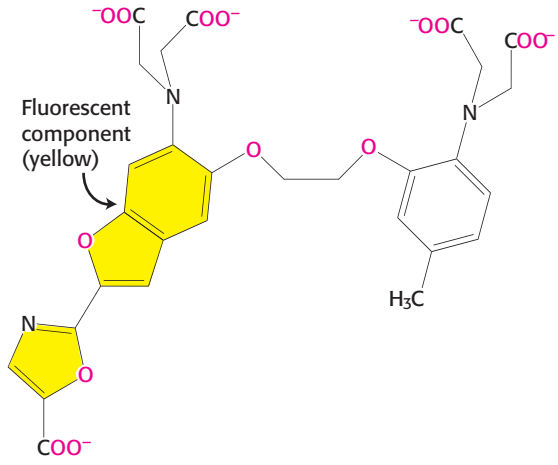
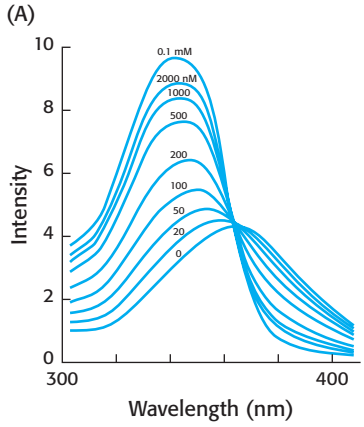


Figure 14.14 Calcium-binding site. In one common mode of binding, calcium is coordinated to six oxygen atoms of a protein and one (top) of water.



Fura-2

When such a dye is introduced into cells, changes in available Ca^{2+} concentration can be monitored with microscopes capable of detecting changes in fluorescence (Figure 14.15). Probes for sensing other second messengers such as cAMP also have been developed. These *molecular-imaging agents* are greatly enhancing our understanding of signal-transduction processes.



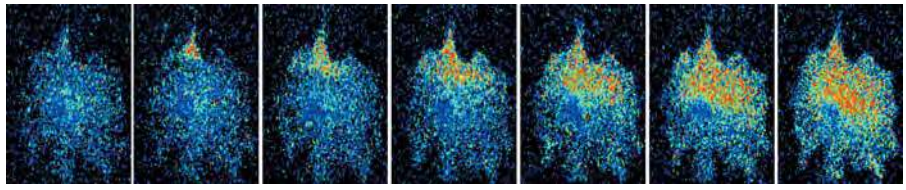


Figure 14.15 Calcium imaging. (A) The fluorescence spectra of the calcium-binding dye Fura-2 can be used to measure available calcium ion concentrations in solution and in cells. (B) A series of images show Ca^{2+} spreading across a cell. These images were obtained through the use of a fluorescent calcium-binding dye. The images are false colored: red represents high Ca^{2+} concentrations, and blue represents low Ca^{2+} concentrations. [(A) After S. J. Lippard and J. M. Berg, *Principles of Bioinorganic Chemistry* (University Science Books, 1994), p. 193; (B) courtesy of Dr. Masashi Isshiki, Department of Nephrology, University of Tokyo, and Dr. G. W. Anderson, Department of Cell Biology, University of Texas Southwestern Medical School.]

Calcium ion often activates the regulatory protein calmodulin

Calmodulin (CaM), a 17-kd protein with four Ca^{2+}-binding sites, serves as a calcium sensor in nearly all eukaryotic cells. *At cytoplasmic concentrations above about 500 nM, Ca^{2+} binds to and activates calmodulin.* Calmodulin is a member of the *EF-hand protein family.* The *EF hand* is a Ca^{2+}-binding motif that consists of a helix, a loop, and a second helix. This motif, originally discovered in the protein parvalbumin, was named the EF hand because the two key helices designated E and F in parvalbumin are positioned like the forefinger and thumb of the right hand (Figure 14.16). These two helices and the intervening loop form the Ca^{2+}-binding motif. Seven

oxygen atoms are coordinated to each Ca^{2+}, six from the protein and one from a bound water molecule. Calmodulin is made up of four EF-hand motifs, each of which can bind a single Ca^{2+} ion.

The binding of Ca^{2+} to calmodulin induces substantial conformational changes in its EF hands, exposing hydrophobic surfaces that can be used to bind other proteins. Using its two sets of two EF hands, calmodulin clamps down around specific regions of target proteins, usually exposed α helices with appropriately positioned hydrophobic and charged groups (Figure 14.17). The Ca^{2+}–calmodulin complex stimulates a wide variety of enzymes, pumps, and other target proteins by inducing structural rearrangements in these binding partners. An especially noteworthy set of targets are several *calmodulin-dependent protein kinases* (CaM kinases) that phosphorylate many different proteins and regulate fuel metabolism, ionic permeability, neurotransmitter synthesis, and neurotransmitter release. We see here a recurring theme in signal-transduction pathways: the concentration of a second messenger is increased (in this case, Ca^{2+}); the signal is sensed by a second-messenger-binding protein (in this case, calmodulin); and the second-messenger-binding protein acts to generate changes in enzymes (in this case, calmodulin-dependent kinases) that control effectors.

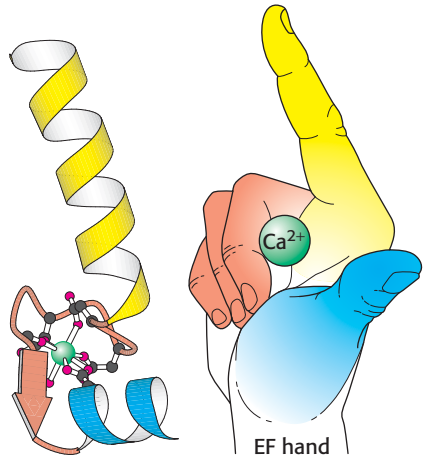


Figure 14.16 EF hand. Formed by a helix-loop-helix unit, an EF hand is a binding site for Ca^{2+} in many calcium-sensing proteins. Here, the E helix is yellow, the F helix is blue, and calcium is represented by the green sphere. *Notice* that the calcium ion is bound in a loop connecting two nearly perpendicular helices. [Drawn from 1CLL.pdb.]

14.2 Insulin Signaling: Phosphorylation Cascades Are Central to Many Signal-Transduction Processes

The signaling pathways that we have examined so far have activated a protein kinase as a downstream component of the pathway. We now turn to a class of signal-transduction pathways that are *initiated by receptors that include protein kinases as part of their structures*. The activation of these protein kinases sets in motion other processes that ultimately modify the effectors of these pathways.

An example is the signal-transduction pathway initiated by *insulin*, the hormone released in response to increased blood-glucose levels after a meal. In all of its detail, this multifaceted pathway is quite complex. Hence, we will focus solely on the major branch, which leads to the mobilization of glucose transporters to the cell surface. These transporters allow the cell to take up the glucose that is plentiful in the bloodstream after a meal.

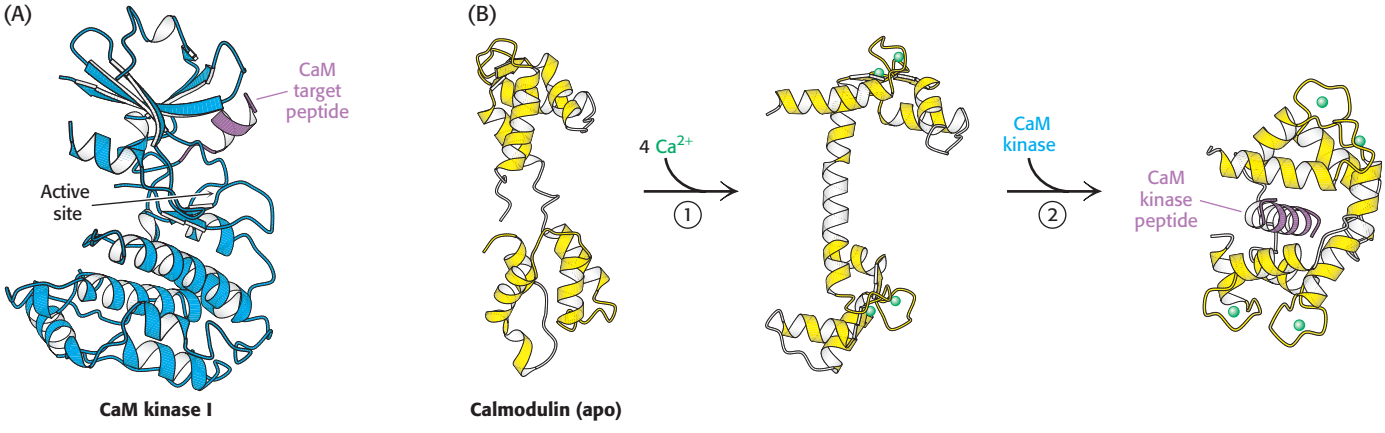


Figure 14.17 Calmodulin binds to α helices. (A) An α helix (purple) in CaM kinase I is a target for calmodulin. (B) On Ca^{2+} binding to the apo, or calcium-free, form of calmodulin (1), the two halves of calmodulin clamp down around the target helix (2), binding it through hydrophobic and ionic interactions. In CaM kinase I, this interaction allows the enzyme to adopt an active conformation. [Drawn from 1A06, 1CFD, 1CLL, and 1CM1.pdb.]

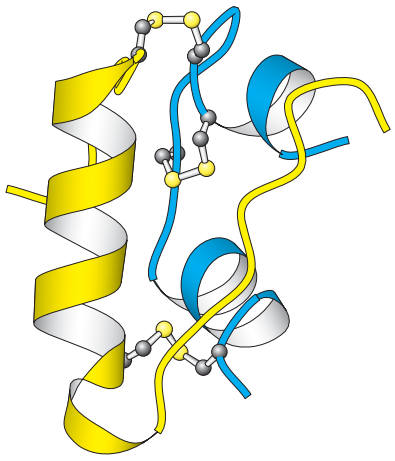


Figure 14.18 Insulin structure. *Notice* that insulin consists of two chains (shown in blue and yellow) linked by two interchain disulfide bonds. The α chain (blue) also has an intrachain disulfide bond. [Drawn from 1B2F.pdb.]

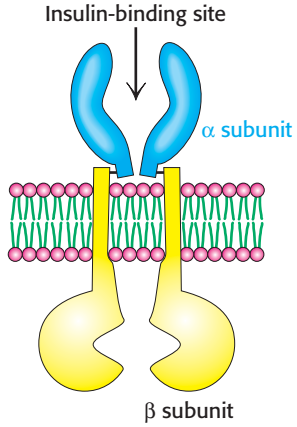
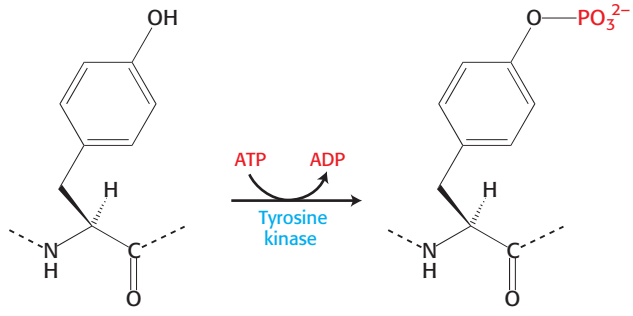


Figure 14.19 The insulin receptor. The receptor consists of two units, each of which consists of an α subunit and a β subunit linked by a disulfide bond. Two α subunits, which lie outside the cell, come together to form a binding site for insulin. Each β subunit lies primarily inside the cell and includes a protein kinase domain.

The insulin receptor is a dimer that closes around a bound insulin molecule

Insulin is a peptide hormone that consists of two chains, linked by three disulfide bonds (Figure 14.18). Its receptor has a quite different structure from that of the β-AR. The *insulin receptor* is a dimer of two identical units. Each unit consists of one α chain and one β chain linked to one another by a single disulfide bond (Figure 14.19). Each α subunit lies completely outside the cell, whereas each β subunit lies primarily inside the cell, spanning the membrane with a single transmembrane segment. The two α subunits move together to form a binding site for a single insulin molecule, a surprising occurrence because two different surfaces on the insulin molecule must interact with the two identical insulin-receptor chains. The moving together of the dimeric units in the presence of an insulin molecule sets the signaling pathway in motion. *The closing up of an oligomeric receptor or the oligomerization of monomeric receptors around a bound ligand is a strategy used by many receptors to initiate a signal, particularly by those containing a protein kinase.*

Each β subunit consists primarily of a protein kinase domain, homologous to protein kinase A. However, this kinase differs from protein kinase A in two important ways. First, the insulin-receptor kinase is a *tyrosine kinase;* that is, it catalyzes the transfer of a phosphoryl group from ATP to the hydroxyl group of tyrosine, rather than serine or threonine.



Because this tyrosine kinase is a component of the receptor itself, the insulin receptor is referred to as a *receptor tyrosine kinase*. Second, the insulin receptor kinase is in an inactive conformation when the domain is not covalently modified. The kinase is rendered inactive by the position of an unstructured loop (called the *activation loop*) that lies in the center of the structure.

Insulin binding results in the cross-phosphorylation and activation of the insulin receptor

When the two α subunits move together to surround an insulin molecule, the two protein kinase domains on the inside of the cell also are drawn together. Importantly, as they come together, the flexible activation loop of one kinase subunit is able to fit into the active site of the other kinase subunit within the dimer. With the two β subunits forced together, the kinase domains catalyze the addition of phosphoryl groups from ATP to tyrosine residues in the activation loops. When these tyrosine residues are phosphorylated, a striking conformational change takes place (Figure 14.20). The rearrangement of the activation loop converts the kinase into an active conformation. Thus, *insulin binding on the outside of the cell results in the activation of a membrane-associated kinase within the cell.*

The activated insulin-receptor kinase initiates a kinase cascade

On phosphorylation, the insulin-receptor tyrosine kinase is activated. Because the two units of the receptor are held in close proximity to one

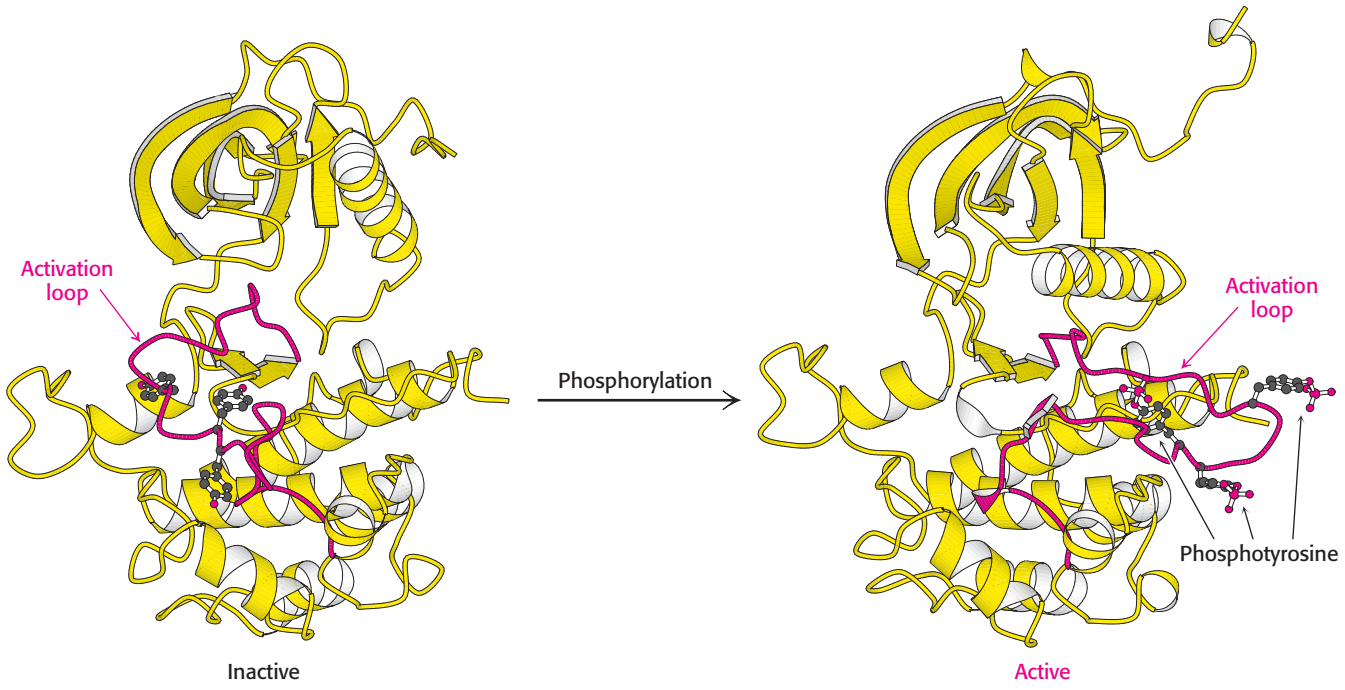


Figure 14.20 Activation of the insulin receptor by phosphorylation. The activation loop is shown in red in this model of the protein kinase domain of the β subunit of the insulin receptor. The unphosphorylated structure on the left is not catalytically active. *Notice* that, when three tyrosine residues in the activation loop are phosphorylated, the activation loop swings across the structure and the kinase structure adopts a more compact conformation. This conformation is catalytically active. [Drawn from 1IRK.pdb and 1IR3.pdb.]

another, additional sites within the receptor also are phosphorylated. These phosphorylated sites act as docking sites for other substrates, including a class of molecules referred to as *insulin-receptor substrates* (IRS; Figure 14.21). IRS-1 and IRS-2 are two homologous proteins with a common modular structure (Figure 14.22). The amino-terminal part includes a

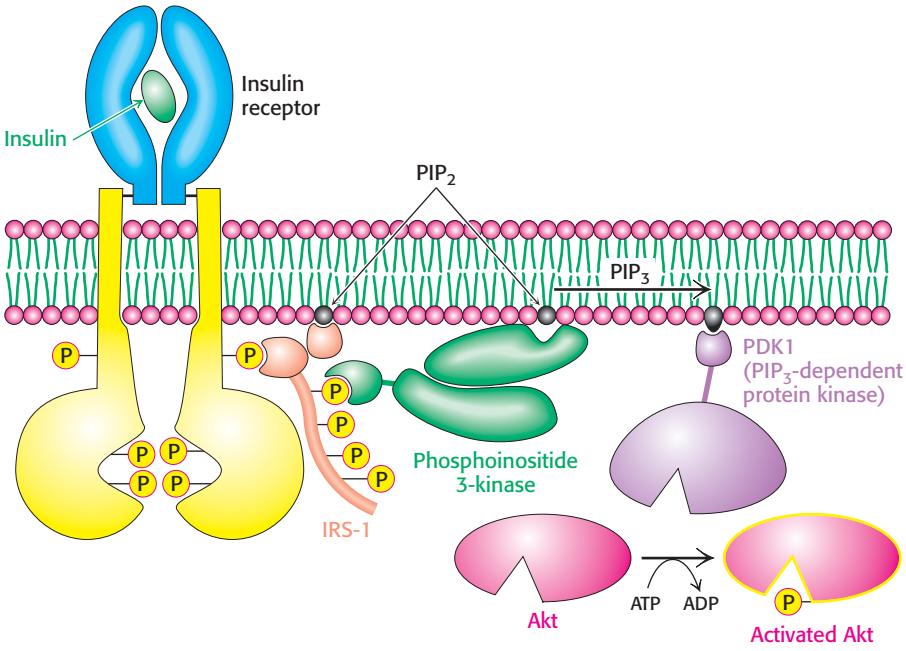


Figure 14.21 Insulin signaling. The binding of insulin results in the cross-phosphorylation and activation of the insulin receptor. Phosphorylated sites on the receptor act as binding sites for insulin-receptor substrates such as IRS-1. The lipid kinase phosphoinositide 3-kinase binds to phosphorylated sites on IRS-1 through its regulatory domain, then converts PIP_2 into PIP_3. Binding to PIP_3 activates PIP_3-dependent protein kinase, which phosphorylates and activates kinases such as Akt1. Activated Akt1 can then diffuse throughout the cell to continue the signal-transduction pathway.

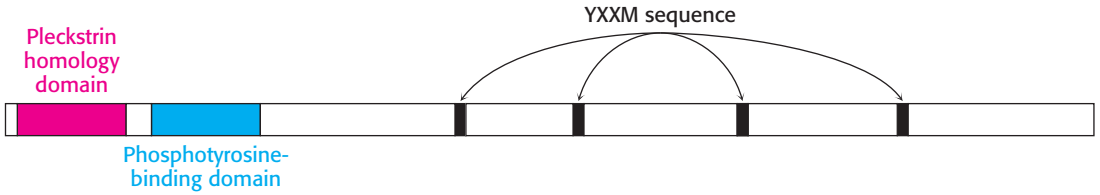


Figure 14.22 The modular structure of insulin-receptor substrates IRS-1 and IRS-2. This schematic view represents the amino acid sequence common to IRS-1 and IRS-2. Each protein contains a pleckstrin homology domain (which binds phosphoinositide lipids), a phosphotyrosine-binding domain, and four sequences that approximate Tyr-X-X-Met (YXXM). The four sequences are phosphorylated by the insulin-receptor tyrosine kinase.

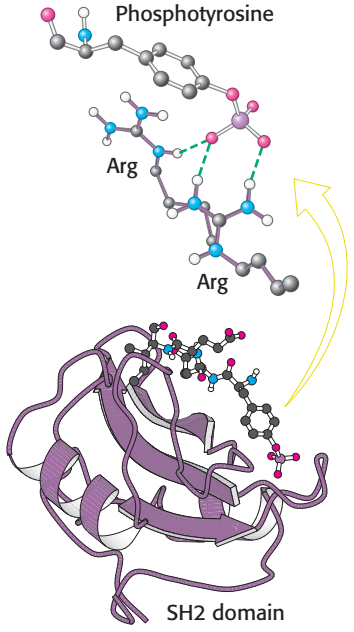


Figure 14.23 Structure of the SH2 domain. The domain is shown bound to a phosphotyrosine-containing peptide. *Notice* at the top that the negatively charged phosphotyrosine residue interacts with two arginine residues that are conserved in essentially all SH2 domains. [Drawn from 1SPS.pdb.]

pleckstrin homology domain, which binds phosphoinositide, and a phosphotyrosine-binding domain. These domains act together to anchor the IRS protein to the insulin receptor and the associated membrane. Each IRS protein contains four sequences that approximate the form Tyr-X-X-Met. These sequences are also substrates for the activated insulin-receptor kinase. When the tyrosine residues within these sequences are phosphorylated to become phosphotyrosine residues, IRS molecules can act as *adaptor proteins:* they are not enzymes but serve to tether the downstream components of this signaling pathway to the membrane.

Phosphotyrosine residues, such as those in the IRS proteins, are recognized most often by *Src homology 2* (SH2) domains (Figure 14.23). These domains, present in many signal-transduction proteins, bind to stretches of polypeptide that contain phosphotyrosine residues. Each specific SH2 domain shows a binding preference for phosphotyrosine in a particular sequence context. Which proteins contain SH2 domains that bind to phosphotyrosine-containing sequences in the IRS proteins? The most important of them are in a class of lipid kinases, called phosphoinositide 3-kinases (PI3Ks), that add a phosphoryl group to the 3-position of inositol in phosphatidylinositol 4,5-bisphosphate (PIP_2; Figure 14.24). These enzymes are heterooligomers that consist of 110-kd catalytic subunits and 85-kd regulatory subunits. Through SH2 domains in the regulatory subunits, these enzymes bind to the IRS proteins and are drawn to the membrane where they can phosphorylate PIP_2 to form phosphatidyl-inositol 3,4,5-trisphosphate (PIP_3). PIP_3, in turn, activates a protein kinase, PDK1, by virtue of a pleckstrin homology domain present in this kinase that is specific for PIP_3. The activated PDK1 phosphorylates and activates Akt, another protein kinase. Akt is not membrane anchored and moves through the cell to phosphorylate targets that include components that control the trafficking of the glucose receptor GLUT4 to the cell surface as well as enzymes that stimulate glycogen synthesis (Section 21.4).

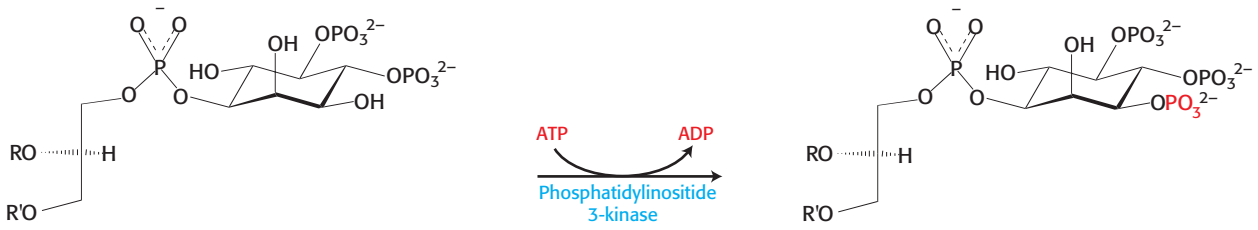


Figure 14.24 Action of a lipid kinase in insulin signaling. Phosphorylated IRS-1 and IRS-2 activate the enzyme phosphatidylinositide 3-kinase, an enzyme that converts PIP_2 into PIP_3.

The cascade initiated by the binding of insulin to the insulin receptor is summarized in Figure 14.25. The signal is amplified at several stages along this pathway. Because the activated insulin receptor itself is a protein kinase, each activated receptor can phosphorylate multiple IRS molecules. Activated enzymes further amplify the signal in at least two of the subsequent steps. Thus, a small increase in the concentration of circulating insulin can produce a robust intracellular response. Note that, although the insulin pathway described here may seem complicated, it is substantially less elaborate than the full signaling network initiated by insulin.

Insulin signaling is terminated by the action of phosphatases

We have seen that the activated G protein promotes its own inactivation by the release of a phosphoryl group from GTP. In contrast, proteins phosphorylated on serine, threonine, or tyrosine residues are extremely stable kinetically. Specific enzymes, called protein phosphatases, are required to hydrolyze these phosphorylated proteins and return them to their initial states. Similarly, lipid phosphatases are required to remove phosphoryl groups from inositol lipids that had been activated by lipid kinases. In insulin signaling, three classes of enzymes are of particular importance in shutting off the signaling pathway: protein tyrosine phosphatases that remove phosphoryl groups from tyrosine residues on the insulin receptor and the IRS adaptor proteins, lipid phosphatases that hydrolyze PIP_3 to PIP_2, and protein serine phosphatases that remove phosphoryl groups from activated protein kinases such as Akt. Many of these phosphatases are activated or recruited as part of the response to insulin. Thus, the binding of the initial signal sets the stage for the eventual termination of the response.

14.3 EGF Signaling: Signal-Transduction Pathways Are Poised to Respond

Our consideration of the signal-transduction cascades initiated by epinephrine and insulin included examples of how components of signal-transduction pathways are poised for action, ready to be activated by minor modifications. For example, G-protein subunits require only the binding of GTP in exchange for GDP to transmit a signal. This exchange reaction is thermodynamically favorable, but it is quite slow in the absence of an appropriate activated 7TM receptor. Similarly, the tyrosine kinase domains of the dimeric insulin receptor are ready for phosphorylation and activation but require insulin bound between two α subunits to draw the activation loop of one tyrosine kinase into the active site of a partner tyrosine kinase to initiate this process.

We now examine a signal-transduction pathway that reveals another clear example of how these signaling cascades are poised to respond. This pathway is activated by the signal molecule *epidermal growth factor* (EGF). Like that of the insulin receptor, the initiator of this pathway is a receptor tyrosine kinase. Both the extracellular and the intracellular domains of this receptor are ready for action, held in check only by a specific structure that prevents receptors from coming together. Furthemore, in the EGF pathway, we will encounter several additional classes of signaling components that participate in many other signaling networks.

EGF binding results in the dimerization of the EGF receptor

Epidermal growth factor is a 6-kd polypeptide that stimulates the growth of epidermal and epithelial cells (Figure 14.26). The *EGF receptor* (EGFR),

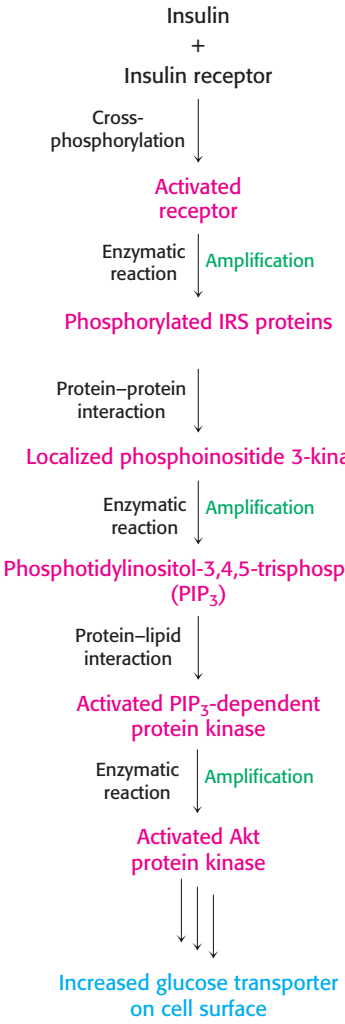
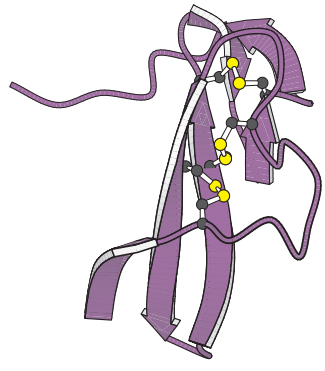


Figure 14.25 Insulin signaling pathway. Key steps in the signal-transduction pathway initiated by the binding of insulin to the insulin receptor.



Epidermal growth factor (EGF)

Figure 14.26 Structure of epidermal growth factor. *Notice* that three intrachain disulfide bonds stabilize the compact three-dimensional structure of the growth factor. [Drawn from 1EGF.pdb.]

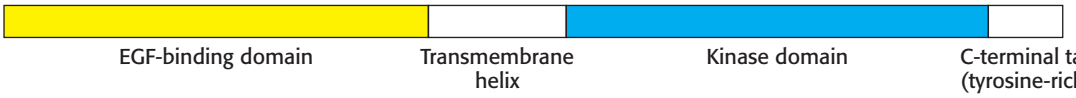


Figure 14.27 Modular structure of the EGF receptor. This schematic view of the amino acid sequence of the EGF receptor shows the EGF-binding domain that lies outside the cell, a single transmembrane helix-forming region, the intracellular tyrosine kinase domain, and the tyrosine-rich domain at the carboxyl terminus.

like the insulin receptor, is a dimer of two identical subunits. Each subunit contains an intracellular protein tyrosine kinase domain that participates in cross-phosphorylation reactions (Figure 14.27). Unlike those of the insulin receptor, however, these units exist as monomers until they bind EGF. Moreover, each EGF receptor monomer binds a single molecule of EGF in its extracellular domain (Figure 14.28). Thus the dimer binds two ligand molecules, in contrast with the insulin-receptor dimer, which binds only one ligand. Note that each EGF molecule lies far away from the dimer interface. This interface includes a so-called *dimerization arm* from each monomer that reaches out and inserts into a binding pocket on the other monomer.

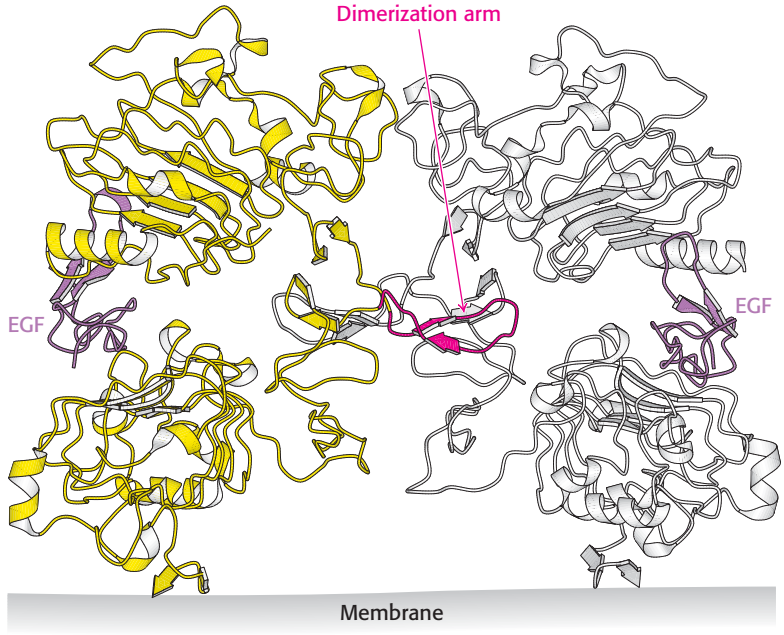


Figure 14.28 EGF receptor dimerization. The structure of the extracellular region of the EGF receptor is shown bound to EGF. *Notice* that the structure is dimeric with one EGF molecule bound to each receptor molecule and that the dimerization is mediated by a dimerization arm that extends from each receptor molecule. [Drawn from 1IVO.pdb.]

Although this structure nicely reveals the interactions that support the formation of a receptor dimer favoring cross-phosphorylation, it raises another question: Why doesn't the receptor dimerize and signal in the absence of EGF? This question has been addressed by examining the structure of the EGF receptor in the absence of bound ligand (Figure 14.29). This structure is, indeed, monomeric and each monomer is in a conformation that is quite different from that observed in the ligand-bound dimer. In particular, the dimerization arm binds to a domain *within the same monomer* that holds the receptor in a closed configuration. In essence, the

receptor is poised in a spring-loaded conformation held in position by the contact between the interaction loop and another part of the structure, ready to bind ligand and change into a conformation active for dimerization and signaling.

This observation suggests that a receptor that exists in the extended conformation even in the absence of bound ligand would be constitutively active. Remarkably, such a receptor exists. This receptor, Her2, is approximately 50% identical in amino acid sequence with the EGF receptor and has the same domain structure. Her2 does not bind any known ligand, yet crystallographic studies reveal that it adopts an extended structure very similar to that observed for the ligand-bound EGF receptor. Under normal conditions, Her2 forms heterodimers with the EGF receptor and other members of the EGF receptor family and participates in cross-phosphorylation reactions with these receptors. Her2 is overexpressed in some cancers, presumably contributing to tumor growth by forming homodimers that signal even in the absence of ligand. We will return to Her2 when we consider approaches to cancer treatment based on knowledge of signaling pathways (Section 14.5).

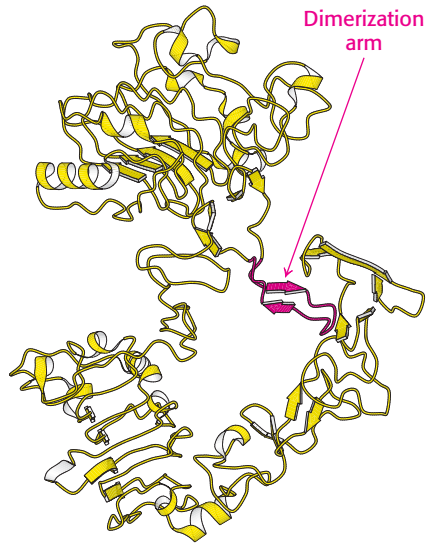


Figure 14.29 Structure of the unactivated EGF receptor. The extracellular domain of the EGF receptor is shown in the absence of bound EGF. *Notice* that the dimerization arm is bound to a part of the receptor that makes it unavailable for interaction with the other receptor. [Drawn from 1NQL.pdb.]

The EGF receptor undergoes phosphorylation of its carboxyl-terminal tail

Like the insulin receptor, the EGF receptor undergoes cross-phosphorylation of one unit by another unit within a dimer. However, unlike that of the insulin receptor, the site of this phosphorylation is not within the activation loop of the kinase, but rather in a region that lies on the C-terminal side of the kinase domain. As many as five tyrosines residues in this region are phosphorylated. The dimerization of the EGF receptor brings the C-terminal region on one receptor into the active site of its partner's kinase. The kinase itself is in an active conformation without phosphorylation, revealing again how this signaling system is poised to respond.

EGF signaling leads to the activation of Ras, a small G protein

The phosphotyrosines on the EGF receptors act as docking sites for SH2 domains on other proteins. The intracellular signaling cascade begins with the binding of *Grb-2,* a key adaptor protein that contains one SH2 domain and two *Src homology 3* (SH3) domains. On phosphorylation of the receptor, the SH2 domain of Grb-2 binds to the phosphotyrosine residues of the receptor tyrosine kinase. Through its two SH3 domains, Grb-2 then binds polyproline-rich polypeptides within a protein called *Sos*. Sos, in turn, binds to Ras and activates it. A very prominent signal-transduction component, Ras is a member of a class of proteins called the *small G proteins.* Like the G proteins described in Section 14.1, the small G proteins contain bound GDP in their unactivated forms. Sos opens up the nucleotide-binding pocket of Ras, allowing GDP to escape and GTP to enter in its place. Because of its effect on Ras, Sos is referred to as a *guanine-nucleotide-exchange factor* (GEF). Thus, the binding of EGF to its receptor leads to the conversion of Ras into its GTP form through the intermediacy of Grb-2 and Sos (Figure 14.30).

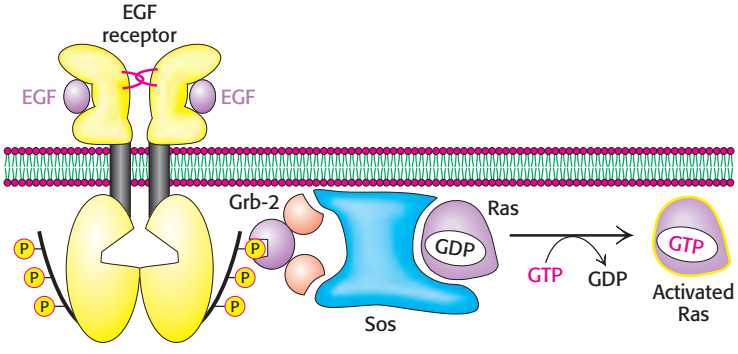


Figure 14.30 Ras activation mechanism. The dimerization of the EGF receptor due to EGF binding leads to the phosphorylation of the C-terminal tails of the receptor, the subsequent recruitment of Grb-2 and Sos, and the exchange of GTP for GDP in Ras. This signal-transduction pathway results in the conversion of Ras into its activated GTP-bound form.

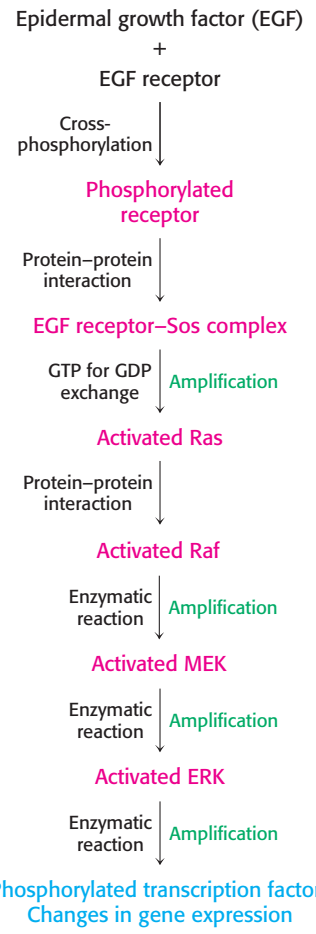


Figure 14.31 EGF signaling pathway. The key steps in the pathway initiated by EGF binding to the EGF receptor. A kinase cascade leads to the phosphorylation of transcription factors and concomitant changes in gene expression.

Activated Ras initiates a protein kinase cascade

Ras changes conformation when it is transformed from its GDP into its GTP form. In the GTP form, Ras binds other proteins, including a protein kinase termed *Raf*. When bound to Ras, Raf undergoes a conformational change that activates the Raf protein kinase domain. Both Ras and Raf are anchored to the membrane through covalently bound isoprene lipids. Activated Raf then phosphorylates other proteins, including protein kinases termed MEKs. In turn, MEKs activate kinases called *extracellular signal-regulated kinases* (ERKs). ERKs then phosphorylate numerous substrates, including transcription factors in the nucleus as well as other protein kinases. The complete flow of information from the arrival of EGF at the cell surface to changes in gene expression is summarized in Figure 14.31.

Small G proteins, or small GTPases, constitute a large superfamily of proteins—grouped into subfamilies called Ras, Rho, Arf, Rab, and Ran—that play a major role in a host of cell functions including growth, differentiation, cell motility, cytokinesis, and the transport of materials throughout the cell (Table 14.2). As with the heterotrimeric G proteins, the small G proteins cycle between an active GTP-bound form and an inactive GDP-bound form. They differ from the heterotrimeric G proteins in being smaller (20–25 kd versus 30–35 kd) and monomeric. Nonetheless, the two families are related by divergent evolution, and small G proteins have many key mechanistic and structural motifs in common with the G_α subunit of the heterotrimeric G proteins.

EGF signaling is terminated by protein phosphatases and the intrinsic GTPase activity of Ras

Because so many components of the EGF signal-transduction pathway are activated by phosphorylation, we can expect protein phosphatases to play key roles in the termination of EGF signaling. Indeed, crucial phosphatases remove phosphoryl groups from tyrosine residues on the EGF receptor and from serine, threonine, and tyrosine residues in the protein kinases that participate in the signaling cascade. The signaling process itself sets in motion the events that activate many of these phosphatases. Consequently, signal activation also initiates signal termination.

Like the G proteins activated by 7TM receptors, Ras possesses intrinsic GTPase activity. Thus, the activated GTP form of Ras spontaneously converts into the inactive GDP form. The rate of conversion can be accelerated in the presence of *GTPase-activating proteins* (GAPs), proteins that interact with small G proteins in the GTP form and facilitate GTP hydrolysis. Thus, the lifetime of activated Ras is regulated by accessory proteins in the cell. The GTPase activity of Ras is crucial for shutting off signals leading to cell growth, and so it is not surprising that mutations in Ras are found in many types of cancer, as will be discussed in Section 14.5.

Table 14.2 Ras superfamily of GTPases

Subfamily	Function
Ras	Regulates cell growth through serine–threonine protein kinases
Rho	Reorganizes cytoskeleton through serine–threonine protein kinases
Arf	Activates the ADP-ribosyltransferase of the cholera toxin A subunit; regulates vesicular trafficking pathways; activates phospholipase D
Rab	Plays a key role in secretory and endocytotic pathways
Ran	Functions in the transport of RNA and protein into and out of the nucleus

14.4 Many Elements Recur with Variation in Different Signal-Transduction Pathways

We can begin to make sense of the complexity of signal-transduction pathways by taking note of several common themes that have appeared consistently in the pathways described in this chapter and underlie many additional signaling pathways not considered herein.

1. *Protein kinases are central to many signal-transduction pathways.* Protein kinases are central to all three signal-transduction pathways described in this chapter. In the epinephrine-initiated pathway, cAMP-dependent protein kinase (PKA) lies at the end of the pathway, transducing information represented by an increase in cAMP concentration into covalent modifications that alter the activity of key metabolic enzymes. In the insulin- and EGF-initiated pathways, the receptors themselves are protein kinases and several additional protein kinases participate downstream in the pathways. Signal amplification due to protein kinase cascades is a feature common to all three pathways. Although not presented in this chapter, protein kinases often phosphorylate multiple substrates and are thus able to generate a diversity of responses.

2. *Second messengers participate in many signal-transduction pathways.* We have encountered several second messengers, including cAMP, Ca^{2+} IP_3, and the lipid DAG. Because second messengers are activated by enzymes or by the opening of ion channels, their concentrations can be tremendously amplified compared with the signals that lead to their generation. Specialized proteins sense the concentrations of these second messengers and continue the flow of information along signal-transduction pathways.

The second messengers that we have seen recur in many additional signal-transduction pathways. For example, in a consideration of the sensory systems in Chapter 33, we will see how Ca^{2+}-based signaling and cyclic nucleotide-based signaling play key roles in vision and olfaction.

3. *Specialized domains that mediate specific interactions are present in many signaling proteins.* The "wiring" of many signal-transduction pathways is based on particular protein domains that mediate the interactions between protein components of a particular signaling cascade. We have encountered several of them, including: pleckstrin homology domains, which facilitate protein interactions with the lipid PIP_3; SH2 domains, which mediate interactions with polypeptides containing phosphorylated tyrosine residues; and SH3 domains, which interact with peptide sequences that contain multiple proline residues. Many other such domain families exist. In many cases, individual members of each domain family have unique features that allow them to bind to their targets only within a particular sequence context, making them specific for a given signaling pathway and avoiding unwanted cross-talk. *Signal-transduction pathways have evolved in large part by the incorporation of DNA fragments encoding these domains into genes encoding pathway components.*

The presence of these domains is tremendously helpful to scientists trying to unravel signal-transduction pathways. When a protein in a signal-transduction pathway is identified, its amino acid sequence can be analyzed for the presence of these specialized domains by the methods described in Chapter 6. If one or more domains of known function is found, it is often possible to develop clear hypotheses about potential binding partners and signal-transduction mechanisms.

14.5 Defects in Signal-Transduction Pathways Can Lead to Cancer and Other Diseases

In light of their complexity, it comes as no surprise that signal-transduction pathways occasionally fail, leading to pathological or disease states. Cancer, a set of diseases characterized by uncontrolled or inappropriate cell growth, is strongly associated with defects in signal-transduction proteins. Indeed, the study of cancer, particularly cancers caused by certain viruses, has contributed greatly to our understanding of signal-transduction proteins and pathways.

For example, Rous sarcoma virus is a retrovirus that causes sarcoma (a cancer of tissues of mesodermal origin such as muscle or connective tissue) in chickens. In addition to the genes necessary for viral replication, this virus carries a gene termed v-*src*. The v-*src* gene is an *oncogene;* it leads to the generation of cancerlike characteristics in susceptible cell types. The protein encoded by the v-*src* gene, v-Src, is a protein tyrosine kinase that includes SH2 and SH3 domains. The v-Src protein is similar in amino acid sequence to a protein normally found in chicken-muscle cells referred to as c-Src (for cellular Src; Figure 14.32A). The c-*src* gene does not induce cell transformation and is termed a *proto-oncogene,* referring to the fact that this gene, when mutated, can be converted into an oncogene. The protein that it encodes is a signal-transduction protein that regulates cell growth.

Why is the biological activity of the v-Src protein so different from that of c-Src? c-Src contains a key tyrosine residue near its C-terminal end that, when phosphorylated, is bound intramolecularly by the upstream SH2 domain (Figure 14.32B). This interaction maintains the kinase domain in an inactive conformation. However, in v-Src, the C-terminal 19 amino acids of c-Src are replaced by a completely different stretch of 11 amino acids that lacks this critical tyrosine residue. Thus, v-Src is always active and can promote unregulated cell growth. Since the discovery of Src, many other mutated protein kinases have been identified as oncogenes.

The gene encoding Ras, a component of the EGF-initiated pathway, is one of the genes most commonly mutated in human tumors. Mammalian cells contain three 21-kd Ras proteins (H-, K-, and N-Ras), each of which cycles between inactive GDP and active GTP forms. The most common mutations in tumors lead to a loss of the ability to hydrolyze GTP. Thus, the Ras protein is trapped in the "on" position and continues to stimulate cell growth, even in the absence of a continuing signal.

Other genes can contribute to cancer development only when both copies of the gene normally present in a cell are deleted or otherwise damaged. Such genes are called *tumor-suppressor genes*. For example, genes for some of the phosphatases that participate in the termination of EGF signaling are tumor suppressors. Without any functional phosphatase present, EGF signaling persists once initiated, stimulating inappropriate cell growth.

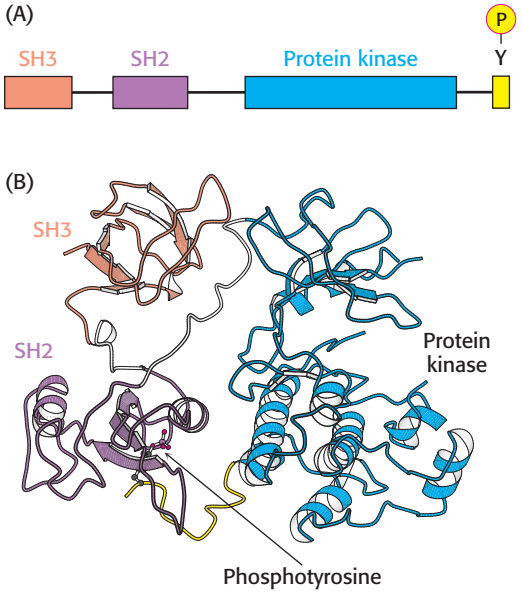


Figure 14.32 Src structure. (A) Cellular Src includes an SH3 domain, an SH2 domain, a protein kinase domain, and a carboxyl-terminal tail that includes a key tyrosine residue. (B) Structure of c-Src in an inactivated form with the key tyrosine residue phosphorylated. *Notice* how the three domains work together to keep the enzyme in an inactive conformation: phosphotyrosine residue is bound in the SH2 domain and the linker between the SH2 domain and the protein kinase domain is bound by the SH3 domain. [Drawn from 2PTK.pdb.]

Monoclonal antibodies can be used to inhibit signal-transduction pathways activated in tumors

Mutated or overexpressed receptor tyrosine kinases are frequently observed in tumors. For instance, the epidermal-growth-factor receptor (EGFR) is overexpressed in some human epithelial cancers, including breast, ovarian, and colorectal cancer.

Because some small amount of the receptor can dimerize and activate the signaling pathway even without binding to EGF, overexpression of the receptor increases the likelihood that a "grow and divide" signal will be inappropriately sent to the cell. This understanding of cancer-related signal-transduction pathways has led to a therapeutic approach that targets the EGFR. The strategy is to produce monoclonal antibodies to the extracellular domains of the offending receptors. One such antibody, cetuximab (Erbitux), has effectively targeted the EGFR in colorectal cancers. Cetuximab inhibits the EGFR by competing with EGF for the binding site on the receptor. Because the antibody sterically blocks the change in conformation that exposes the dimerization arm, the antibody itself cannot induce dimerization. The result is that the EGFR-controlled pathway is not initiated.

Cetuximab is not the only monoclonal antibody that has been developed to target a receptor tyrosine kinase. Trastuzumab (Herceptin) inhibits another EGFR family member, Her2, that is overexpressed in approximately 30% of breast cancers. Recall that this protein can signal even in the absence of ligand; so it is especially likely that overexpression will stimulate cell proliferation. Breast-cancer patients are now being screened for Her2 overexpression and treated with Herceptin as appropriate. Thus, this cancer treatment is tailored to the genetic characteristics of the tumor.

Protein kinase inhibitors can be effective anticancer drugs

The widespread occurrence of overactive protein kinases in cancer cells suggests that molecules that inhibit these enzymes might act as antitumor agents. For example, more than 90% of patients with chronic myelogenous leukemia (CML) show a specific chromosomal defect in cancer cells (Figure 14.33). The translocation of genetic material between chromosomes 9 and 22 causes the c-*abl* gene, which encodes a tyrosine kinase of the Src family, to be inserted into the *bcr* gene on chromosome 22. The result is the production of a fusion protein called Bcr-Abl that consists primarily of sequences for the c-Abl kinase. However, the *bcr-abl* gene is not regulated appropriately; it is expressed at higher levels than that of the gene encoding the normal c-Abl kinase, stimulating a growth-promoting pathway. Because of this overexpression, leukemia cells express a unique target for chemotherapy. A specific inhibitor of the Bcr-Abl kinase, Gleevec (STI-571, imatinib mesylate), has proved to be a highly effective treatment for patients suffering from CML. This approach to cancer chemotherapy is fundamentally distinct from most approaches, which target all rapidly growing cells, including normal ones. *Thus, our understanding of signal-transduction pathways is leading to conceptually new disease treatments.*

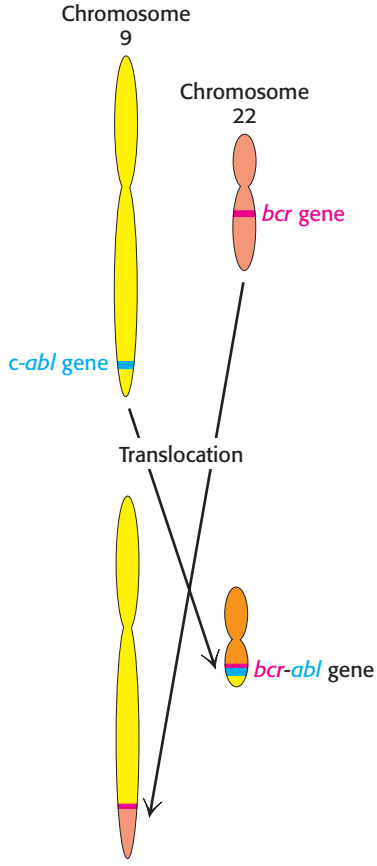


Figure 14.33 Formation of the *bcr-abl* gene by translocation. In chronic myelogenous leukemia, parts of chromosomes 9 and 22 are reciprocally exchanged, causing the *bcr* and *abl* genes to fuse. The protein kinase encoded by the *bcr-abl* gene is expressed at higher levels in cells having this translocation than is the *c-abl* gene in normal cells.

Cholera and whooping cough are due to altered G-protein activity

Although defects in signal-transduction pathways have been most extensively studied in the context of cancer, such defects are important in many other diseases. Cholera and whooping cough are two pathologies of the G-protein-dependent signal pathways. Let us first consider the mechanism of action of the cholera toxin, secreted by the intestinal bacterium *Vibrio cholerae*. Cholera is a potentially life threatening, acute diarrheal disease transmitted through contaminated water and food. It causes the voluminous secretion of electrolytes and fluids from the intestines of infected persons. The cholera toxin, *choleragen,* is a protein composed of two functional units—a β subunit that binds to G_{M1} gangliosides (p. 765) of the intestinal epithelium and a catalytic A subunit that enters the cell. The A subunit catalyzes the covalent modification of a $G_{\alpha s}$ protein: the α subunit is modified by the attachment of an ADP-ribose to an arginine

residue. This modification stabilizes the GTP-bound form of $G_{\alpha s}$, trapping the molecule in its active conformation. The active G protein, in turn, continuously activates protein kinase A. PKA opens a chloride channel and inhibits sodium absorption by the Na^+–H^+ exchanger by phosphorylating both the channel and the exchanger. The net result of the phosphorylation is an excessive loss of NaCl and the loss of large amounts of water into the intestine. Patients suffering from cholera may pass as much as twice their body weight in fluid in 4 to 6 days. Treatment consists of rehydration with a glucose–electrolyte solution.

Whereas cholera is a result of a G protein trapped in the active conformation, causing the signal-transduction pathway to be perpetually stimulated, pertussis, or whooping cough, is a result of the opposite situation. Pertussis toxin also adds an ADP-ribose moiety—in this case, to a $G_{\alpha i}$ protein, a G_α protein that inhibits adenylate cyclase, closes Ca^{2+} channels, and opens K^+ channels. The effect of this modification, however, is to lower the G protein's affinity for GTP, effectively trapping it in the "off" conformation. The pulmonary symptoms have not yet been traced to a particular target of the $G_{\alpha i}$ protein. Pertussis toxin is secreted by *Bordetella pertussis,* the bacterium responsible for whooping cough.

Summary

In human beings and other multicellular organisms, specific signal molecules are released from cells in one organ and are sensed by cells in other organs throughout the body. The message initiated by an extracellular ligand is converted into specific changes in metabolism or gene expression by means of often complex networks referred to as signal-transduction pathways. These pathways amplify the initial signal and lead to changes in the properties of specific effector molecules.

14.1 Heterotrimeric G Proteins Transmit Signals and Reset Themselves

Epinephrine binds to a cell-surface protein called the β-adrenergic receptor. This receptor is a member of the seven-transmembrane-helix receptor family, so named because each receptor has seven α helices that span the cell membrane. When epinephrine binds to the β-adrenergic receptor on the outside of the cell, the receptor undergoes a conformational change that is sensed inside the cell by a signaling protein termed a heterotrimeric G protein. The α subunit of the G protein exchanges a bound GDP molecule for GTP and concomitantly releases the heterodimer consisting of the β and γ subunits. The α subunit in the GTP form then binds to adenylate cyclase and activates it, leading to an increase in the concentration of the second messenger cyclic AMP. This increase in cyclic AMP concentration, in turn, activates protein kinase A. Other 7TM receptors also signal through heterotrimeric G proteins, although these pathways often include enzymes other than adenylate cyclase. One prominent pathway, the phosphoinositide pathway, leads to the activation of phospholipase C, which cleaves a membrane lipid to produce two secondary messengers, diacylglycerol and inositol 1,4,5-trisphosphate. An increased IP_3 concentration leads to the release of calcium ion, another important second messenger, into the cell. G-protein signaling is terminated by the hydrolysis of the bound GTP to GDP.

14.2 Insulin Signaling: Phosphorylation Cascades Are Central to Many Signal-Transduction Processes

Protein kinases are key components in many signal-transduction pathways, including some for which the protein kinase is an integral component of the initial receptor. An example of such a receptor is the membrane tyrosine kinase bound by insulin. Insulin binding causes one subunit within the dimeric receptor to phosphorylate specific tyrosine residues in the other subunit. The resulting conformational changes dramatically increase the kinase activity of the receptor. The activated receptor kinase initiates a kinase cascade that includes both lipid kinases and protein kinases. This cascade eventually leads to the mobilization of glucose transporters to the cell surface, increasing glucose uptake. Insulin signaling is terminated through the action of phosphatases.

14.3 EGF Signaling: Signal-Transduction Systems Are Poised to Respond

Only minor modifications are necessary to transform many signal-transduction proteins from their inactive into their active forms. Epidermal growth factor also signals through a receptor tyrosine kinase. EGF binding induces a conformational change that allows receptor dimerization and cross-phosphorylation. The phosphorylated receptor binds adaptor proteins that mediate the activation of Ras, a small G protein. Activated Ras initiates a protein kinase cascade that eventually leads to the phosphorylation of transcription factors and changes in gene expression. EGF signaling is terminated by the action of phosphatases and the hydrolysis of GTP by Ras.

14.4 Many Elements Recur with Variation in Different Signal-Transduction Pathways

Protein kinases are components of many signal-transduction pathways, both as components of receptors and in other roles. Second messengers, including cyclic nucleotides, calcium, and lipid derivatives, are common in many signaling pathways. The changes in the concentrations of second messengers are often much larger than the changes associated with the initial signal owing to amplification processes. Small domains that recognize phosphotyrosine residues or specific lipids are present in many signaling proteins and are essential to determining the specificity of interactions.

14.5 Defects in Signal-Transduction Pathways Can Lead to Cancer and Other Diseases

Genes encoding components of signal-transduction pathways that control cell growth are often mutated in cancer. Some genes can be mutated to forms called oncogenes that are active regardless of appropriate signals. Monoclonal antibodies directed against cell-surface receptors that participate in signaling have been developed for use in cancer treatment. Our understanding of the molecular basis of cancer is leading to the development of anticancer drugs directed against specific targets, such as the specific kinase inhibitor Gleevec.

Key Terms

primary messenger (p. 402)
ligand (p. 402)
second messenger (p. 402)
cross talk (p. 403)
β-adrenergic receptor (β-AR) (p. 404)
seven-transmembrane-helix (7TM) receptor (p. 404)
rhodopsin (p. 404)

G protein (p. 405)
G-protein-coupled receptor (GPCR) (p. 406)
adenylate cyclase (p. 406)
protein kinase A (PKA) (p. 406)
β-adrenergic receptor kinase (p. 408)
phosphoinositide cascade (p. 408)
phosphatidylinositol 4,5-bisphosphate (PIP_2) (p. 408)
phospholipase C (p. 408)
protein kinase C (PKC) (p. 409)
calmodulin (CaM) (p. 410)
EF hand (p. 410)
calmodulin-dependent protein kinase (CaM kinase) (p. 411)
insulin (p. 411)
insulin receptor (p. 412)
tyrosine kinase (p. 412)
receptor tyrosine kinase (p. 412)
insulin-receptor substate (IRS) (p. 413)
adaptor protein (p. 414)
Src homology 2 (SH2) domain (p. 414)
epidermal growth factor (EGF) (p. 415)
EGF receptor (EGFR) (p. 415)
dimerization arm (p. 416)
Src homology 3 (SH3) domain (p. 417)
Ras (p. 417)
small G protein (p. 417)
guanine-nucleotide-exchange factor (GEF) (p. 417)
extracellular signal-regulated kinase (ERK) (p. 418)
GTPase-activating protein (GAP) (p. 418)
oncogene (p. 420)
proto-oncogene (p. 420)
tumor-suppressor gene (p. 420)

Problems

1. *Active mutants.* Some protein kinases are inactive unless they are phosphorylated on key serine or threonine residues. In some cases, active enzymes can be generated by mutating these serine or threonine residues to glutamate. Explain.

2. *In the pocket.* SH2 domains bind phosphotyrosine residues in deep pockets on their surfaces. Would you expect SH2 domains to bind phosphoserine or phosphothreonine with high affinity? Why or why not?

3. *On–off.* Why is the GTPase activity of G proteins crucial to the proper functioning of a cell? Why have G proteins not evolved to catalyze GTP hydrolysis more efficiently?

4. *Viva la différence.* Why is the fact that a monomeric hormone binds to two identical receptor molecules, thus promoting the formation of a dimer of the receptor, considered remarkable?

5. *Antibodies mimicking hormones.* Antibodies have two identical antigen-binding sites. Remarkably, antibodies to the extracellular parts of growth-factor receptors often lead to the same cellular effects as does exposure to growth factors. Explain this observation.

6. *Facile exchange.* A mutated form of the α subunit of the heterotrimeric G protein has been identified; this form readily exchanges nucleotides even in the absence of an activated receptor. What would be the effect on a signaling pathway containing the mutated α subunit?

7. *Making connections.* Suppose that you were investigating a newly discovered growth-factor signal-transduction pathway. You found that, if you added GTPγS, a nonhydrolyzable analog of GTP, the duration of the hormonal response increased. What can you conclude?

8. *Diffusion rates.* Normally, rates of diffusion vary inversely with molecular weights; so smaller molecules diffuse faster than do larger ones. In cells, however, calcium ion diffuses more slowly than does cAMP. Propose a possible explanation.

9. *Negativity abounds.* Fura-2 is not effective for the study of calcium levels in intact, living cells. On the basis of how Fura-2 is depicted on p. 410, why is it ineffective?

10. *Awash with glucose.* Glucose is mobilized for ATP generation in muscle in response to epinephrine, which activates $G_{\alpha s}$. Cyclic AMP phosphodiesterase is an enzyme that converts cAMP into AMP. How would inhibitors of cAMP phosphodiesterase affect glucose mobilization in muscle?

11. *Getting it started.* The insulin receptor, on dimerization, cross-phosphorylates the activation loop of the other receptor molecule, leading to activation of the kinase. Propose how this phosphorylation event can take place if the kinase starts in an inactive conformation.

12. *Many defects.* Considerable effort has been directed toward determining the genes in which sequence variation contributes to the development of type 2 diabetes. Approximately 800 genes have been implicated. Propose an explanation for this observation.

13. *Growth-factor signaling.* Human growth hormone binds to a cell-surface membrane protein that is not a receptor tyrosine kinase. The intracellular domain of the receptor can bind other proteins inside the cell. Furthermore, studies indicate that the receptor is monomeric in the absence of hormone but dimerizes on hormone binding. Propose a possible mechanism for growth-hormone signaling.

14. *Receptor truncation.* You prepare a cell line that overexpresses a mutant form of EGFR in which the entire intracellular region of the receptor has been deleted. Predict the effect of overexpression of this construct on EGF signaling in this cell line.

15. *Hybrid.* Suppose that, through genetic manipulations, a chimeric receptor is produced that consists of the extracellular domain of the insulin receptor and the transmembrane and intracellular domains of the EGF receptor. Cells expressing this receptor are exposed to insulin, and the level of phosphorylation of the chimeric receptor is examined. What would you expect to observe and why? What would you expect to observe if these cells were exposed to EGF?

16. *Total amplification.* Suppose that each β-adrenergic receptor bound to epinephrine converts 100 molecules of $G_{\alpha s}$ into their GTP forms and that each molecule of activated adenylate cyclase produces 1000 molecules of cAMP per second. With the assumption of a full response, how many molecules of cAMP will be produced in 1 s after the formation of a single complex between epinephrine and the β-adrenergic receptor?

Chapter Integration Problems

17. *Nerve-growth-factor pathway.* Nerve-growth factor (NGF) binds to a protein tyrosine kinase receptor. The amount of diacylglycerol in the plasma membrane increases in cells expressing this receptor when treated with NGF. Propose a simple signaling pathway and identify the isoform of any participating enzymes. Would you expect the concentrations of any other common second messengers to increase on NGF treatment?

18. *Redundancy.* Because of the high degree of genetic variability in tumors, typically no single anticancer therapy is universally effective for all patients, even within a given tumor type. Hence, it is often desirable to inhibit a particular pathway at more than one point in the signaling cascade. In addition to the EGFR-directed monoclonal antibody cetuximab, propose alternative strategies for targeting the EGF signaling pathway for antitumor drug development.

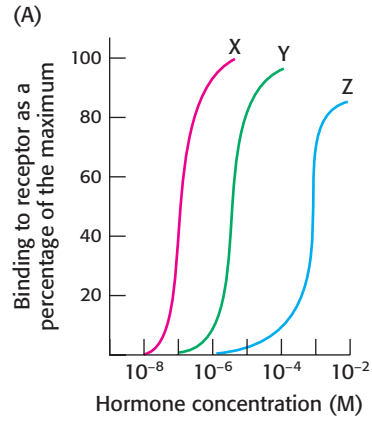
Mechanism Problems

19. *Distant relatives.* The structure of adenylate cyclase is similar to the structures of some types of DNA polymerases, suggesting that these enzymes derived from a common ancestor. Compare the reactions catalyzed by these two enzymes. In what ways are they similar?

20. *Kinase inhibitors as drugs.* Functional and structural analysis indicates that Gleevec is an ATP-competitive inhibitor of the Bcr-Abl kinase. In fact, many kinase inhibitors under investigation or currently marketed as drugs are ATP competitive. Can you suggest a potential drawback of drugs that utilize this particular mechanism of action?

Data Interpretation Problems

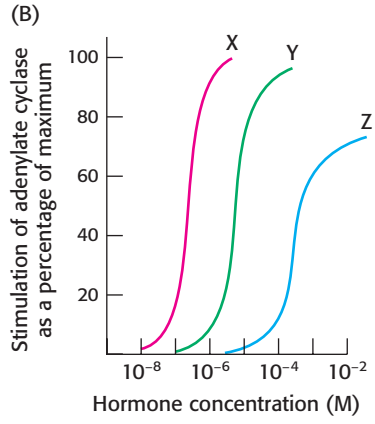
21. *Establishing specificity.* You wish to determine the hormone-binding specificity of a newly identified membrane receptor. Three different hormones, X, Y, and Z, were mixed with the receptor in separate experiments, and the percentage of binding capacity of the receptor was determined as a function of hormone concentration, as shown in graph A.



(a) What concentrations of each hormone yield 50% maximal binding?

(b) Which hormone shows the highest binding affinity for the receptor?

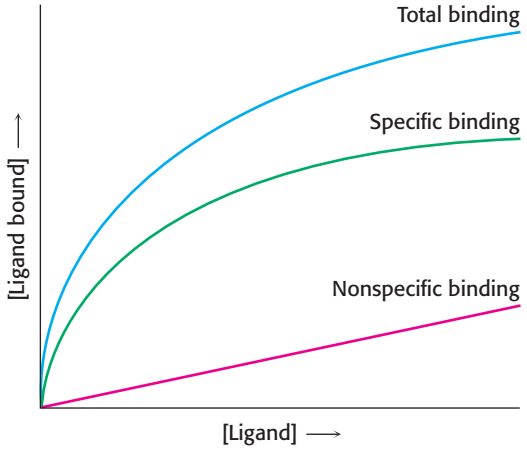
You next wish to determine whether the hormone–receptor complex stimulates the adenylate cyclase cascade. To do so, you measure adenylate cyclase activity as a function of hormone concentration, as shown in graph B.



(c) What is the relation between the binding affinity of the hormone–receptor complex and the ability of the hormone to enhance adenylate cyclase activity? What can you conclude about the mechanism of action of the hormone–receptor complex?

(d) Suggest experiments that would determine whether a $G_{\alpha s}$ protein is a component of the signal-transduction pathway.

22. *Binding issues.* A scientist wishes to determine the number of receptors specific for a ligand X, which he has in both radioactive and nonradioactive form. In one experiment, he adds increasing amounts of radioactive X and measures how much of it is bound to the cells. The result is shown as total activity in the following graph. Next, he performs the same experiment, except that he includes a several hundredfold excess of nonradioactive X. This result is shown as nonspecific binding. The difference between the two curves is the specific binding.



(a) Why is the total binding not an accurate representation of the number of receptors on the cell surface?

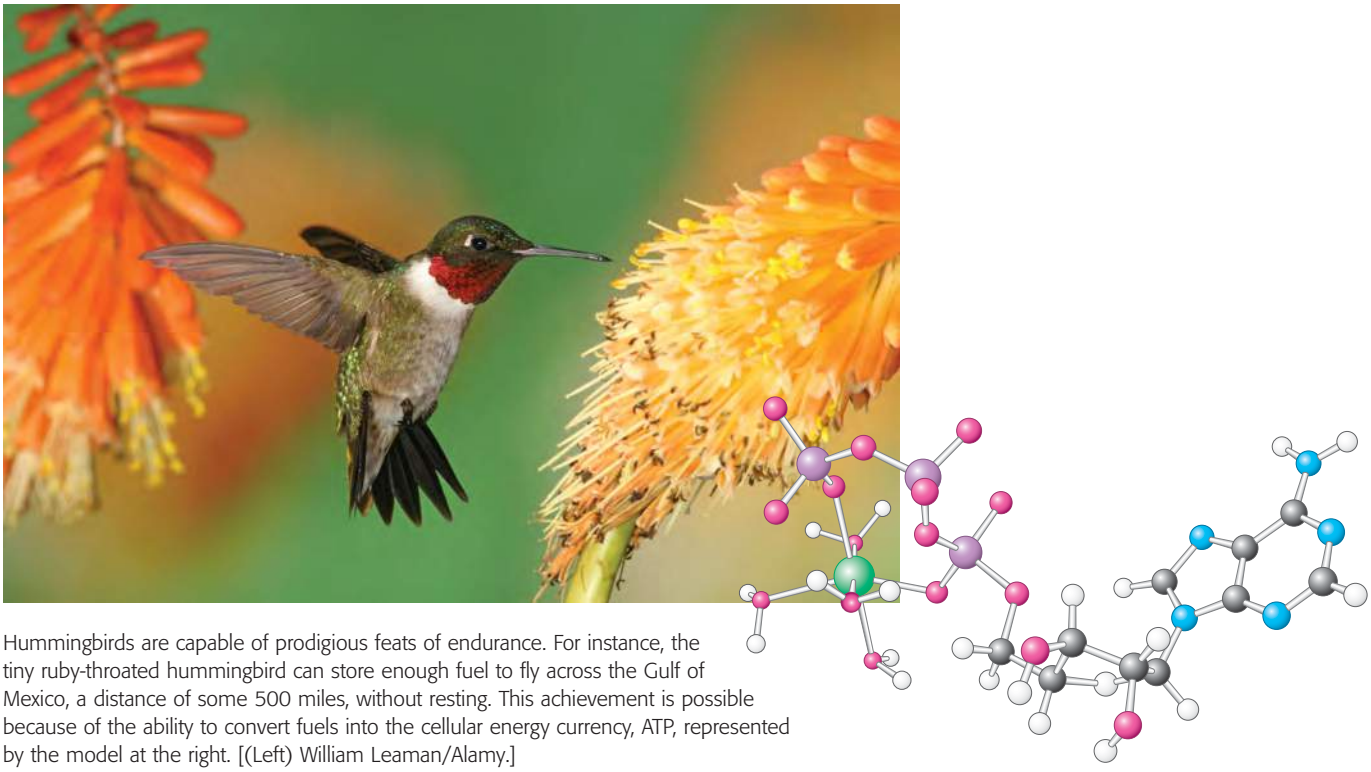
(b) What is the purpose of performing the experiment in the presence of excess nonradioactive ligand?

(c) What is the significance of the fact that specific binding attains a plateau?

23. *Counting receptors.* With the use of experiments such as those described in Problems 21 and 22, the number of receptors in the cell membrane can be calculated. Suppose that the specific activity of the ligand is 10^{12} cpm per millimole and that the maximal specific binding is 10^4 cpm per milligram of membrane protein. There are 10^{10} cells per milligram of membrane protein. Assume that one ligand binds per receptor. Calculate the number of receptor molecules present per cell.

CHAPTER 15

Metabolism: Basic Concepts and Design



Hummingbirds are capable of prodigious feats of endurance. For instance, the tiny ruby-throated hummingbird can store enough fuel to fly across the Gulf of Mexico, a distance of some 500 miles, without resting. This achievement is possible because of the ability to convert fuels into the cellular energy currency, ATP, represented by the model at the right. [(Left) William Leaman/Alamy.]

The concepts of conformation and dynamics developed in Part I—especially those dealing with the specificity and catalytic power of enzymes, the regulation of their catalytic activity, and the transport of molecules and ions across membranes—enable us to now ask questions fundamental to biochemistry:

1. *How does a cell extract energy and reducing power from its environment?*

2. *How does a cell synthesize the building blocks of its macromolecules and then the macromolecules themselves?*

These processes are carried out by a highly integrated network of chemical reactions that are collectively known as *metabolism* or *intermediary metabolism*.

More than a thousand chemical reactions take place in even as simple an organism as *Escherichia coli*. The array of reactions may seem overwhelming at first glance. However, closer scrutiny reveals that metabolism has a *coherent design containing many common motifs*. These motifs include the use of an energy currency and the repeated appearance of a limited number of activated intermediates. In fact, a group of about 100 molecules play central

OUTLINE