CHAPTER **THREE** 3

Energy, Catalysis, and Biosynthesis

One property above all makes living things seem almost miraculously different from nonliving matter: they create and maintain order in a universe that is tending always toward greater disorder. To accomplish this remarkable feat, the cells in a living organism must carry out a never-ending stream of chemical reactions that produce the molecules the organism requires to meet its metabolic needs. In some of these reactions, small organic molecules—amino acids, sugars, nucleotides, and lipids—are taken apart or modified to supply the many other small molecules that the cell requires. In other reactions, these small molecules are used to construct an enormously diverse range of larger molecules, including the proteins, nucleic acids, and other macromolecules that endow living systems with all of their most distinctive properties. Each cell can be viewed as a tiny chemical factory, performing many millions of these reactions every second.

To carry out the tremendous number of chemical reactions needed to sustain it, a living organism requires both a source of atoms in the form of food molecules and a source of energy. The atoms and the energy must both come, ultimately, from the nonliving environment. In this chapter, we discuss why cells require energy, and how they use energy and atoms from their environment to create the molecular order that makes life possible.

Most of the chemical reactions that cells perform would normally occur only at temperatures that are much higher than those inside a cell. Each reaction therefore requires a major boost in chemical reactivity to enable it to proceed rapidly within the cell. This boost is provided by specialized proteins called *enzymes*, each of which accelerates, or *catalyzes*, just one

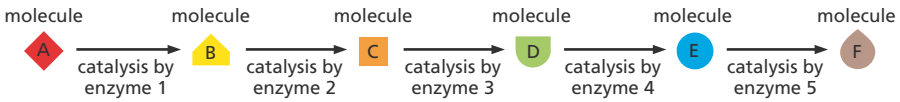


Figure 3–1 A series of enzyme-catalyzed reactions forms a metabolic pathway. Each enzyme catalyzes a chemical reaction involving a particular molecule. In this example, a set of enzymes acting in series converts molecule A to molecule F, forming a metabolic pathway.

of the many possible kinds of reactions that a particular molecule might undergo. These enzyme-catalyzed reactions are usually connected in series, so that the product of one reaction becomes the starting material for the next (**Figure 3–1**). The long linear reaction pathways, or *metabolic pathways*, that result are in turn linked to one another, forming a complex web of interconnected reactions.

Rather than being an inconvenience, the necessity for *catalysis* is a benefit, as it allows the cell to precisely control its **metabolism**—the sum total of all the chemical reactions it needs to carry out to survive, grow, and reproduce. This control is central to the chemistry of life.

Two opposing streams of chemical reactions occur in cells, the *catabolic* pathways and the *anabolic* pathways. The catabolic pathways (**catabolism**) break down foodstuffs into smaller molecules, thereby generating both a useful form of energy for the cell and some of the small molecules that the cell needs as building blocks. The anabolic, or *biosynthetic*, pathways (**anabolism**) use the energy harnessed by catabolism to drive the synthesis of the many molecules that form the cell. Together, these two sets of reactions constitute the metabolism of the cell (**Figure 3–2**).

The details regarding the individual reactions that comprise cell metabolism are part of the subject matter of *biochemistry*, and they need not concern us here. But the general principles by which cells obtain energy from their environment and use it to create order are central to cell biology. We begin this chapter with a discussion of why a constant input of energy is needed to sustain living organisms. We then discuss how enzymes catalyze the reactions that produce biological order. Finally, we describe the molecules that carry the energy that makes life possible.

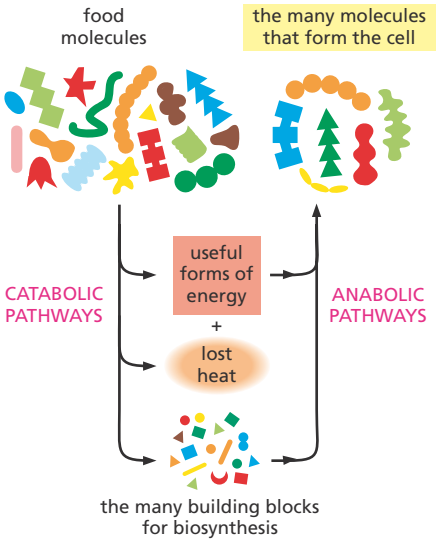


Figure 3–2 Catabolic and anabolic pathways together constitute the cell's metabolism. Note that a major portion of the energy stored in the chemical bonds of food molecules is dissipated as heat. Thus, only some of this energy can be converted to the useful forms of energy needed to drive the synthesis of new molecules.

THE USE OF ENERGY BY CELLS

Nonliving things left to themselves eventually become disordered: buildings crumble and dead organisms decay. Living cells, by contrast, not only maintain, but actually generate order at every level, from the large-scale structure of a butterfly or a flower down to the organization of the molecules that make up these organisms (**Figure 3–3**). This property of life is made possible by elaborate molecular mechanisms that extract energy from the environment and convert it into the energy stored in chemical bonds. Biological structures are therefore able to maintain their form, even though the materials of which they are made are continually being broken down, replaced, and recycled. Your body has the same basic structure it had 10 years ago, even though you now contain atoms that, for the most part, were not in your body then.

Biological Order Is Made Possible by the Release of Heat Energy from Cells

The universal tendency of things to become disordered is expressed in a fundamental law of physics, the *second law of thermodynamics.* This law states that, in the universe or in any isolated system (a collection of matter that is completely isolated from the rest of the universe), the degree of disorder can only increase. The second law of thermodynamics has such profound implications for living things that it is worth restating in several ways.

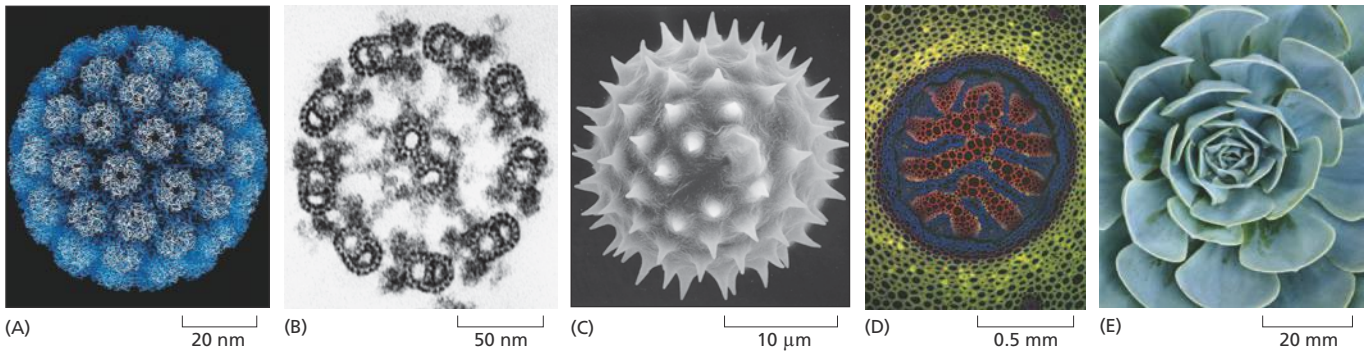


Figure 3–3 Biological structures are highly ordered. Well-defined, ornate, and beautiful spatial patterns can be found at every level of organization in living organisms. In order of increasing size: (A) protein molecules in the coat of a virus (a parasite that, although not technically alive, contains the same types of molecules as those found in living cells); (B) the regular array of microtubules seen in a cross section of a sperm tail; (C) surface contours of a pollen grain (a single cell); (D) cross section of a fern stem, showing the patterned arrangement of cells; and (E) flower with a spiral array of petals, each made of millions of cells. (A, courtesy of Robert Grant, Stéphane Crainic, and James M. Hogle; B, courtesy of Lewis Tilney; C, courtesy of Colin MacFarlane and Chris Jeffree; D, courtesy of Jim Haseloff.)

We can express the second law in terms of probability by stating that *systems will change spontaneously toward those arrangements that have the greatest probability.* Consider a box of 100 coins all lying heads up. A series of events that disturbs the box will tend to move the arrangement toward a mixture of 50 heads and 50 tails. The reason is simple: there are a huge number of possible arrangements of the individual coins that can achieve the 50–50 result, but only one possible arrangement that keeps them all oriented heads up. Because the 50–50 mixture accommodates a greater number of possibilities and places fewer constraints on the orientation of each individual coin, we say that it is more "disordered." For the same reason, one's living space will become increasingly disordered without an intentional effort to keep it organized. Movement toward disorder is a spontaneous process, requiring a periodic input of energy to reverse it (**Figure 3–4**).

The measure of a system's disorder is called the **entropy** of the system, and the greater the disorder, the greater the entropy. Thus another way to express the second law of thermodynamics is to say that systems will change spontaneously toward arrangements with greater entropy. Living cells—by surviving, growing, and forming complex communities and even whole organisms—generate order and thus might appear to defy the second law of thermodynamics. This is not the case, however,

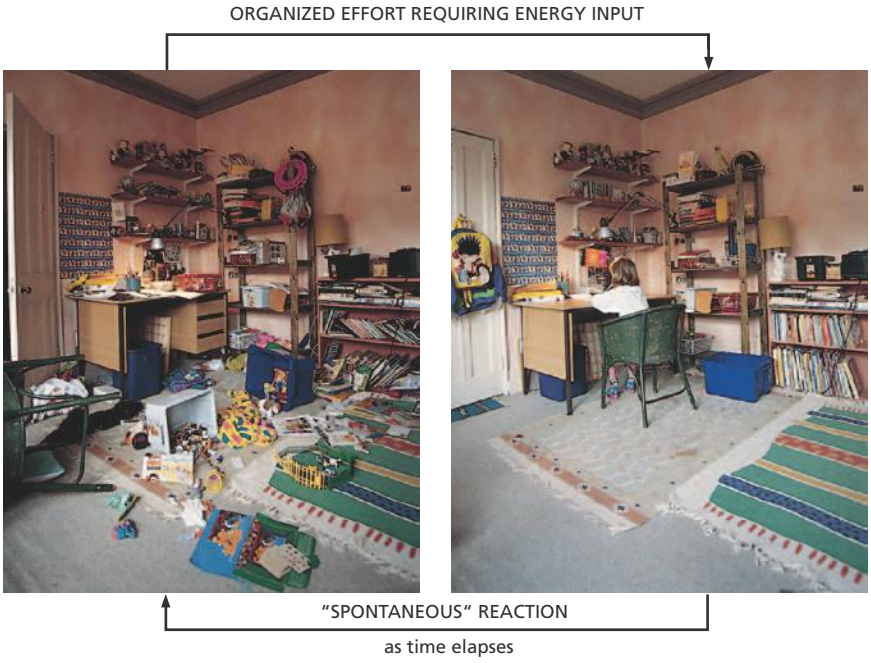


Figure 3–4 The spontaneous tendency toward disorder is an everyday experience. Reversing this natural tendency toward disorder requires an intentional effort and an input of energy. In fact, from the second law of thermodynamics, we can be certain that the human intervention required will release enough heat to the environment to more than compensate for the reestablishment of order in this room.

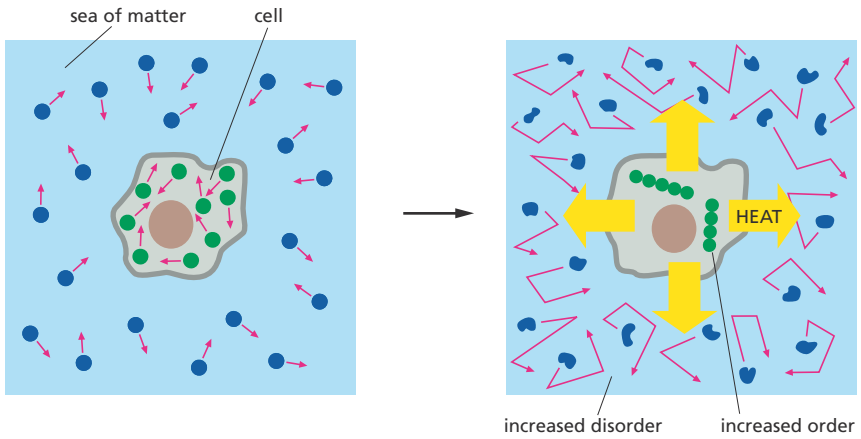


Figure 3–5 Living cells do not defy the second law of thermodynamics. In the diagram on the *left*, the molecules of both the cell and the rest of the universe (the environment) are depicted in a relatively disordered state. In the diagram on the *right*, the cell has taken in energy from food molecules and released heat by carrying out a reaction that orders the molecules that the cell contains. Because the heat increases the disorder in the environment around the cell—as depicted by the *longer, jagged red arrows*, which represent increased thermal motion, and the distorted molecules, which indicate enhanced molecular vibration and rotation—the second law of thermodynamics is satisfied, even as the cell grows and constructs larger molecules.

because a cell is not an isolated system. Rather, it takes in energy from its environment—in the form of food, inorganic molecules, or photons of light from the sun—and it then uses this energy to generate order within itself, forging new chemical bonds and building large macromolecules. In the course of performing the chemical reactions that generate order, some energy is lost in the form of heat. Heat is energy in its most disordered form—the random jostling of molecules (analogous to the random jostling of the coins in the box). Because the cell is not an isolated system, the heat energy that its reactions generate is quickly dispersed into the cell's surroundings. There, the heat increases the intensity of the thermal motions of nearby molecules, thereby increasing the entropy of the environment (**Figure 3–5**).

The amount of heat released by a cell must be great enough that the increased order generated inside the cell is more than compensated for by the increased disorder generated in the environment. Only in this case is the second law of thermodynamics satisfied, because the total entropy of the system—that of the cell plus its environment—increases as a result of the chemical reactions inside the cell.

Cells Can Convert Energy from One Form to Another

According to the *first law of thermodynamics*, energy cannot be created or destroyed—but it can be converted from one form to another (**Figure 3–6**). Cells take advantage of this law of thermodynamics, for example, when they convert the energy from sunlight into the energy in the chemical bonds of sugars and other small organic molecules during photosynthesis. Although chemical reactions that power such energy conversions can change how much energy is present in one form or another, the first law tells us that the total amount of energy in the universe must always be the same.

When an animal cell breaks down foodstuffs, some of the energy in the chemical bonds in the food molecules (chemical-bond energy) is converted into the thermal motion of molecules (heat energy). This conversion of chemical energy into heat energy causes the universe as a whole to become more disordered—as required by the second law of thermodynamics. But the cell cannot derive any benefit from the heat energy it produces unless the heat-generating reactions are directly linked to processes that maintain molecular order inside the cell. It is the tight coupling of heat production to an increase in order that distinguishes the metabolism of a cell from the wasteful burning of fuel in a fire. Later in this chapter, we illustrate how this coupling occurs. For the moment, it is

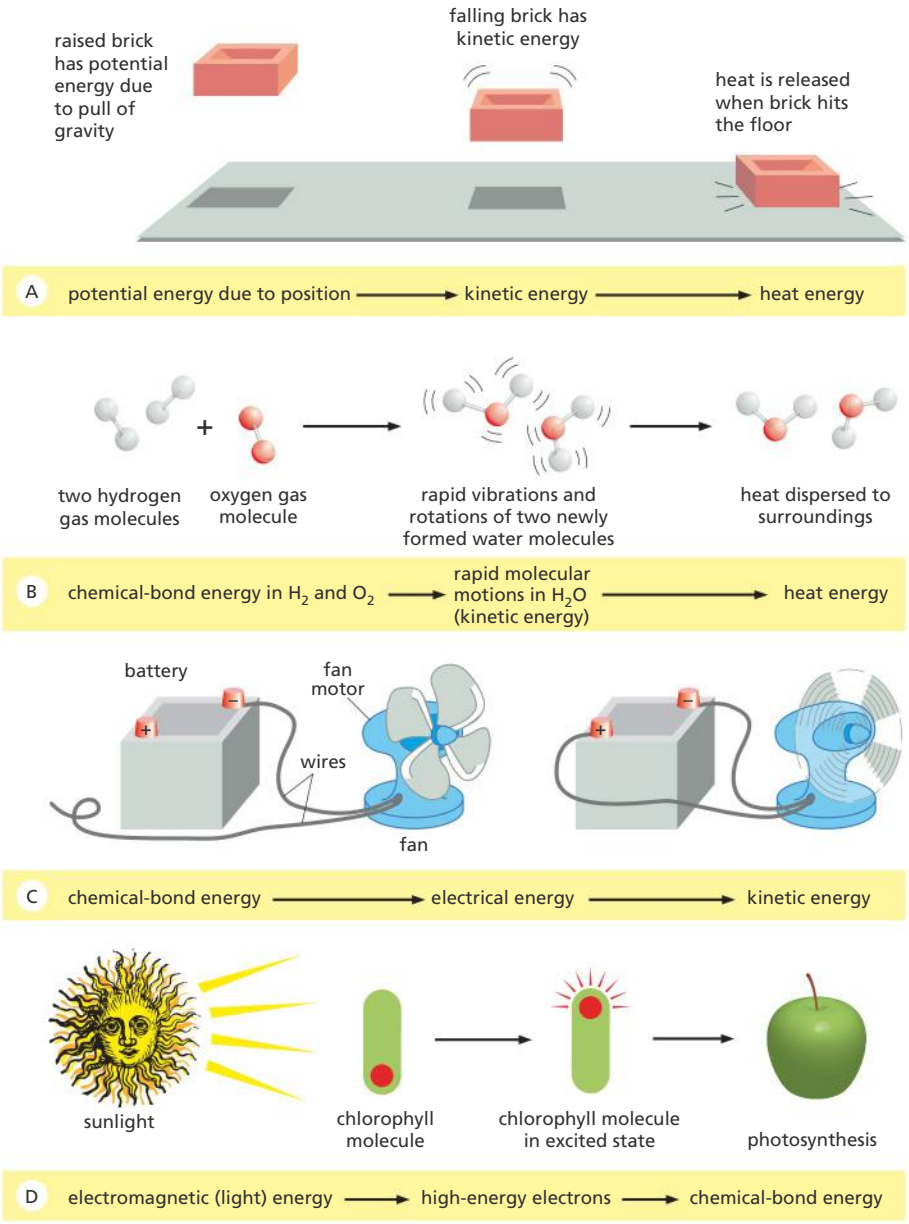


Figure 3–6 Different forms of energy are interconvertible, but the total amount of energy must be conserved. In (A), we can use the height and weight of the brick to predict exactly how much heat will be released when it hits the floor. In (B), the large amount of chemical-bond energy released when water (H_2O) is formed from H_2 and O_2 is initially converted to very rapid thermal motions in the two new H_2O molecules; however, collisions with other H_2O molecules almost instantaneously spread this kinetic energy evenly throughout the surroundings (heat transfer), making the new H_2O molecules indistinguishable from all the rest. (C) Cells can convert chemical-bond energy into kinetic energy to drive, for example, molecular motor proteins; however, this occurs without the intermediate conversion to electrical energy that a man-made appliance such as this fan requires. (D) Some cells can also harvest the energy from sunlight to form chemical bonds via photosynthesis.

sufficient to recognize that—by directly linking the "burning" of food molecules to the generation of biological order—cells are able to create and maintain an island of order in a universe tending toward chaos.

Photosynthetic Organisms Use Sunlight to Synthesize Organic Molecules

All animals live on energy stored in the chemical bonds of organic molecules, which they take in as food. These food molecules also provide the atoms that animals need to construct new living matter. Some animals obtain their food by eating other animals, others by eating plants. Plants, by contrast, obtain their energy directly from sunlight. Thus, the energy animals obtain by eating plants—or by eating animals that have eaten plants—ultimately comes from the sun (Figure 3–7).

Solar energy enters the living world through **photosynthesis**, a process that converts the electromagnetic energy in sunlight into chemical-bond energy in cells. Photosynthetic organisms—including plants, algae, and

Figure 3–7 With few exceptions, the radiant energy of sunlight sustains all life. Trapped by plants and some microorganisms through photosynthesis, light from the sun is the ultimate source of all energy for humans and other animals. (*Wheat Field Behind Saint-Paul Hospital with a Reaper* by Vincent van Gogh. Courtesy of Museum Folkwang, Essen.)

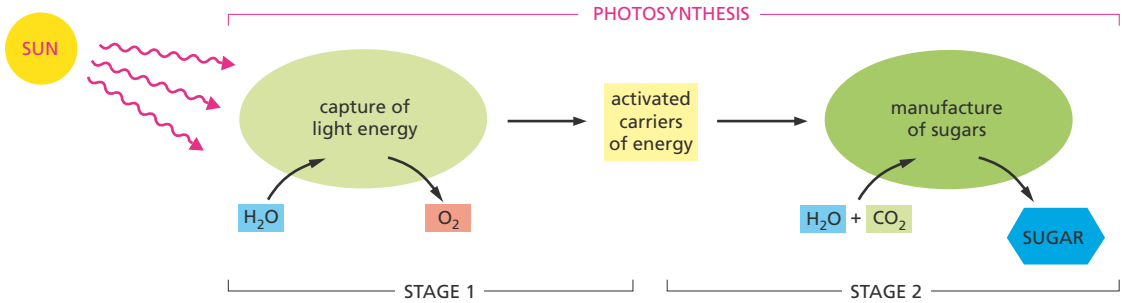


Figure 3–8 Photosynthesis takes place in two stages. The activated carriers generated in the first stage are two molecules that we will discuss shortly: ATP and NADPH.

some bacteria—use the energy they derive from sunlight to synthesize small chemical building blocks such as sugars, amino acids, nucleotides, and fatty acids. These small molecules in turn are converted into the macromolecules—the proteins, nucleic acids, polysaccharides, and lipids—that form the plant.

We describe the elegant mechanisms that underlie photosynthesis in detail in Chapter 14. Generally speaking, the reactions of photosynthesis take place in two stages. In the first stage, energy from sunlight is captured and transiently stored as chemical-bond energy in specialized molecules called *activated carriers*, which we discuss in more detail later in the chapter. All of the oxygen (O_2) in the air we breathe is generated by the splitting of water molecules during this first stage of photosynthesis.

In the second stage, the activated carriers are used to help drive a *carbon-fixation* process, in which sugars are manufactured from carbon dioxide gas (CO_2). In this way, photosynthesis generates an essential source of stored chemical-bond energy and other organic materials—for the plant itself and for any animals that eat it. The two stages of photosynthesis are summarized in **Figure 3–8**.

QUESTION 3–1

Consider the equation

light energy + CO_2 + H_2O → sugars + O_2 + heat energy

Would you expect this reaction to occur in a single step? Why must heat be generated in the reaction? Explain your answers.

Cells Obtain Energy by the Oxidation of Organic Molecules

All animal and plant cells require the chemical energy stored in the chemical bonds of organic molecules—either the sugars that a plant has produced by photosynthesis as food for itself or the mixture of large and small molecules that an animal has eaten. To use this energy to live, grow, and reproduce, organisms must extract it in a usable form. In both plants and animals, energy is extracted from food molecules by a process of gradual *oxidation*, or controlled burning.

Earth's atmosphere is about 21% oxygen. In the presence of oxygen, the most energetically stable form of carbon is CO_2 and that of hydrogen is H_2O. A cell is therefore able to obtain energy from sugars or other organic molecules by allowing the carbon and hydrogen atoms in these molecules to combine with oxygen—that is, become *oxidized*—to produce CO_2 and H_2O, respectively—a process known as cellular **respiration**.

Photosynthesis and cellular respiration are complementary processes (**Figure 3–9**). This means that the transactions between plants and animals are not all one way. Plants, animals, and microorganisms have existed together on this planet for so long that they have become an essential part of each other's environments. The oxygen released by photosynthesis is consumed by nearly all organisms for the oxidative breakdown of organic molecules. And some of the CO_2 molecules that today are incorporated into organic molecules by photosynthesis in a green leaf were released yesterday into the atmosphere by the respiration of an animal, a fungus, or the plant itself, or by the burning of

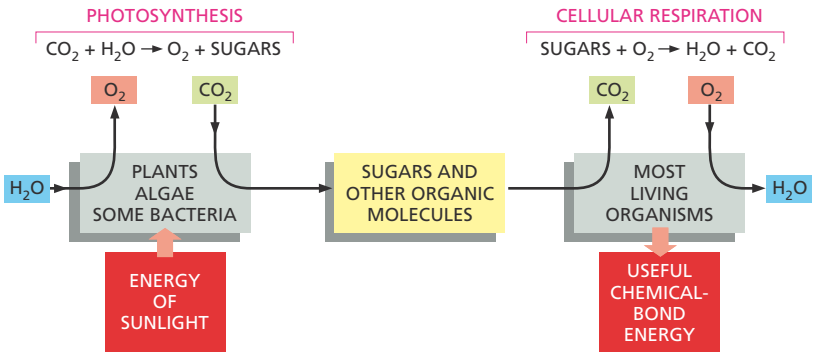


Figure 3–9 Photosynthesis and cellular respiration are complementary processes in the living world. The *left* side of the diagram shows how photosynthesis—carried out by plants and photosynthetic microorganisms—uses the energy of sunlight to produce sugars and other organic molecules from the carbon atoms in CO_2 in the atmosphere. In turn, these molecules serve as food for other organisms. The *right* side of the diagram shows how cellular respiration in most organisms—including plants and photosynthetic microorganisms—uses O_2 to oxidize food molecules, releasing the same carbon atoms in the form of CO_2 back to the atmosphere. In the process, the organisms obtain the useful chemical-bond energy that they need to survive. The first cells on Earth are thought to have been capable of neither photosynthesis nor cellular respiration (discussed in Chapter 14). However, photosynthesis must have preceded respiration on the Earth, because there is strong evidence that billions of years of photosynthesis were required to release enough O_2 to create an atmosphere that could support respiration.

fossil fuels. Carbon utilization therefore forms a huge cycle that involves the *biosphere* (all of the living organisms on Earth) as a whole, crossing boundaries between individual organisms (**Figure 3–10**).

Oxidation and Reduction Involve Electron Transfers

The cell does not oxidize organic molecules in one step, as occurs when organic material is burned in a fire. Through the use of enzyme catalysts, metabolism directs the molecules through a large number of reactions, few of which actually involve the direct addition of oxygen. Thus, before we consider some of these reactions, we should explain what is meant by oxidation.

The term **oxidation** literally means the addition of oxygen atoms to a molecule. More generally, though, oxidation is said to occur in any reaction in which electrons are transferred from one atom to another. Oxidation, in this sense, refers to the removal of electrons from an atom. The converse reaction, called **reduction**, involves the addition of electrons to an atom. Thus, Fe^{2+} is oxidized when it loses an electron to become Fe^{3+}, whereas a chlorine atom is reduced when it gains an electron to become Cl^-. Because the number of electrons is conserved in a chemical reaction (there is no net loss or gain), oxidation and reduction always occur simultaneously: that is, if one molecule gains an electron in a reaction (reduction), a second molecule must lose the electron (oxidation). When a sugar molecule is oxidized to CO_2 and H_2O, for example, the O_2 molecules involved in forming H_2O gain electrons and thus are said to have been reduced.

The terms oxidation and reduction apply even when there is only a partial shift of electrons between atoms linked by a covalent bond. When a carbon atom becomes covalently bonded to an atom with a strong affinity for electrons—oxygen, chlorine, or sulfur, for example—it gives up more

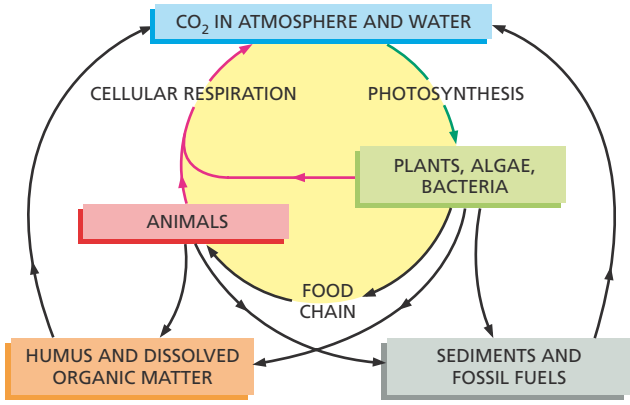


Figure 3–10 Carbon atoms cycle continuously through the biosphere. Individual carbon atoms are incorporated into organic molecules of the living world by the photosynthetic activity of plants, algae, and bacteria. They then pass to animals and microorganisms—as well as into organic material in soil and oceans—and are ultimately restored to the atmosphere in the form of CO_2 when organic molecules are oxidized by cells during respiration or burned by humans as fossil fuels.

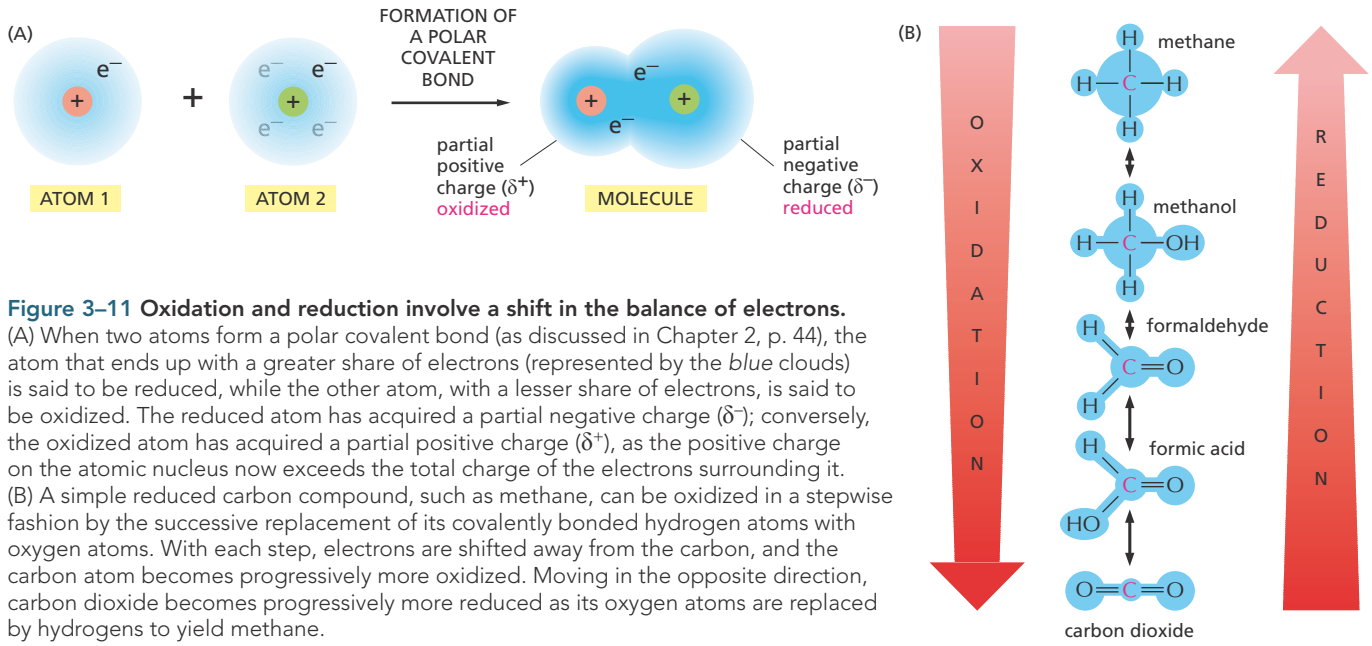


Figure 3–11 Oxidation and reduction involve a shift in the balance of electrons. (A) When two atoms form a polar covalent bond (as discussed in Chapter 2, p. 44), the atom that ends up with a greater share of electrons (represented by the *blue* clouds) is said to be reduced, while the other atom, with a lesser share of electrons, is said to be oxidized. The reduced atom has acquired a partial negative charge (δ^-); conversely, the oxidized atom has acquired a partial positive charge (δ^+), as the positive charge on the atomic nucleus now exceeds the total charge of the electrons surrounding it. (B) A simple reduced carbon compound, such as methane, can be oxidized in a stepwise fashion by the successive replacement of its covalently bonded hydrogen atoms with oxygen atoms. With each step, electrons are shifted away from the carbon, and the carbon atom becomes progressively more oxidized. Moving in the opposite direction, carbon dioxide becomes progressively more reduced as its oxygen atoms are replaced by hydrogens to yield methane.

than its equal share of electrons and forms a *polar covalent bond*. The positive charge of the carbon nucleus now slightly exceeds the negative charge of its electrons, so that the carbon atom acquires a partial positive charge (δ^+) and is said to be oxidized. Conversely, the carbon atom in a C–H bond has somewhat more than its share of electrons; it acquires a partial negative charge (δ^-), and so is said to be reduced (**Figure 3–11A**).

When a molecule in a cell picks up an electron (e^-), it often picks up a proton (H^+) at the same time (protons being freely available in water). The net effect in this case is to add a hydrogen atom to the molecule:

$$A + e^- + H^+ \rightarrow AH$$

Even though a proton plus an electron is involved (instead of just an electron), such *hydrogenation* reactions are reductions, and the reverse, *dehydrogenation*, reactions are oxidations. An easy way to tell whether an organic molecule is being oxidized or reduced is to count its C–H bonds: reduction occurs when the number of C–H bonds increases, whereas oxidation occurs when the number of C–H bonds decreases (**Figure 3–11B**).

As we will see later in this chapter—and again in Chapter 13—cells use enzymes to catalyze the oxidation of organic molecules in small steps, through a sequence of reactions that allows energy to be harvested in useful forms.

FREE ENERGY AND CATALYSIS

Enzymes, like cells, obey the second law of thermodynamics. Although they can speed up energetically favorable reactions—those that produce disorder in the universe—enzymes cannot by themselves force energetically unfavorable reactions to occur. Cells, however, must do just that in order to grow and divide—or just to survive. They must build highly ordered and energy-rich molecules from small and simple ones—a process that requires an input of energy.

To understand how enzymes promote **catalysis**—the acceleration of the specific chemical reactions needed to sustain life—we first need to examine the energetics involved. In this section, we consider how the free energy of molecules contributes to their chemistry, and we see how

free-energy changes—which reflect how much total disorder is generated in the universe by a reaction—influence whether and how the reaction will proceed. We then discuss how enzymes lower the activation energy needed to initiate reactions in the cell. And we describe how enzymes can exploit differences in the free-energy changes of different reactions to drive the energetically unfavorable reactions that produce biological order. Such enzyme-assisted catalysis is crucial for cells: without it, life could not exist.

Chemical Reactions Proceed in the Direction that Causes a Loss of Free Energy

Paper burns readily, releasing into the atmosphere water and carbon dioxide as gases, while simultaneously releasing energy as heat:

$$\text{paper} + O_2 \rightarrow \text{smoke} + \text{ashes} + \text{heat} + CO_2 + H_2O$$

This reaction occurs in only one direction: smoke and ashes never spontaneously gather carbon dioxide and water from the heated atmosphere and reconstitute themselves into paper. When paper burns, much of its chemical energy is dissipated as heat: it is not lost from the universe, since energy can never be created or destroyed; instead, it is irretrievably dispersed in the chaotic random thermal motions of molecules. At the same time, the atoms and molecules of the paper become dispersed and disordered. In the language of thermodynamics, there has been a release of *free energy*—that is, energy that can be harnessed to do work or drive chemical reactions. This release reflects a loss of orderliness in the way the energy and molecules had been stored in the paper. We will discuss free energy in more detail shortly, but the general principle can be summarized as follows: chemical reactions proceed only in the direction that leads to a loss of free energy. In other words, the spontaneous direction for any reaction is the direction that goes "downhill." A "downhill" reaction in this sense is said to be energetically favorable.

Enzymes Reduce the Energy Needed to Initiate Spontaneous Reactions

Although the most energetically favorable form of carbon under ordinary conditions is CO_2, and that of hydrogen is H_2O, a living organism will not disappear in a puff of smoke, and the book in your hands will not burst spontaneously into flames. This is because the molecules in both the living organism and the book are in a relatively stable state, and they cannot be changed to lower-energy states without an initial input of energy. In other words, a molecule requires a boost over an energy barrier before it can undergo a chemical reaction that moves it to a lower-energy (more stable) state (**Figure 3–12A**). This boost is known as the

QUESTION 3–2

In which of the following reactions does the *red* atom undergo an oxidation?

A. $Na \rightarrow Na^+$ (Na atom → Na^+ ion)
B. $Cl \rightarrow Cl^-$ (Cl atom → Cl^- ion)
C. $CH_3CH_2OH \rightarrow CH_3CHO$ (ethanol → acetaldehyde)
D. $CH_3CHO \rightarrow CH_3COO^-$ (acetaldehyde → acetic acid)
E. $CH_2{=}CH_2 \rightarrow CH_3CH_3$ (ethene → ethane)

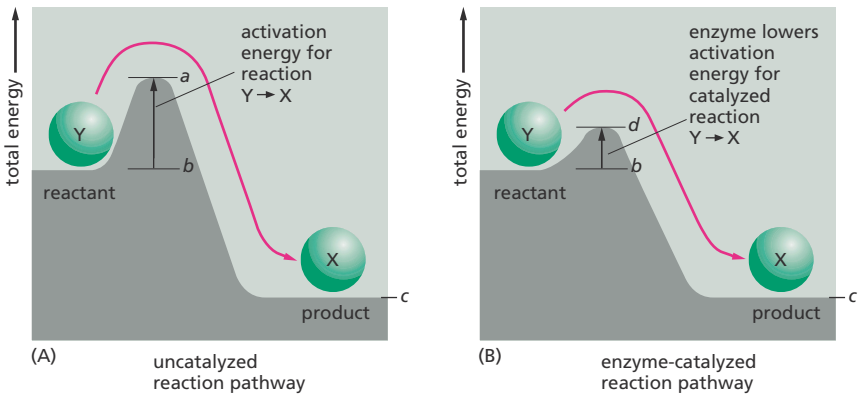


Figure 3–12 Even energetically favorable reactions require activation energy to get them started. (A) Compound Y (a reactant) is in a relatively stable state; thus energy is required to convert it to compound X (a product), even though X is at a lower overall energy level than Y. This conversion will not take place, therefore, unless compound Y can acquire enough *activation energy* (*energy a* minus *energy b*) from its surroundings to undergo the reaction that converts it into compound X. This energy may be provided by means of an unusually energetic collision with other molecules. For the reverse reaction, X → Y, the activation energy required will be much larger (*energy a* minus *energy c*); this reaction will therefore occur much more rarely. Activation energies are always positive. The total energy change for the energetically favorable reaction Y → X, is *energy c* minus *energy b*, a negative number, which corresponds to a loss of free energy. (B) Energy barriers for specific reactions can be lowered by catalysts, as indicated by the line marked *d*. Enzymes are particularly effective catalysts because they greatly reduce the activation energy for the reactions they catalyze.

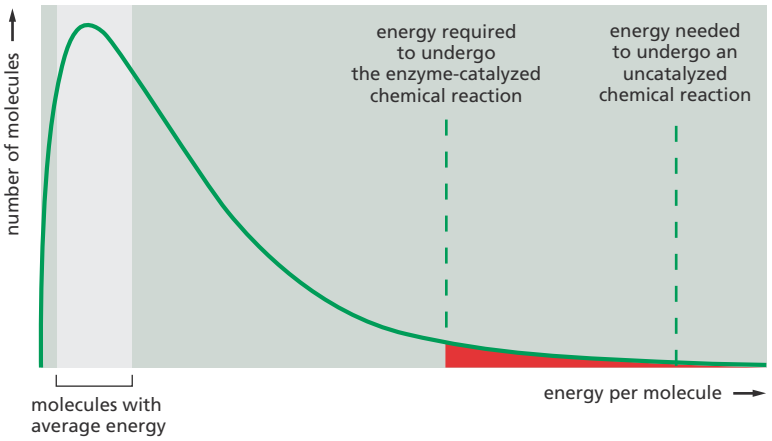


Figure 3–13 Lowering the activation energy greatly increases the probability that a reaction will occur. At any given instant, a population of identical substrate molecules will have a range of energies, distributed as shown on the graph. The varying energies come from collisions with surrounding molecules, which make the substrate molecules jiggle, vibrate, and spin. For a molecule to undergo a chemical reaction, the energy of the molecule must exceed the activation energy barrier for that reaction (*dashed lines*); for most biological reactions, this almost never happens without enzyme catalysis. Even with enzyme catalysis, only a small fraction of substrate molecules reach an energy state that is high enough for them to undergo a reaction (*red shaded* area).

activation energy. In the case of a burning book, the activation energy is provided by the heat of a lighted match. But cells can't raise their temperature to drive biological reactions. Inside cells, the push over the energy barrier is aided by specialized proteins called **enzymes**.

Each enzyme binds tightly to one or two molecules, called **substrates**, and holds them in a way that greatly reduces the activation energy needed to facilitate a specific chemical interaction between them (**Figure 3–12B**). A substance that can lower the activation energy of a reaction is termed a **catalyst**; catalysts increase the rate of chemical reactions because they allow a much larger proportion of the random collisions with surrounding molecules to kick the substrates over the energy barrier, as illustrated in **Figure 3–13** and **Figure 3–14A**. Enzymes are among the most effective catalysts known. They can speed up reactions by a factor of as much as 10^{14} (that is, trillions of times faster than the same reactions would proceed without an enzyme catalyst). Enzymes therefore allow reactions that would not otherwise occur to proceed rapidly at the normal temperature inside cells.

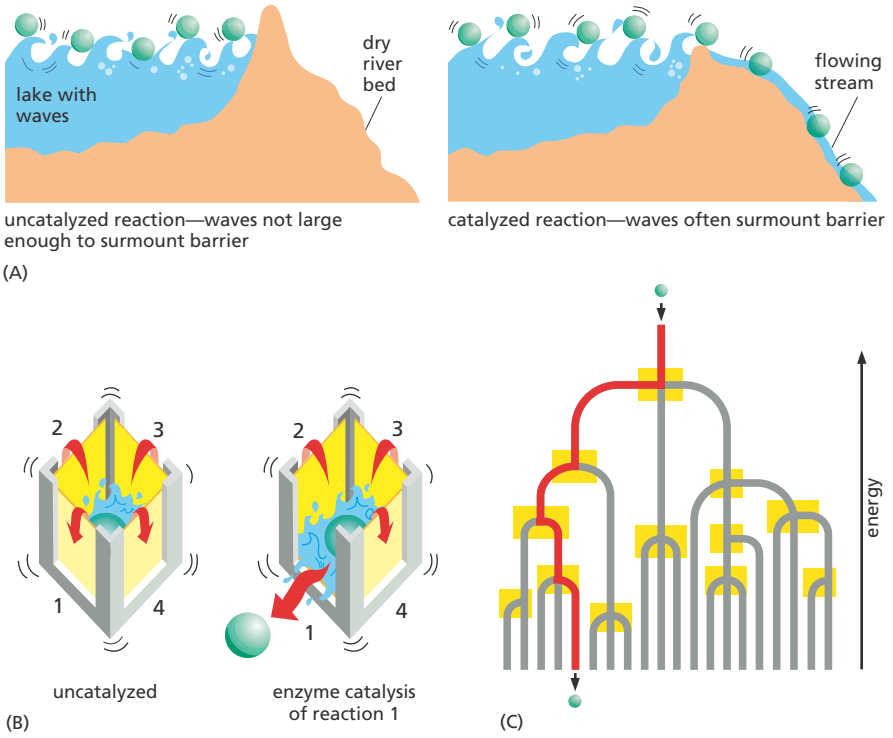


Figure 3–14 Enzymes catalyze reactions by lowering the activation energy barrier. (A) The dam represents the activation energy, which is lowered by enzyme catalysis. Each *green ball* represents a potential substrate molecule that is bouncing up and down in energy level owing to constant encounters with waves, an analogy for the thermal bombardment of substrate molecules by surrounding water molecules. When the barrier—the activation energy—is lowered significantly, the balls (substrate molecules) with sufficient energy can roll downhill, an energetically favorable movement. (B) The four walls of the box represent the activation energy barriers for four different chemical reactions that are all energetically favorable because the products are at lower energy levels than the substrates. In the left-hand box, none of these reactions occurs because even the largest waves are not large enough to surmount any of the energy barriers. In the right-hand box, enzyme catalysis lowers the activation energy for reaction number 1 only; now the jostling of the waves allows the substrate molecule to pass over this energy barrier, allowing reaction 1 to proceed (**Movie 3.1**). (C) A branching river with a set of barrier dams (*yellow* boxes) serves to illustrate how a series of enzyme-catalyzed reactions determines the exact reaction pathway followed by each molecule inside the cell by controlling specifically which reaction will be allowed at each junction.

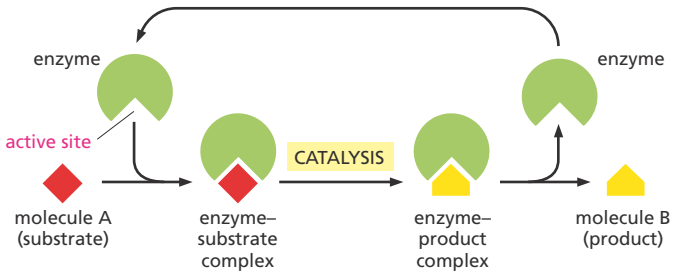


Figure 3–15 Enzymes convert substrates to products while remaining unchanged themselves. Each enzyme has an active site to which one or two substrate molecules bind, forming an enzyme–substrate complex. A reaction occurs at the active site, generating an enzyme–product complex. The product is then released, allowing the enzyme to bind additional substrate molecules and repeat the reaction. An enzyme thus serves as a catalyst, and it usually forms or breaks a single covalent bond in a substrate molecule.

Unlike the effects of temperature, enzymes are highly selective. Each enzyme usually speeds up only one particular reaction out of the several possible reactions that its substrate molecules could undergo. In this way, enzymes direct each of the many different molecules in a cell along specific reaction pathways (**Figure 3–14B and C**), thereby producing the compounds that the cell actually needs.

Like all catalysts, enzyme molecules themselves remain unchanged after participating in a reaction and therefore can function over and over again (**Figure 3–15**). In Chapter 4, we will discuss further how enzymes work, after we have looked in detail at the molecular structure of proteins.

The Free-Energy Change for a Reaction Determines Whether It Can Occur

According to the second law of thermodynamics, a chemical reaction can proceed only if it results in a net (overall) increase in the disorder of the universe (see Figure 3–5). Disorder increases when useful energy that could be harnessed to do work is dissipated as heat. The useful energy in a system is known as its **free energy**, or ***G***. And because chemical reactions involve a transition from one molecular state to another, the term that is of most interest to chemists and cell biologists is the **free-energy change**, denoted **ΔG** ("Delta *G*").

Let's consider a collection of molecules. ΔG measures the amount of disorder created in the universe when a reaction involving these molecules takes place. *Energetically favorable* reactions, by definition, are those that create disorder by decreasing the free energy of the system to which they belong; in other words, they have a *negative* ΔG (**Figure 3–16**).

A reaction can occur spontaneously only if ΔG is negative. On a macroscopic scale, an energetically favorable reaction with a negative ΔG is the relaxation of a compressed spring into an expanded state, releasing its stored elastic energy as heat to its surroundings. On a microscopic scale, an energetically favorable reaction with a negative ΔG occurs when salt (NaCl) dissolves in water. Note that, just because a reaction can occur spontaneously, does not mean it will occur quickly. The decay of diamonds into graphite is a spontaneous process—but it takes millions of years.

Energetically unfavorable reactions, by contrast, create order in the universe; they have a *positive* ΔG. Such reactions—for example, the formation of a peptide bond between two amino acids—cannot occur spontaneously; they take place only when they are coupled to a second reaction with a negative ΔG large enough that the net ΔG of the entire process is negative (**Figure 3–17**). Life is possible because enzymes can create biological order by coupling energetically unfavorable reactions with energetically favorable ones. These critical concepts are summarized, with examples, in **Panel 3–1** (pp. 96–97).

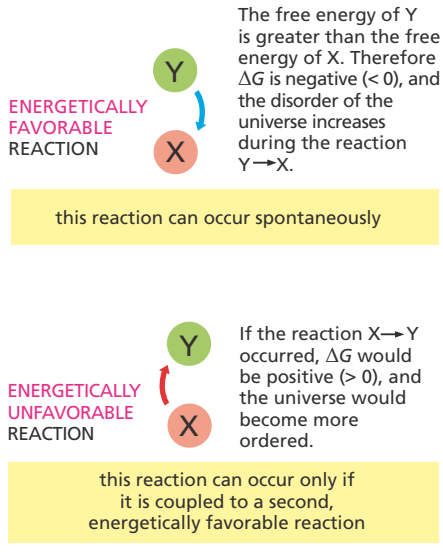


Figure 3–16 Energetically favorable reactions have a negative ΔG, whereas energetically unfavorable reactions have a positive ΔG.

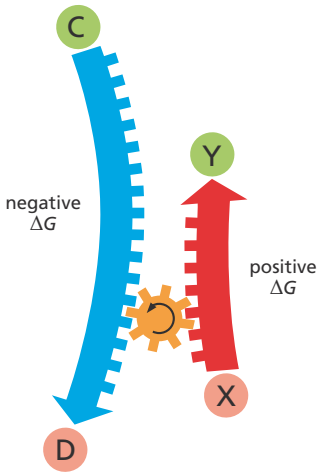


Figure 3–17 Reaction coupling can drive an energetically unfavorable reaction. The energetically unfavorable ($\Delta G > 0$) reaction X → Y cannot occur unless it is coupled to an energetically favorable ($\Delta G < 0$) reaction C → D, such that the net free-energy change for the coupled reactions is negative (less than 0).

ΔG Changes As a Reaction Proceeds Toward Equilibrium

It's easy to see how a tensed spring, when left to itself, will relax and release its stored energy to the environment as heat. But chemical reactions are a bit more complex—and harder to intuit. That's because whether a reaction will proceed depends not only on the energy stored in each individual molecule, but also on the concentrations of the molecules in the reaction mixture. Recalling our coin analogy, more coins in a jiggling box will flip from a head to a tail orientation when the box contains 90 heads and 10 tails, than when the box contains 10 heads and 90 tails.

The same is true for a chemical reaction. As the energetically favorable reaction Y → X proceeds, the concentration of the product X will increase and the concentration of the substrate Y will decrease. This change in relative concentrations of substrate and product will cause the ratio of Y to X to shrink, making the initially favorable ΔG less and less negative. Unless more Y is added, the reaction will slow and eventually stop.

Because ΔG changes as products accumulate and substrates are depleted, chemical reactions will generally proceed until they reach a state of **equilibrium**. At that point, the rates of the forward and reverse reactions are equal, and there is no further net change in the concentrations of substrate or product (**Figure 3–18**). For reactions at chemical equilibrium, $\Delta G = 0$, so the reaction will not proceed forward or backward, and no work can be done.

Such a state of chemical inactivity would be incompatible with life. Living cells avoid reaching a state of complete chemical equilibrium because they are constantly exchanging materials with their environment: replenishing nutrients and eliminating waste products. Many of the individual reactions in the cell's complex metabolic network also exist in disequilibrium because the products of one reaction are continually being siphoned off to become the substrates in a subsequent reaction. Rarely do products and substrates reach concentrations at which the forward and reverse reaction rates are equal.

The Standard Free-Energy Change, ΔG°, Makes it Possible to Compare the Energetics of Different Reactions

Because ΔG depends on the concentrations of the molecules in the reaction mixture at any given time, it is not a particularly useful value for comparing the relative energies of different types of reactions. But such energetic assessments are necessary, for example, to predict whether an energetically favorable reaction is likely to have a ΔG negative enough to drive an energetically unfavorable reaction. To compare reactions in this way, we need to turn to the *standard free-energy change* of a reaction, $\Delta G°$. The $\Delta G°$ is independent of concentration; it depends only on the intrinsic characters of the reacting molecules, based on their behavior under ideal conditions where the concentrations of all the reactants are set to the same fixed value of 1 mole/liter.

A large body of thermodynamic data has been collected from which $\Delta G°$ can be calculated for most metabolic reactions. Some common reactions are compared in terms of their $\Delta G°$ in Panel 3–1 (pp. 96–97).

The ΔG of a reaction can be calculated from $\Delta G°$ if the concentrations of the reactants and products are known. For the simple reaction Y → X, their relationship follows this equation:

$$\Delta G = \Delta G° + RT \ln \frac{[X]}{[Y]}$$

where ΔG is in kilocalories per mole, [Y] and [X] denote the concentrations

QUESTION 3–3

Consider the analogy of the jiggling box containing coins that was described on page 85. The reaction, the flipping of coins that either face heads up (H) or tails up (T), is described by the equation H ↔ T, where the rate of the forward reaction equals the rate of the reverse reaction.

A. What are ΔG and $\Delta G°$ in this analogy?

B. What corresponds to the temperature at which the reaction proceeds? What corresponds to the activation energy of the reaction? Assume you have an "enzyme," called jigglase, which catalyzes this reaction. What would the effect of jigglase be and what, mechanically, might jigglase do in this analogy?

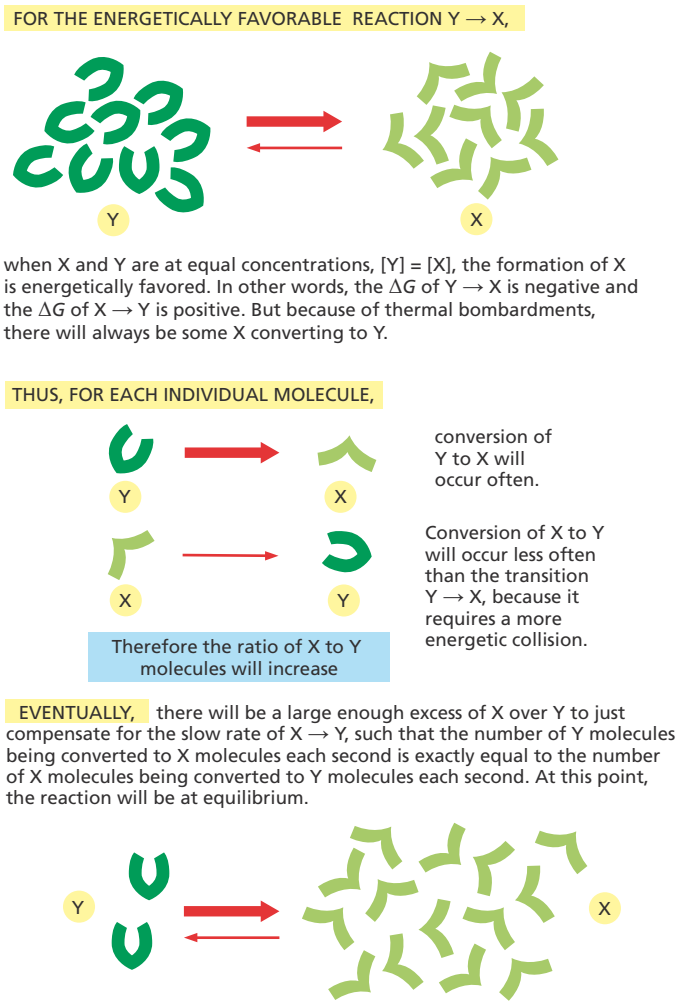


Figure 3–18 Reactions will eventually reach a chemical equilibrium. At that point, the forward and the backward fluxes of reacting molecules are equal and opposite. The widths of the arrows indicate the relative rates at which an *individual molecule* converts.

of Y and X in moles/liter, ln is the natural logarithm, and RT is the product of the gas constant, R, and the absolute temperature, T. At 37°C, $RT = 0.616$. (A mole is 6×10^{23} molecules of a substance.)

From this equation, we can see that when the concentrations of reactants and products are equal, in other words, [X]/[Y] = 1, the value of ΔG equals the value of $\Delta G°$ (because ln 1 = 0). Thus when the reactants and products are present in equal concentrations, the direction of the reaction depends entirely on the intrinsic properties of the molecules.

The Equilibrium Constant Is Directly Proportional to $\Delta G°$

As mentioned earlier, all chemical reactions tend to proceed toward equilibrium. Knowing where that equilibrium lies for any given reaction will tell you which way the reaction will proceed—and how far it will go. For example, if a reaction is at equilibrium when the concentration of the product is ten times the concentration of the substrate, and we begin with a surplus of substrate and little or no product, the reaction will proceed forward for some time. For the simple reaction Y → X, that value—the ratio of substrate to product at equilibrium—is called the reaction's **equilibrium constant**, ***K***. Expressed as an equation:

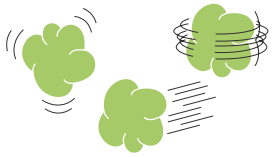
$$K = \frac{[X]}{[Y]}$$

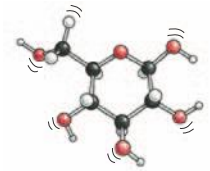
where [X] is the concentration of the product and [Y] is the concentration of the substrate at equilibrium.

FREE ENERGY

This panel reviews the concept of free energy and offers examples showing how changes in free energy determine whether—and how—biological reactions occur.

The molecules of a living cell possess energy because of their vibrations, rotations, and movement through space, and because of the energy that is stored in the bonds between individual atoms.



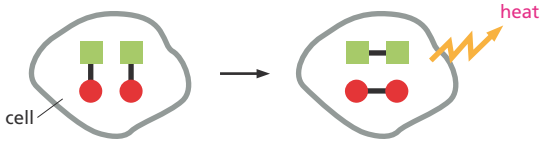


The free energy, *G* (in kcal/mole), measures the energy of a molecule which could in principle be used to do useful work at constant temperature, as in a living cell. Energy can also be expressed in joules (1 cal = 4.184 joules).

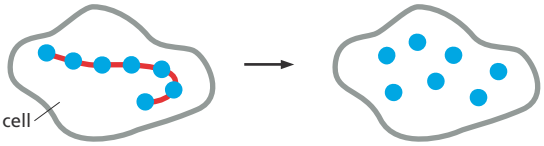
REACTIONS CAUSE DISORDER

Think of a chemical reaction occurring in a cell that has a constant temperature and volume. This reaction can produce disorder in two ways.

1 Changes of bond energy of the reacting molecules can cause heat to be released, which disorders the environment around the cell.



2 The reaction can decrease the amount of order in the cell—for example, by breaking apart a long chain of molecules, or by disrupting an interaction that prevents bond rotations.



PREDICTING REACTIONS

To predict the outcome of a reaction (Will it proceed to the right or to the left? At what point will it stop?), we must measure its standard free-energy change ($\Delta G°$). This quantity represents the gain or loss of free energy as one mole of reactant is converted to one mole of product under "standard conditions" (all molecules present at a concentration of 1 M and pH 7.0).

driving force

$\Delta G°$ for some reactions

glucose-1-P → glucose-6-P	–1.7 kcal/mole
sucrose → glucose + fructose	–5.5 kcal/mole
ATP → ADP + P_i	–7.3 kcal/mole
glucose + $6O_2$ → $6CO_2$ + $6H_2O$	–686 kcal/mole

ΔG ("DELTA G")

Changes in free energy occurring in a reaction are denoted by ΔG, where "Δ" indicates a difference. Thus, for the reaction

$$A + B \longrightarrow C + D$$

ΔG = free energy (C + D) minus free energy (A + B)

ΔG measures the amount of disorder caused by a reaction: the change in order inside the cell, plus the change in order of the surroundings caused by the heat released.

ΔG is useful because it measures how far away from equilibrium a reaction is. Thus the reaction

$$ATP \longrightarrow ADP + P_i$$

has a large negative ΔG because cells keep the reaction a long way from equilibrium by continually making fresh ATP. However, if the cell dies, then most of its ATP will be hydrolyzed, until equilibrium is reached; at equilibrium, the forward and backward reactions occur at equal rates and $\Delta G = 0$.

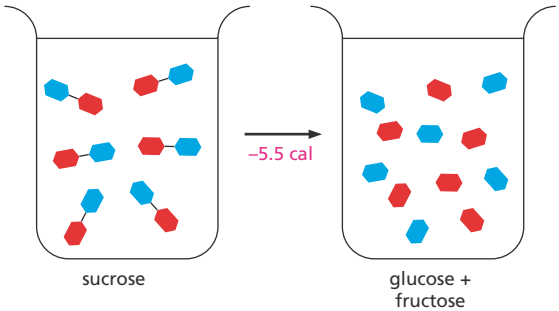
SPONTANEOUS REACTIONS

From the second law of thermodynamics, we know that the disorder of the universe can only increase. ΔG is *negative* if the disorder of the universe (reaction plus surroundings) *increases*.

In other words, a chemical reaction that occurs spontaneously must have a negative ΔG:

$$G_{products} - G_{reactants} = \Delta G < 0$$

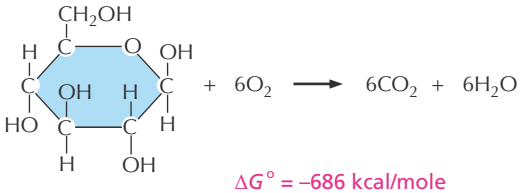
EXAMPLE: The difference in free energy of 100 ml of 10 mM sucrose (common sugar) and 100 ml of 10 mM glucose plus 10 mM fructose is about –5.5 calories. Therefore, the hydrolysis reaction that produces two monosaccharides from a disaccharide (sucrose → glucose + fructose) can proceed spontaneously.



In contrast, the reverse reaction (glucose + fructose → sucrose), which has a ΔG of +5.5 calories, could not occur without an input of energy from a coupled reaction.

REACTION RATES

A spontaneous reaction is not necessarily an instantaneous reaction: a reaction with a negative free-energy change (ΔG) will not necessarily occur rapidly by itself. Consider, for example, the combustion of glucose in oxygen:



$\Delta G^\circ = -686$ kcal/mole

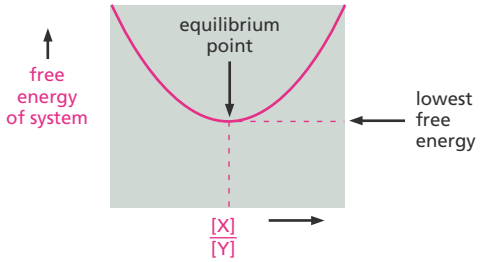
Even this highly favorable reaction may not occur for centuries unless there are enzymes to speed up the process.
Enzymes are able to catalyze reactions and speed up their rate, but they cannot change the ΔG° of a reaction.

CHEMICAL EQUILIBRIA

A fixed relationship exists between the standard free-energy change of a reaction, ΔG°, and its equilibrium constant K. For example, the reversible reaction

$$Y \rightleftarrows X$$

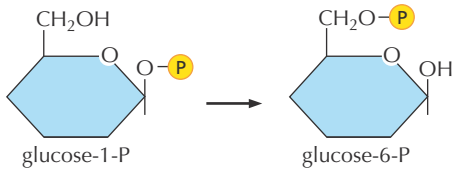
will proceed until the ratio of concentrations [X]/[Y] is equal to K (note: square brackets [ ] indicate concentration). At this point, the free energy of the system will have its lowest value.



At 37°C, $\Delta G^\circ = -1.42 \log_{10} K$ (see text, p. 98)

$K = 10^{-\Delta G^\circ/1.42}$

For example, the reaction



has $\Delta G^\circ = -1.74$ kcal/mole. Therefore, its equilibrium constant

$$K = 10^{(1.74/1.42)} = 10^{(1.23)} = 17$$

So the reaction will reach steady state when

[glucose-6-P]/[glucose-1-P] = 17

COUPLED REACTIONS

Reactions can be "coupled" together if they share one or more intermediates. In this case, the overall free-energy change is simply the sum of the individual ΔG° values. A reaction that is unfavorable (has a positive ΔG°) can for this reason be driven by a second, highly favorable reaction.

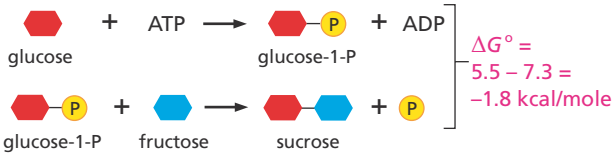
SINGLE REACTION

NET RESULT: reaction will not occur

ATP → ADP + P $\Delta G^\circ = -7.3$ kcal/mole

NET RESULT: reaction is highly favorable

COUPLED REACTIONS



NET RESULT: sucrose is made in a reaction driven by the hydrolysis of ATP

HIGH-ENERGY BONDS

One of the most common reactions in the cell is hydrolysis, in which a covalent bond is split by adding water.

The ΔG° for this reaction is sometimes loosely termed the "bond energy." Compounds such as acetyl phosphate and ATP, which have a large negative ΔG° of hydrolysis in an aqueous solution, are said to have "high-energy" bonds.

	ΔG° (kcal/mole)
acetyl P → acetate + P_i	–10.3
ATP → ADP + P_i	–7.3
glucose-6-P → glucose + P_i	–3.3

(Note that, for simplicity, H_2O is omitted from the above equations.)

But how do we know at what concentrations of substrate and product a reaction will reach equilibrium? It goes back to the intrinsic properties of the molecules involved, as expressed by $\Delta G°$. Let's see why.

At equilibrium, the rate of the forward reaction is exactly balanced by the rate of the reverse reaction. At that point, $\Delta G = 0$, and there is no net change of free energy to drive the reaction in either direction (see Panel 3–1, pp. 96–97).

Now, if we return to the equation presented on p. 94,

$$\Delta G = \Delta G° + RT \ln \frac{[X]}{[Y]}$$

we can see that, at equilibrium at 37°C, where $\Delta G = 0$ and the constant $RT = 0.616$, this equation becomes:

$$\Delta G° = -0.616 \ln \frac{[X]}{[Y]}$$

In other words, $\Delta G°$ is directly proportional to the equilibrium constant, K:

$$\Delta G° = -0.616 \ln K$$

If we convert this equation from natural log (ln) to the more commonly used base–10 logarithm (log), we get

$$\Delta G° = -1.42 \log K$$

This equation reveals how the equilibrium ratio of Y to X, expressed as the equilibrium constant K, depends on the intrinsic character of the molecules, as expressed in the value of $\Delta G°$ (**Table 3–1**). It tells us that for every 1.42 kcal/mole difference in free energy at 37°C, the equilibrium constant changes by a factor of 10. Thus, the more energetically favorable the reaction, the more product will accumulate if the reaction proceeds to equilibrium.

TABLE 3–1 RELATIONSHIP BETWEEN THE STANDARD FREE-ENERGY CHANGE, $\Delta G°$, AND THE EQUILIBRIUM CONSTANT

Equilibrium Constant $\frac{[X]}{[Y]}$	Standard Free Energy ($\Delta G°$) of X minus Free Energy of Y in kcal/mole
10^5	–7.1
10^4	–5.7
10^3	–4.3
10^2	–2.8
10	–1.4
1	0
10^{-1}	1.4
10^{-2}	2.8
10^{-3}	4.3
10^{-4}	5.7
10^{-5}	7.1

Values of the equilibrium constant were calculated for the simple chemical reaction Y ↔ X, using the equation given in the text.

The $\Delta G°$ values given here are in kilocalories per mole at 37°C. As explained in the text, $\Delta G°$ represents the free-energy difference under standard conditions (where all components are present at a concentration of 1 mole/liter).

From this table, we see that, if there is a favorable free-energy change of –4.3 kcal/mole for the transition Y → X, there will be 1000 times more molecules of X than of Y at equilibrium.

In Complex Reactions, the Equilibrium Constant Includes the Concentrations of All Reactants and Products

We have so far discussed the simplest of reactions, Y → X, in which a single substrate is converted into a single product. But inside cells, it is more common for two reactants to combine to form a single product: A + B ⇌ AB. How can we predict how this reaction will proceed?

The same principles apply, except that in this case the equilibrium constant K includes the concentrations of both of the reactants, in addition to the concentration of the product:

$$K = [AB]/[A][B]$$

As illustrated in **Figure 3–19**, the concentrations of both reactants are multiplied because the formation of product AB depends on the collision of A and B, and these encounters occur at a rate that is proportional to [A] × [B]. As with single-substrate reactions, $\Delta G° = -1.42 \log K$ at 37°C.

The Equilibrium Constant Indicates the Strength of Molecular Interactions

The concept of free-energy change does not only apply to chemical reactions where covalent bonds are being broken and formed, but also to interactions where one molecule binds to another by means of noncovalent interactions (see Chapter 2, p. 63). Noncovalent interactions are immensely important to cells. They include the binding of substrates to enzymes, the binding of gene regulatory proteins to DNA, and the binding of one protein to another to make the many different structural and functional protein complexes that operate in a living cell.

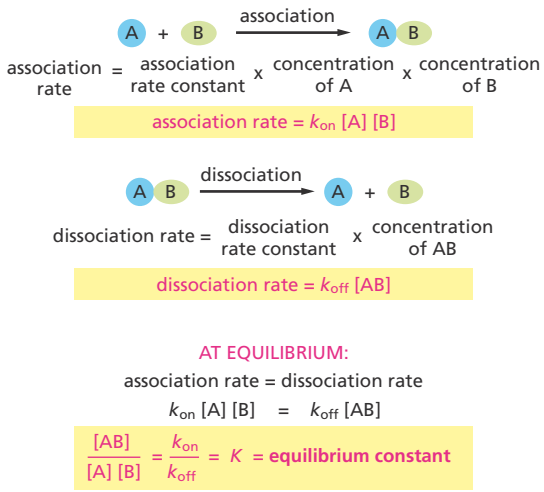


Figure 3–19 The equilibrium constant (*K*) for the reaction A + B → AB depends on both the association and dissociation rate constants. Molecules A and B must collide in order to interact, and the association rate is therefore proportional to the product of their individual concentrations [A] × [B]. As shown, the ratio of the rate constants k_{on} and k_{off} for the association and the dissociation reactions, respectively, is equal to the equilibrium constant, *K*, for the interaction. For two interacting components, *K* involves the concentrations of both substrates, in addition to that of the product. However, the relationship between *K* and $\Delta G°$ is the same as that shown in Table 3–1. The larger the value of *K*, the stronger is the binding between A and B.

Two molecules will bind to each other if the free-energy change for the interaction is negative; that is, the free energy of the resulting complex is lower than the sum of the free energies of the two partners when unbound. Because the equilibrium constant of a reaction is related directly to $\Delta G°$, *K* is commonly employed as a measure of the binding strength of a noncovalent interaction between two molecules. The binding strength is a very useful quantity to know because it also indicates how specific the interaction is between the two molecules.

Consider the reaction that was shown in Figure 3–19, where molecule A interacts with molecule B to form the complex AB. The reaction proceeds until it reaches equilibrium, at which point the number of association events precisely equals the number of dissociation events; at this point, the concentrations of reactants A and B, and of the complex AB, can be used to determine the equilibrium constant *K*.

K becomes larger as the *binding energy*—that is, the energy released in the binding interaction—increases. In other words, the larger *K* is, the greater is the drop in free energy between the dissociated and associated states, and the more tightly the two molecules will bind. Even a change of a few noncovalent bonds can have a striking effect on a binding interaction, as illustrated in **Figure 3–20**. In this example, eliminating a few hydrogen bonds from a binding interaction can be seen to cause a dramatic decrease in the amount of complex that exists at equilibrium.

For Sequential Reactions, the Changes in Free Energy Are Additive

Now we return to our original concern: how can enzymes catalyze reactions that are energetically unfavorable? One way they do so is by directly coupling energetically unfavorable reactions with energetically favorable ones. Consider, for example, two sequential reactions,

$$X \rightarrow Y \text{ and } Y \rightarrow Z$$

where the $\Delta G°$ values are +5 and –13 kcal/mole, respectively. (Recall that a mole is 6×10^{23} molecules of a substance.) The unfavorable reaction, X → Y, will not occur spontaneously. However, it can be driven by the favorable reaction Y → Z, provided that the second reaction follows the first. That's because the overall free-energy change for the coupled reaction is equal to the sum of the free-energy changes for each individual reaction. In this case, the $\Delta G°$ for the coupled reaction will be –8 kcal/mole, making the overall pathway energetically favorable.

Consider 1000 molecules of A and 1000 molecules of B in the cytosol of a eukaryotic cell. The concentration of both will be about 10^{-9} M.

If the equilibrium constant (*K*) for A + B ↔ AB is 10^{10} liters/mole, then at equilibrium there will be

270	270	730
A molecules	B molecules	AB complexes

If the equilibrium constant is a little weaker, say 10^{8} liters/mole—a value that represents a loss of 2.8 kcal/mole of binding energy from the example above, or 2–3 fewer hydrogen bonds—then there will be

915	915	85
A molecules	B molecules	AB complexes

Figure 3–20 Small changes in the number of weak bonds can have drastic effects on a binding interaction. This example illustrates the dramatic effect of the presence or absence of a few weak noncovalent bonds in the interaction between two cytosolic proteins.

Cells can therefore cause the energetically unfavorable transition, X → Y, to occur if an enzyme catalyzing the X → Y reaction is supplemented by a second enzyme that catalyzes the energetically favorable reaction, Y → Z. In effect, the reaction Y → Z acts as a "siphon," pulling the conversion of all of molecule X to molecule Y, and then to molecule Z (**Figure 3–21**). For example, several of the reactions in the long pathway that converts sugars into CO_2 and H_2O are energetically unfavorable. The pathway nevertheless proceeds rapidly to completion, however, because the total $\Delta G°$ for the series of sequential reactions has a large negative value.

Forming a sequential pathway, however, is not the answer for all metabolic needs. Often the desired reaction is simply X → Y, without further conversion of Y to some other product. Fortunately, there are other, more general ways of using enzymes to couple reactions together, involving the production of activated carriers that can shuttle energy from one reaction site to another. We discuss these systems shortly. Before we do, let's pause to look at how enzymes find and recognize their substrates and how enzyme-catalyzed reactions proceed. After all, thermodynamic considerations merely establish whether chemical reactions can occur; enzymes actually make them happen.

QUESTION 3–4

For the reactions shown in Figure 3–21, sketch an energy diagram similar to that in Figure 3–12 for the two reactions alone and for the combined reactions. Indicate the standard free-energy changes for the reactions X → Y, Y → Z, and X → Z in the graph. Indicate how enzymes that catalyze these reactions would change the energy diagram.

Thermal Motion Allows Enzymes to Find Their Substrates

Enzymes and their substrates are both present in relatively small amounts in the cytosol of a cell, yet a typical enzyme can capture and process about a thousand substrate molecules every second. This means that an enzyme can release its product and bind a new substrate in a fraction of a millisecond. How do these molecules find each other so quickly in the crowded cytosol of the cell?

Rapid binding is possible because molecular motions are enormously fast. Because of heat energy, molecules are in constant motion and consequently will explore the cytosolic space very efficiently by wandering

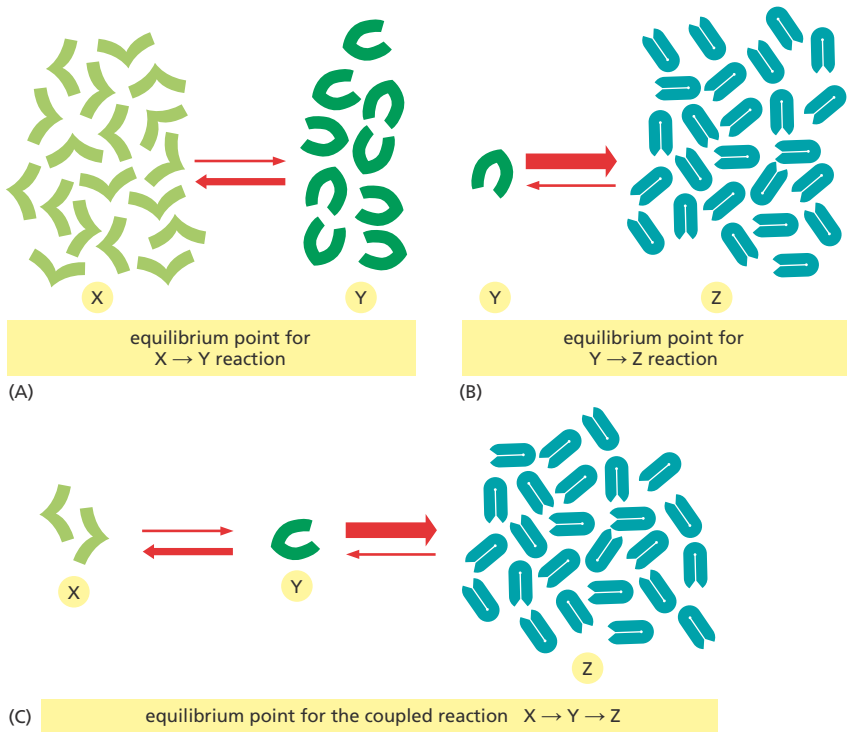


Figure 3–21 An energetically unfavorable reaction can be driven by an energetically favorable follow-on reaction that acts as a chemical siphon. (A) At equilibrium, there are twice as many X molecules as Y molecules. (B) At equilibrium, there are 25 times more Z molecules than Y molecules. (C) If the reactions in (A) and (B) are coupled, nearly all of the X molecules will be converted to Z molecules, as shown. In terms of energetics, the $\Delta G°$ of the Y → Z reaction is so negative that, when coupled to the X → Y reaction, it lowers the ΔG of X → Y, because the ΔG of X → Y decreases as the ratio of Y to X declines. As shown in Figure 3–18, arrow widths reflect the relative rates at which an individual molecule converts; the arrow lengths are the same in both directions here, indicating that there is no net flux.

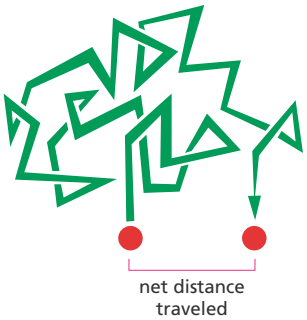


Figure 3–22 A molecule traverses the cytosol by taking a random walk. Molecules in solution move in a random fashion due to the continual buffeting they receive in collisions with other molecules. This movement allows small molecules to diffuse rapidly throughout the cell cytosol (**Movie 3.2**).

randomly through it—a process called **diffusion**. In this way, every molecule in the cytosol collides with a huge number of other molecules each second. As the molecules in a liquid collide and bounce off one another, an individual molecule moves first one way and then another, its path constituting a *random walk* (**Figure 3–22**).

Although the cytosol of a cell is densely packed with molecules of various shapes and sizes (**Figure 3–23**), experiments in which fluorescent dyes and other labeled molecules are injected into the cell cytosol show that small organic molecules diffuse through this aqueous gel nearly as rapidly as they do through water. A small organic molecule, such as a substrate, takes only about one-fifth of a second on average to diffuse a distance of 10 μm. Diffusion is therefore an efficient way for small molecules to move limited distances in the cell.

Because proteins diffuse through the cytosol much more slowly than do small molecules, the rate at which an enzyme will encounter its substrate depends on the concentration of the substrate. The most abundant substrates are present in the cell at a concentration of about 0.5 mM. Because pure water is 55 M, there is only about one such substrate molecule in the cell for every 10^5 water molecules. Nevertheless, the site on an enzyme that binds this substrate will be bombarded by about 500,000 random collisions with the substrate every second. For a substrate concentration tenfold lower (0.05 mM), the number of collisions drops to 50,000 per second, and so on.

The random encounters between an enzyme and its substrate often lead to the formation of an enzyme–substrate complex. This association is stabilized by the formation of multiple, weak bonds between the enzyme and substrate. These weak interactions—which can include hydrogen bonds, van der Waals attractions, and electrostatic attractions (discussed in Chapter 2)—persist until random thermal motion causes the molecules to dissociate again. When two colliding molecules have poorly matching surfaces, few noncovalent bonds are formed, and their total energy is negligible compared with that of thermal motion. In this case, the two molecules dissociate as rapidly as they come together (see Figure 2–33). This is what prevents incorrect and unwanted associations from forming between mismatched molecules, such as those between an enzyme and the wrong substrate. But when the enzyme and substrate are well matched, they form many weak interactions, which keep them held together long enough for a covalent bond in the substrate molecule to be formed or broken. Knowing the speed at which molecules collide and come apart, as well as how fast bonds can be formed and broken, makes the observed rate of enzymatic catalysis seem a little less amazing.

QUESTION 3–5

The enzyme carbonic anhydrase is one of the speediest enzymes known. It catalyzes the rapid conversion of CO_2 gas into the much more soluble bicarbonate ion (HCO_3^-). The reaction:

$$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+$$

is very important for the efficient transport of CO_2 from tissue, where CO_2 is produced by respiration, to the lungs, where it is exhaled. Carbonic anhydrase accelerates the reaction 10^7-fold, hydrating 10^5 CO_2 molecules per second at its maximal speed. What do you suppose limits the speed of the enzyme? Sketch a diagram analogous to the one shown in Figure 3–13 and indicate which portion of your diagram has been designed to display the 10^7-fold acceleration.

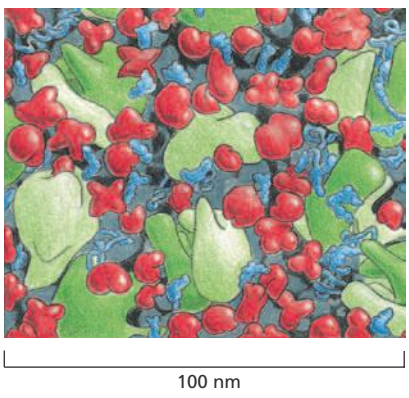


Figure 3–23 The cytosol is crowded with various molecules. Only the macromolecules, which are drawn to scale, are shown. RNAs are *blue*, ribosomes are *green*, and proteins are *red*. Enzymes and other macromolecules diffuse relatively slowly in the cytosol, in part because they interact with so many other macromolecules. Small molecules, by contrast, can diffuse nearly as rapidly as they do in water. (Adapted from D.S. Goodsell, *Trends Biochem. Sci.* 16:203–206, 1991. With permission from Elsevier.)

V_{max} and K_M Measure Enzyme Performance

To catalyze a reaction, an enzyme must first bind its substrate. The substrate then undergoes a reaction to form the product, which initially remains bound to the enzyme. Finally, the product is released and diffuses away, leaving the enzyme free to bind another substrate molecule and catalyze another reaction (see Figure 3–15). The rates of the different steps vary widely from one enzyme to another, and they can be measured by mixing purified enzymes and substrates together under carefully defined conditions in a test tube (see **How We Know**, pp. 104–106).

In such experiments, the substrate is introduced in increasing concentrations to a solution containing a fixed concentration of enzyme. At first, the concentration of the enzyme–substrate complex—and therefore the rate at which product is formed—rises in a linear fashion in direct proportion to substrate concentration. However, as more and more enzyme molecules become occupied by substrate, this rate increase tapers off, until at a very high concentration of substrate it reaches a maximum value, termed V_{max}. At this point, the active sites of all enzyme molecules in the sample are fully occupied by substrate, and the rate of product formation depends only on how rapidly the substrate molecule can undergo a reaction to form the product. For many enzymes, this **turnover number** is of the order of 1000 substrate molecules per second, although turnover numbers between 1 and 100,000 have been measured.

Because there is no clearly defined substrate concentration at which the enzyme can be deemed fully occupied, biochemists instead use a different parameter to gauge the concentration of substrate needed to make the enzyme work efficiently. This value is called the **Michaelis constant**, K_M, named after one of the biochemists who worked out the relationship. The K_M of an enzyme is defined as the concentration of substrate at which the enzyme works at half its maximum speed (**Figure 3–24**). In general, a small K_M indicates that a substrate binds very tightly to the enzyme, and a large K_M indicates weak binding.

Although an enzyme (or any catalyst) functions to lower the activation energy for a reaction such as Y → X, it is important to note that the enzyme will also lower the activation energy for the reverse reaction X → Y to exactly the same degree. The forward and backward reactions will therefore be accelerated by the same factor by an enzyme, and the equilibrium point for the reaction—and thus its $\Delta G°$—remains unchanged (**Figure 3–25**).

QUESTION 3–6

In cells, an enzyme catalyzes the reaction AB → A + B. It was isolated, however, as an enzyme that carries out the opposite reaction A + B → AB. Explain the paradox.

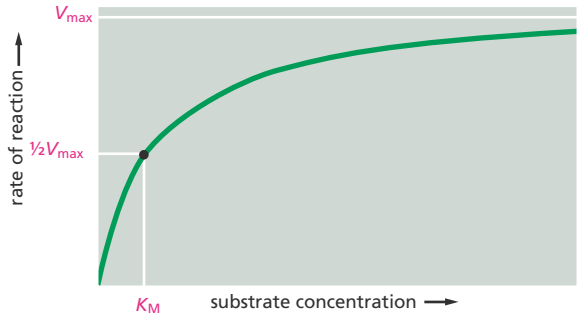


Figure 3–24 An enzyme's performance depends on how rapidly it can process its substrate. The rate of an enzyme reaction (V) increases as the substrate concentration increases, until a maximum value (V_{max}) is reached. At this point, all substrate-binding sites on the enzyme molecules are fully occupied, and the rate of the reaction is limited by the rate of the catalytic process on the enzyme surface. For most enzymes, the concentration of substrate at which the reaction rate is half-maximal (K_M) is a direct measure of how tightly the substrate is bound, with a large value of K_M (a large amount of substrate needed) corresponding to weak binding.

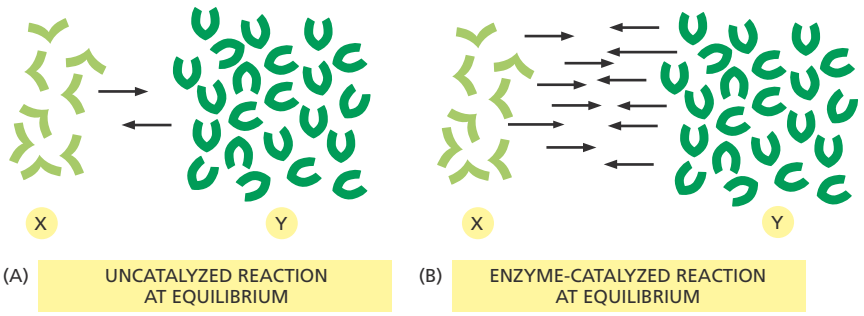


Figure 3–25 Enzymes cannot change the equilibrium point for reactions. Enzymes, like all catalysts, speed up the forward and reverse rates of a reaction by the same amount. Therefore, for both the (A) uncatalyzed and (B) catalyzed reactions shown here, the number of molecules undergoing the transition X ⟶ Y is equal to the number of molecules undergoing the transition Y ⟶ X when the ratio of Y molecules to X molecules is 3.5 to 1, as illustrated. In other words, both the catalyzed and uncatalyzed reactions will eventually reach the same equilibrium point, although the catalyzed reaction will reach equilibrium faster.

ACTIVATED CARRIERS AND BIOSYNTHESIS

The energy released by energetically favorable reactions such as the oxidation of food molecules must be stored temporarily before it can be used by cells to fuel energetically unfavorable reactions, such as the synthesis of all the other molecules needed by the cell. In most cases, the energy is stored as chemical-bond energy in a set of *activated carriers*, small organic molecules that contain one or more energy-rich covalent bonds. These molecules diffuse rapidly and carry their bond energy from the sites of energy generation to the sites where energy is used for **biosynthesis** or for other energy-requiring cell activities (**Figure 3–26**).

Activated carriers store energy in an easily exchangeable form, either as a readily transferable chemical group or as readily transferable ("high energy") electrons. They can serve a dual role as a source of both energy and chemical groups for biosynthetic reactions. The most important activated carriers are *ATP* and two molecules that are closely related to each other, *NADH* and *NADPH*. Cells use activated carriers like money to pay for the energetically unfavorable reactions that otherwise would not take place.

The Formation of an Activated Carrier Is Coupled to an Energetically Favorable Reaction

When a fuel molecule such as glucose is oxidized in a cell, enzyme-catalyzed reactions ensure that a large part of the free energy released is captured in a chemically useful form, rather than being released wastefully as heat. (Oxidizing sugar in a cell allows you to power metabolic reactions, whereas burning a chocolate bar in the street will get you nowhere, producing no metabolically useful energy.) In cells, energy capture is achieved by means of a **coupled reaction**, in which an energetically favorable reaction is used to drive an energetically unfavorable one that produces an activated carrier or some other useful molecule.

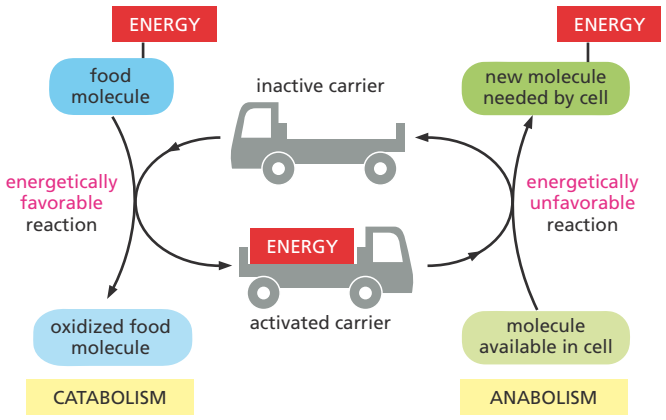


Figure 3–26 Activated carriers can store and transfer energy in a form that cells can use. By serving as intracellular energy shuttles, activated carriers perform their function as go-betweens that link the release of energy from the breakdown of food molecules (catabolism) to the energy-requiring biosynthesis of small and large organic molecules (anabolism).

MEASURING ENZYME PERFORMANCE

At first glance, it seems that a cell's metabolic pathways have been pretty well mapped out, with each reaction proceeding predictably to the next—substrate X is converted to product Y, which is passed along to enzyme Z. So why would anyone need to know exactly how tightly a particular enzyme clutches its substrate or whether it can process 100 or 1000 substrate molecules every second?

In reality, such elaborate metabolic maps merely suggest which pathways a cell might follow as it converts nutrients into small molecules, chemical energy, and the larger building blocks of life. Like a road map, they do not predict the density of traffic under a particular set of conditions: which pathways the cell will use when it is starving, when it is well fed, when oxygen is scarce, when it is stressed, or when it decides to divide. The study of an enzyme's *kinetics*—how fast it operates, how it handles its substrate, how its activity is controlled—makes it possible to predict how an individual catalyst will perform, and how it will interact with other enzymes in a network. Such knowledge leads to a deeper understanding of cell biology, and it opens the door to learning how to harness enzymes to perform desired reactions.

Speed

The first step to understanding how an enzyme performs involves determining the maximal velocity, V_{max}, for the reaction it catalyzes. This is accomplished by measuring, in a test tube, how rapidly the reaction proceeds in the presence of different concentrations of substrate (**Figure 3–27A**): the rate should increase as the amount of substrate rises until the reaction reaches its V_{max}. The velocity of the reaction is measured by monitoring either how quickly the substrate is consumed or how rapidly the product accumulates. In many cases, the appearance of product or the disappearance of substrate can be observed directly with a spectrophotometer. This instrument detects the presence of molecules that absorb light at a particular wavelength; NADH, for example, absorbs light at 340 nm, while its oxidized counterpart, NAD^+, does not. So, a reaction that generates NADH (by reducing NAD^+) can be monitored by following the formation of NADH at 340 nm in a spectrophotometer.

To determine the V_{max} of a reaction, you would set up a series of test tubes, where each tube contains a different concentration of substrate. For each tube, add the same amount of enzyme and then measure the velocity of the reaction—the number of micromoles of substrate consumed or product generated per minute. Because these numbers will tend to decrease over time, the rate used is the velocity measured early in the reaction. These initial velocity values (v) are then plotted against the substrate concentration, yielding a curve like the one shown in **Figure 3–27B**.

Looking at this plot, however, it is difficult to determine the exact value of V_{max}, as it is not clear where the reaction rate will reach its plateau. To get around this problem, the data are converted to their reciprocals

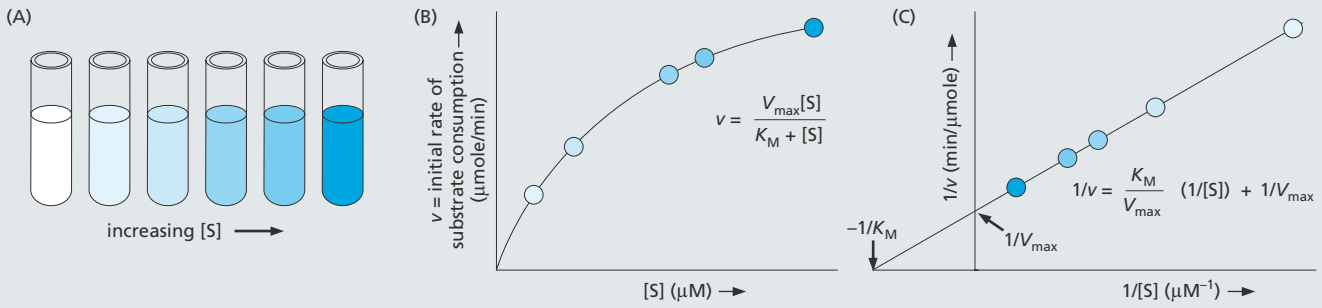


Figure 3–27 Measured reaction rates are plotted to determine V_{max} and K_M of an enzyme-catalyzed reaction. (A) A series of increasing substrate concentrations is prepared, a fixed amount of enzyme is added, and initial reaction rates (velocities) are determined. (B) The initial velocities (v) plotted against the substrate concentrations [S] give a curve described by the general equation $y = ax/(b + x)$. Substituting our kinetic terms, the equation becomes $v = V_{max}[S]/(K_M + [S])$, where V_{max} is the asymptote of the curve (the value of y at an infinite value of x), and K_M is equal to the substrate concentration where v is one-half V_{max}. This is called the *Michaelis–Menten equation*, named for the biochemists who provided evidence for this enzymatic relationship. (C) In a double-reciprocal plot, $1/v$ is plotted against 1/[S]. The equation describing this straight line is $1/v = (K_M/V_{max})(1/[S]) + 1/V_{max}$. When 1/[S] = 0, the y intercept ($1/v$) is $1/V_{max}$. When $1/v = 0$, the x intercept (1/[S]) is $-1/K_M$. Plotting the data this way allows V_{max} and K_M to be calculated more precisely. By convention, lowercase letters are used for variables (hence v for velocity) and uppercase letters are used for constants (hence V_{max}).

and graphed in a "double-reciprocal plot," where the inverse of the velocity ($1/v$) appears on the y axis and the inverse of the substrate concentration ($1/[S]$) on the x axis (**Figure 3–27C**). This graph yields a straight line whose y intercept (the point where the line crosses the y axis) represents $1/V_{max}$ and whose x intercept corresponds to $-1/K_M$. These values are then converted to values for V_{max} and K_M.

Enzymologists use this technique to determine the kinetic parameters of many enzyme-catalyzed reactions (although these days computer programs automatically plot the data and spit out the sought-after values). Some reactions, however, happen too fast to be monitored in this way; the reaction is essentially complete—the substrate entirely consumed—within thousandths of a second. For these reactions, a special piece of equipment must be used to follow what happens during the first few milliseconds after enzyme and substrate meet (**Figure 3–28**).

Control

Substrates are not the only molecules that can influence how well or how quickly an enzyme works. In many cases, products, substrate lookalikes, inhibitors, and other small molecules can also increase or decrease enzyme activity. Such regulation allows cells to control when and how rapidly various reactions occur, a process we will consider in more detail in Chapter 4.

Determining how an inhibitor decreases an enzyme's activity can reveal how a metabolic pathway is regulated—and can suggest how those control points can be circumvented by carefully designed mutations in specific genes.

The effect of an inhibitor on an enzyme's activity is monitored in the same way that we measured the enzyme's kinetics. A curve is first generated showing the velocity of the uninhibited reaction between enzyme and substrate, as described previously. Additional curves are then produced for reactions in which the inhibitor molecule has been included in the mix.

Comparing these curves, with and without inhibitor, can also reveal how a particular inhibitor impedes enzyme activity. For example, some inhibitors bind to the same site on an enzyme as its substrate. These *competitive inhibitors* block enzyme activity by competing directly with the substrate for the enzyme's attention. They resemble the substrate enough to tie up the enzyme, but they differ enough in structure to avoid getting converted to product. This blockage can be overcome by adding enough substrate so that enzymes are more likely to encounter a substrate molecule than an inhibitor molecule. From the kinetic data, we can see that competitive inhibitors do not change the V_{max} of a reaction; in other words, add enough substrate and the enzyme will encounter mostly substrate molecules and will reach its maximum velocity (**Figure 3–29**).

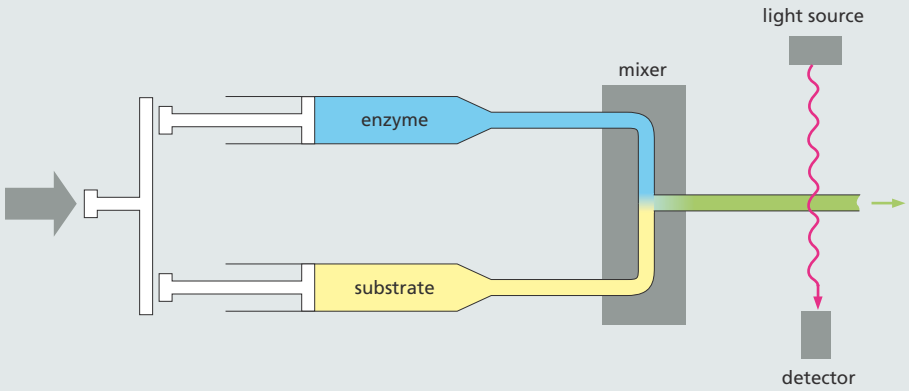


Figure 3–28 A stopped-flow apparatus is used to observe reactions during the first few milliseconds. In this piece of equipment, the enzyme and substrate are rapidly injected into a mixing chamber through two syringes. The enzyme and its substrate meet as they shoot through the mixing tube at flow rates that can easily reach 1000 cm/sec. They then enter another tube and zoom past a detector that monitors, say, the appearance of product. If the detector is located within a centimeter of where the enzyme and substrate meet, it is possible to observe reactions when they are only a few milliseconds old.

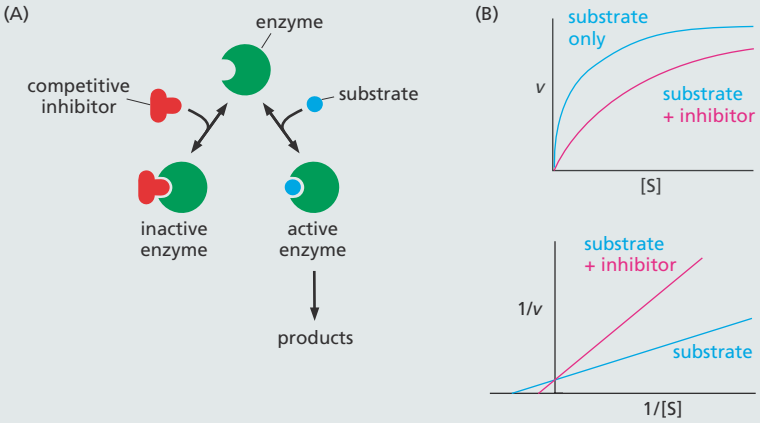


Figure 3–29 A competitive inhibitor directly blocks substrate binding to an enzyme. (A) The active site of the enzyme can bind either the competitive inhibitor or the substrate, but not both together. (B) The upper plot shows that inhibition by a competitive inhibitor can be overcome by increasing the substrate concentration. The double-reciprocal plot below shows that the V_{max} of the reaction is not changed in the presence of the competitive inhibitor: the *y* intercept is identical for both the curves.

Competitive inhibitors can be used to treat patients who have been poisoned by ethylene glycol, an ingredient in commercially available antifreeze. Although ethylene glycol is itself not fatally toxic, a by-product of its metabolism—oxalic acid—can be lethal. To prevent oxalic acid from forming, the patient is given a large (though not quite intoxicating) dose of ethanol. Ethanol competes with the ethylene glycol for binding to alcohol dehydrogenase, the first enzyme in the pathway to oxalic acid formation. As a result, the ethylene glycol goes mostly unmetabolized and is safely eliminated from the body.

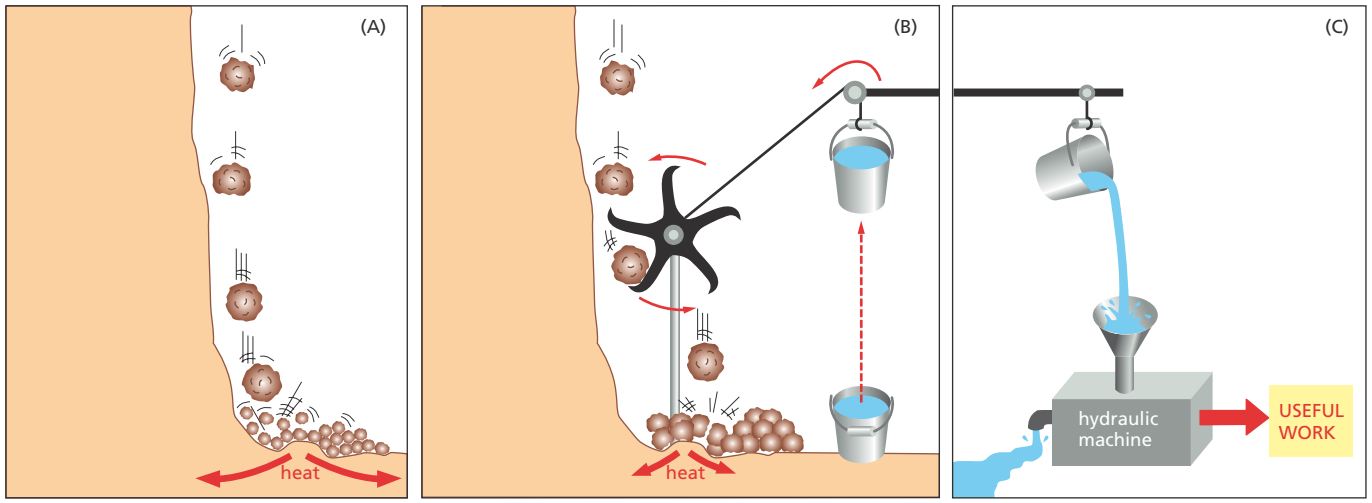
Other types of inhibitors may interact with sites on the enzyme distant from where the substrate binds. As we discuss in Chapter 4, many biosynthetic enzymes are regulated by feedback inhibition, whereby an enzyme early in a pathway will be shut down by a product generated later in the pathway. Because this type of inhibitor binds to a separate regulatory site on the enzyme, the substrate can still bind, but it might do so more slowly than it would in the absence of inhibitor. Such *noncompetitive inhibition* is not overcome by the addition of more substrate.

Design

With the kinetic data in hand, we can use computer modeling programs to predict how an enzyme will perform, and even how a cell will respond when exposed to different conditions—such as the addition of a particular sugar or amino acid to the culture medium, or the addition of a poison or a pollutant. Seeing how a cell manages its resources—which pathways it favors for dealing with particular biochemical challenges—can also suggest strategies for designing better catalysts for reactions of medical or commercial importance (e.g., for producing drugs or detoxifying industrial waste). Using such tactics, bacteria have even been genetically engineered to produce large amounts of indigo—the dye, originally extracted from plants, that makes your blue jeans blue.

Computer programs have been developed to facilitate the dissection of complex reaction pathways. They require information about the components in the pathway, including the K_M and V_{max} of the participating enzymes and the concentrations of enzymes, substrates, products, inhibitors, and other regulatory molecules. The program then predicts how molecules will flow through the pathway, which products will be generated, and where any bottlenecks might be. The process is not unlike balancing an algebraic equation, in which every atom of carbon, nitrogen, oxygen, and so on must be tallied. Such careful accounting makes it possible to rationally design ways to manipulate the pathway, such as re-routing it around a bottleneck, eliminating an important inhibitor, redirecting the reactions to favor the generation of predominantly one product, or extending the pathway to produce a novel molecule. Of course, such computer models must be validated in cells, which may not always behave as predicted.

Producing designer cells that spew out commercial products generally requires using genetic engineering techniques to introduce the gene or genes of choice into a cell, usually a bacterium, that can be manipulated and maintained in the laboratory. We discuss these methods at greater length in Chapter 10. Harnessing the power of cell biology for commercial purposes—even to produce something as simple as the amino acid tryptophan—is currently a multibillion-dollar industry. And, as more genome data come in, presenting us with more enzymes to exploit, it may not be long before vats of custom-made bacteria are churning out drugs and chemicals that represent the biological equivalent of pure gold.



kinetic energy of falling rocks is transformed into heat energy only

part of the kinetic energy is used to lift a bucket of water, and a correspondingly smaller amount is transformed into heat

the potential energy stored in the raised bucket of water can be used to drive hydraulic machines that carry out a variety of useful tasks

Figure 3–30 A mechanical model illustrates the principle of coupled chemical reactions. The spontaneous reaction shown in (A) could serve as an analogy for the direct oxidation of glucose to CO_2 and H_2O, which produces only heat. In (B), the same reaction is coupled to a second reaction, which could serve as an analogy for the synthesis of activated carriers. The energy produced in (B) is in a more useful form than in (A) and can be used to drive a variety of otherwise energetically unfavorable reactions (C).

Such coupling requires enzymes, which are fundamental to all of the energy transactions in the cell.

The nature of a coupled reaction is illustrated by a mechanical analogy in **Figure 3–30**, in which an energetically favorable chemical reaction is represented by rocks falling from a cliff. The kinetic energy of falling rocks would normally be entirely wasted in the form of heat generated by friction when the rocks hit the ground (Figure 3–30A). By careful design, however, part of this energy could be used to drive a paddle wheel that lifts a bucket of water (Figure 3–30B). Because the rocks can now reach the ground only after moving the paddle wheel, we say that the energetically favorable reaction of rocks falling has been directly coupled to the energetically unfavorable reaction of lifting the bucket of water. Because part of the energy is used to do work in (B), the rocks hit the ground with less velocity than in (A), and correspondingly less energy is wasted as heat. The energy saved in the elevated bucket of water can then be used to do useful work (Figure 3–30C).

Analogous processes occur in cells, where enzymes play the role of the paddle wheel in Figure 3–30B. By mechanisms that we discuss in Chapter 13, enzymes couple an energetically favorable reaction, such as the oxidation of foodstuffs, to an energetically unfavorable reaction, such as the generation of activated carriers. As a result, the amount of heat released by the oxidation reaction is reduced by exactly the amount of energy that is stored in the energy-rich covalent bonds of the activated carrier. That saved energy can then be used to power a chemical reaction elsewhere in the cell.

ATP Is the Most Widely Used Activated Carrier

The most important and versatile of the activated carriers in cells is **ATP** (adenosine 5′-triphosphate). Just as the energy stored in the raised bucket of water in Figure 3–30B can be used to drive a wide variety of hydraulic machines, ATP serves as a convenient and versatile store, or currency, of energy that can be used to drive a variety of chemical reactions in cells.

QUESTION 3–7

Use Figure 3–30B to illustrate the following reaction driven by the hydrolysis of ATP:
$X + ATP \rightarrow Y + ADP + P_i$
A. In this case, which molecule or molecules would be analogous to (i) rocks at top of cliff, (ii) broken debris at bottom of cliff, (iii) bucket at its highest point, and (iv) bucket on the ground?
B. What would be analogous to (i) the rocks hitting the ground in the absence of the paddle wheel in Figure 3–30A and (ii) the hydraulic machine in Figure 3–30C?

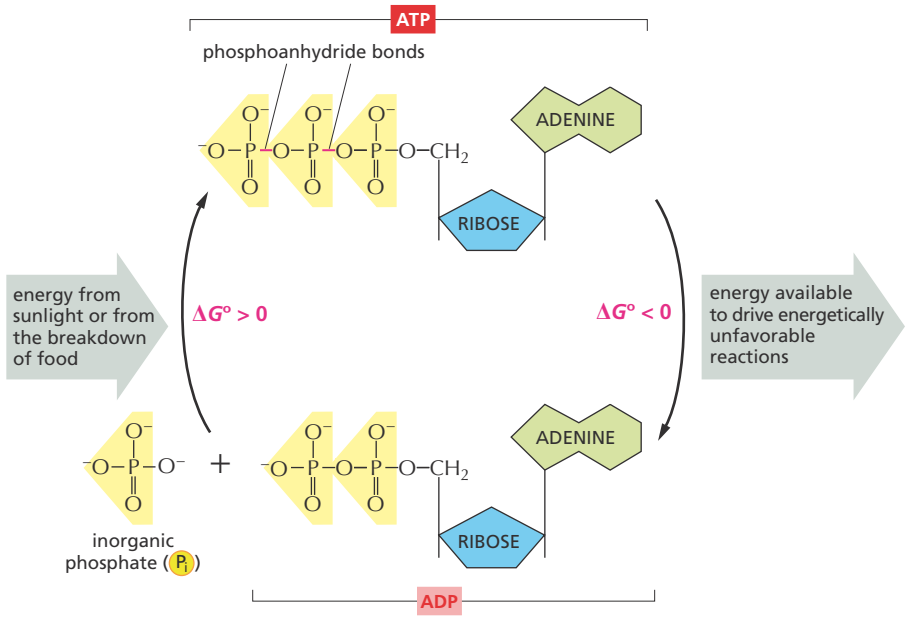


Figure 3–31 The interconversion of ATP and ADP occurs in a cycle. The two outermost phosphate groups in ATP are held to the rest of the molecule by high-energy phosphoanhydride bonds and are readily transferred to other organic molecules. Water can be added to ATP to form ADP and inorganic phosphate (P_i). Inside a cell, this hydrolysis of the terminal phosphate of ATP yields between 11 and 13 kcal/mole of usable energy. Although the $\Delta G°$ of this reaction is –7.3 kcal/mole, the ΔG is much more negative, because the ratio of ATP to the products ADP and P_i is so high inside the cell.

The large negative $\Delta G°$ of the reaction arises from a number of factors. Release of the terminal phosphate group removes an unfavorable repulsion between adjacent negative charges; in addition, the inorganic phosphate ion (P_i) released is stabilized by favorable hydrogen-bond formation with water. The formation of ATP from ADP and P_i reverses the hydrolysis reaction; because this condensation reaction is energetically unfavorable, it must be coupled to an energetically more favorable reaction to occur.

As shown in **Figure 3–31**, ATP is synthesized in an energetically unfavorable *phosphorylation* reaction, in which a phosphate group is added to **ADP** (adenosine 5′-diphosphate). When required, ATP gives up this energy packet in an energetically favorable hydrolysis to ADP and inorganic phosphate (P_i). The regenerated ADP is then available to be used for another round of the phosphorylation reaction that forms ATP, creating an ATP cycle in the cell.

The energetically favorable reaction of ATP hydrolysis is coupled to many otherwise unfavorable reactions through which other molecules are synthesized. We will encounter several of these reactions in this chapter, where we will see exactly how this is done. ATP hydrolysis is often coupled to the transfer of the terminal phosphate in ATP to another molecule, as illustrated in **Figure 3–32**. Any reaction that involves the transfer of a phosphate group to a molecule is termed a phosphorylation reaction. Phosphorylation reactions are examples of condensation reactions (see Figure 2–25), and they occur in many important cell processes: they activate substrates, mediate the exchange of chemical energy, and serve as key constituents of intracellular signaling pathways (discussed in Chapter 16).

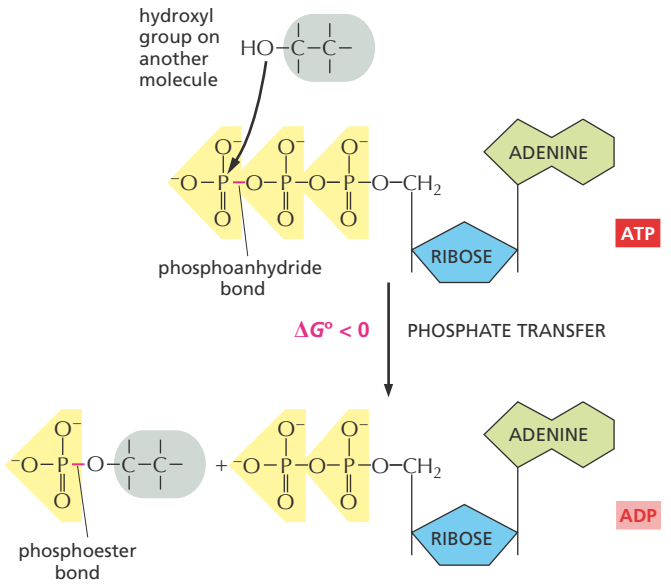


Figure 3–32 The terminal phosphate of ATP can be readily transferred to other molecules. Because an energy-rich phosphoanhydride bond in ATP is converted to a less energy-rich phosphoester bond in the phosphate-accepting molecule, this reaction is energetically favorable, having a large negative $\Delta G°$ (see Panel 3–1, pp. 96–97). Phosphorylation reactions of this type are involved in the synthesis of phospholipids and in the initial steps of the breakdown of sugars, as well as in many other metabolic and intracellular signaling pathways.

ATP is the most abundant activated carrier in cells. It is used, for example, to supply energy for many of the pumps that actively transport substances into or out of the cell (discussed in Chapter 12); it also powers the molecular motors that enable muscle cells to contract and nerve cells to transport materials along their lengthy axons (discussed in Chapter 17). Why evolution selected this particular nucleotide over the others as the major carrier of energy, however, remains a mystery. The nucleotide GTP, although similar, has very different functions in the cell, as we discuss in later chapters.

Energy Stored in ATP Is Often Harnessed to Join Two Molecules Together

A common type of reaction that is needed for biosynthesis is one in which two molecules, A and B, are joined together by a covalent bond to produce A–B in the energetically unfavorable condensation reaction:

$$A\text{–}H + B\text{–}OH \rightarrow A\text{–}B + H_2O$$

ATP hydrolysis can be coupled indirectly to this reaction to make it go forward. In this case, energy from ATP hydrolysis is first used to convert B–OH to a higher-energy intermediate compound, which then reacts directly with A–H to give A–B. The simplest mechanism involves the transfer of a phosphate from ATP to B–OH to make $B\text{–}O\text{–}PO_3$, in which case the reaction pathway contains only two steps:

1. $B\text{–}OH + ATP \rightarrow B\text{–}O\text{–}PO_3 + ADP$
2. $A\text{–}H + B\text{–}O\text{–}PO_3 \rightarrow A\text{–}B + P_i$

Net result: $B\text{–}OH + ATP + A\text{–}H \rightarrow A\text{–}B + ADP + P_i$

The condensation reaction, which by itself is energetically unfavorable, has been forced to occur by being coupled to ATP hydrolysis in an enzyme-catalyzed reaction pathway (**Figure 3–33A**).

A biosynthetic reaction of exactly this type is employed to synthesize the amino acid glutamine, as illustrated in **Figure 3–33B**. We will see later in the chapter that very similar (but more complex) mechanisms are also used to produce nearly all of the large molecules of the cell.

NADH and NADPH Are Both Activated Carriers of Electrons

Other important activated carriers participate in oxidation–reduction reactions and are commonly part of coupled reactions in cells. These activated carriers are specialized to carry both high-energy electrons and hydrogen atoms. The most important of these *electron carriers* are **NADH** (nicotinamide adenine dinucleotide) and the closely related molecule **NADPH** (nicotinamide adenine dinucleotide phosphate). Both NADH and NADPH carry energy in the form of two high-energy electrons plus a proton (H^+), which together form a hydride ion (H^-). When these activated carriers pass their energy (in the form of a hydride ion) to a donor molecule, they become oxidized to form **NAD^+** and **$NADP^+$**, respectively.

Like ATP, NADPH is an activated carrier that participates in many important biosynthetic reactions that would otherwise be energetically unfavorable. NADPH is produced according to the general scheme shown in **Figure 3–34A**. During a special set of energy-yielding catabolic reactions, a hydride ion is removed from the substrate molecule and added to the nicotinamide ring of $NADP^+$ to form NADPH. This is a typical oxidation–reduction reaction; the substrate is oxidized and $NADP^+$ is reduced.

QUESTION 3–8

The phosphoanhydride bond that links two phosphate groups in ATP in a high-energy linkage has a $\Delta G°$ of –7.3 kcal/mole. Hydrolysis of this bond in a cell liberates from 11 to 13 kcal/mole of usable energy. How can this be? Why do you think a range of energies is given, rather than a precise number as for $\Delta G°$?

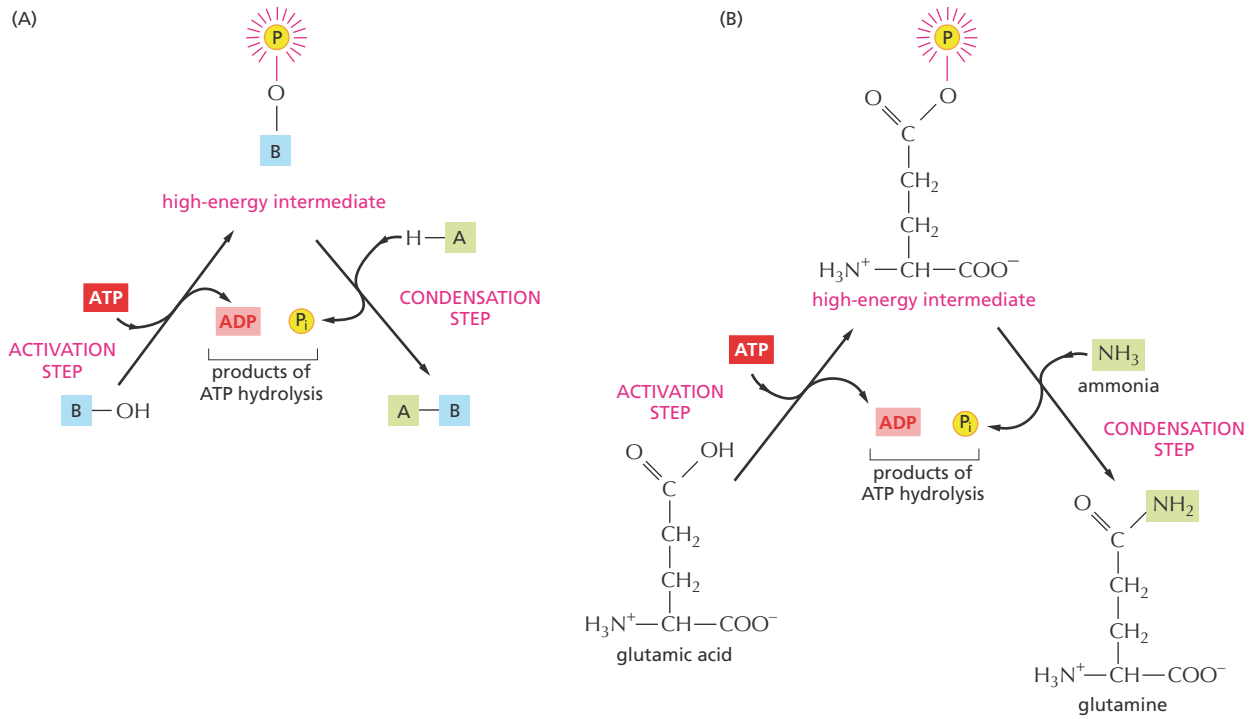


Figure 3–33 An energetically unfavorable biosynthetic reaction can be driven by ATP hydrolysis. (A) Schematic illustration of the formation of A–B in the condensation reaction described in the text. (B) The biosynthesis of the amino acid glutamine from glutamic acid. Glutamic acid is first converted to a high-energy phosphorylated intermediate (corresponding to the compound B–O–PO_3 described in the text), which then reacts with ammonia (corresponding to A–H) to form glutamine. In this example, both steps occur on the surface of the same enzyme, glutamine synthetase (not shown). For clarity, the glutamic acid side chain is shown in its uncharged form. ATP hydrolysis can drive this energetically unfavorable reaction because it yields more energy ($\Delta G°$ of –7.3 kcal/mole) than the energy required for the synthesis of glutamine from glutamic acid plus NH_3 ($\Delta G°$ of +3.4 kcal/mole).

The hydride ion carried by NADPH is given up readily in a subsequent oxidation–reduction reaction, because the ring can achieve a more stable arrangement of electrons without it. In this subsequent reaction, which regenerates $NADP^+$, the NADPH becomes oxidized and the substrate becomes reduced—thus completing the NADPH cycle. NADPH is efficient at donating its hydride ion to other molecules for the same reason that ATP readily transfers a phosphate: in both cases, the transfer is accompanied by a large negative free-energy change. One example of the use of NADPH in biosynthesis is shown in **Figure 3–35**.

NADPH and NADH Have Different Roles in Cells

NADPH and NADH differ in a single phosphate group, which is located far from the region involved in electron transfer in NADPH (Figure 3–34B). Although this phosphate group has no effect on the electron-transfer properties of NADPH compared with NADH, it is nonetheless crucial for their distinctive roles, as it gives NADPH a slightly different shape from NADH. This subtle difference in conformation makes it possible for the two carriers to bind as substrates to different sets of enzymes and thereby deliver electrons (in the form of hydride ions) to different target molecules.

Why should there be this division of labor? The answer lies in the need to regulate two sets of electron-transfer reactions independently. NADPH operates chiefly with enzymes that catalyze anabolic reactions, supplying the high-energy electrons needed to synthesize energy-rich biological molecules. NADH, by contrast, has a special role as an intermediate in

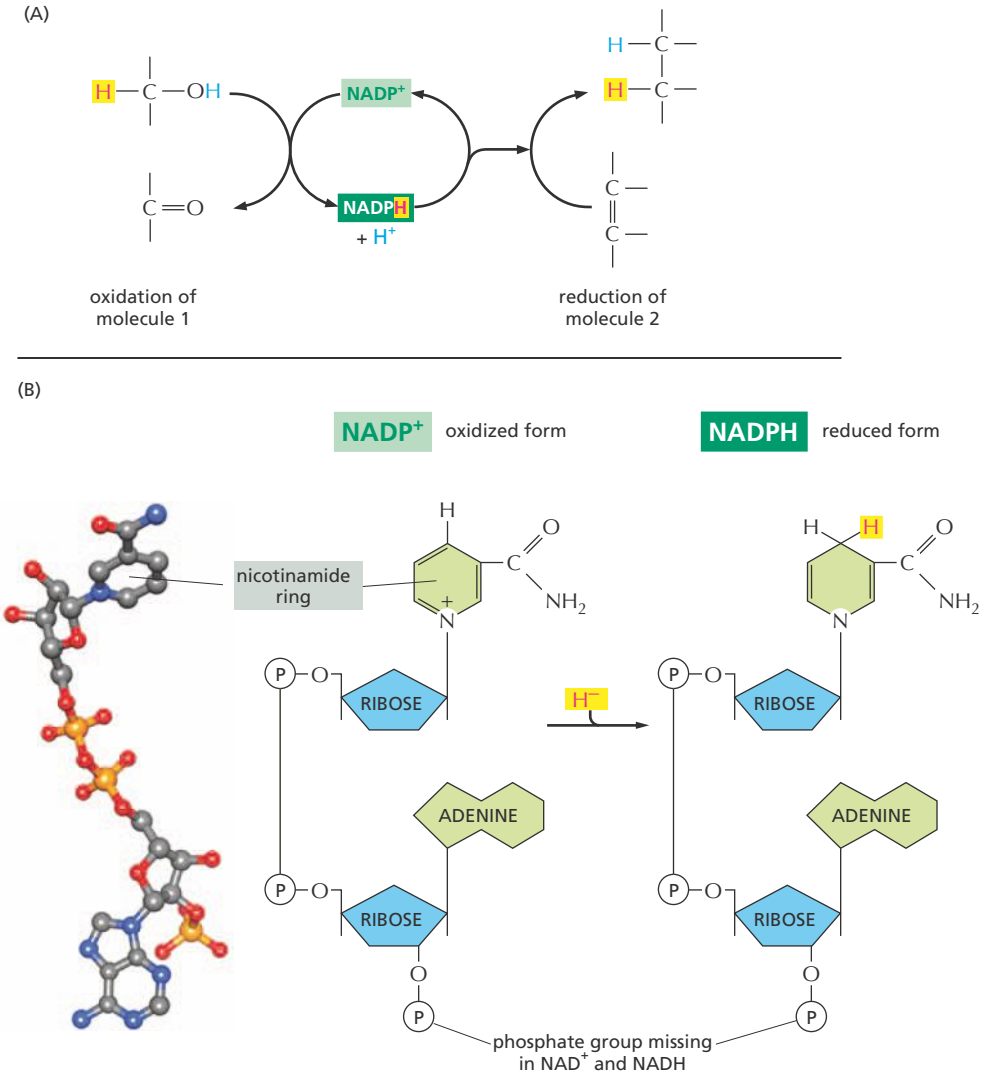


Figure 3–34 NADPH is an activated carrier of electrons. (A) NADPH is produced in reactions of the general type shown on the left, in which two hydrogen atoms are removed from a substrate. The oxidized form of the carrier molecule, $NADP^+$, receives one hydrogen atom plus an electron (a hydride ion), while the proton (H^+) from the other H atom is released into solution. Because NADPH holds its hydride ion in a high-energy linkage, the ion can easily be transferred to other molecules, as shown on the right. (B) The structure of $NADP^+$ and NADPH. On the left is a ball-and-stick model of NADP. The part of the $NADP^+$ molecule known as the nicotinamide ring accepts two electrons, together with a proton (the equivalent of a hydride ion, H^-), forming NADPH. NAD^+ and NADH are identical in structure to $NADP^+$ and NADPH, respectively, except that they lack the phosphate group, as indicated.

the catabolic system of reactions that generate ATP through the oxidation of food molecules, as we discuss in Chapter 13. The genesis of NADH from NAD^+ and that of NADPH from $NADP^+$ occurs by different pathways that are independently regulated, so that the cell can adjust the supply of electrons for these two contrasting purposes. Inside the cell, the ratio of NAD^+ to NADH is kept high, whereas the ratio of $NADP^+$ to NADPH is kept low. This arrangement provides plenty of NAD^+ to act as an oxidizing agent and plenty of NADPH to act as a reducing agent—as required for their special roles in catabolism and anabolism, respectively.

Cells Make Use of Many Other Activated Carriers

In addition to ATP (which transfers a phosphate) and NADPH and NADH (which transfer electrons and hydrogen), cells make use of other activated carriers that pick up and carry a chemical group in an easily transferred, high-energy linkage. *$FADH_2$*, like NADH and NADPH, carries hydrogen and high-energy electrons (see Figure 13–13B). But other important reactions involve the transfers of acetyl, methyl, carboxyl, and glucose groups from activated carriers for the purpose of biosynthesis (**Table 3–2**). Coenzyme A, for example, can carry an acetyl group in a readily transferable linkage. This activated carrier, called **acetyl CoA** (acetyl coenzyme A), is shown in **Figure 3–36**. It is used, for example, to add sequentially two-carbon units in the biosynthesis of the hydrocarbon tails of fatty acids.

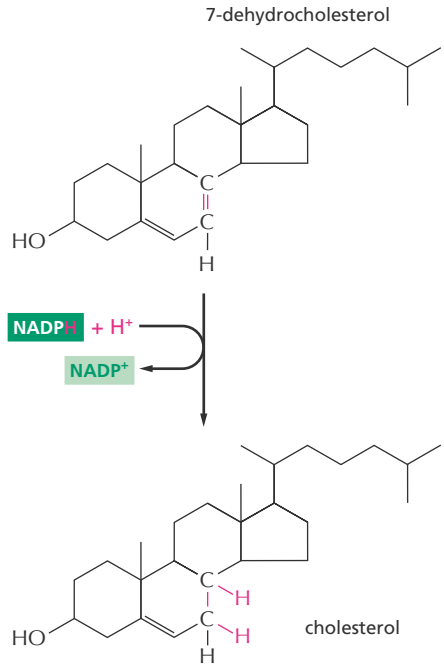


Figure 3–35 NADPH participates in the final stage of one of the biosynthetic routes leading to cholesterol. As in many other biosynthetic reactions, the reduction of the C=C bond is achieved by the transfer of a hydride ion from the activated carrier NADPH, plus a proton (H^+) from solution.

TABLE 3–2 SOME ACTIVATED CARRIERS WIDELY USED IN METABOLISM

Activated Carrier	Group Carried in High-Energy Linkage
ATP	phosphate
NADH, NADPH, $FADH_2$	electrons and hydrogens
Acetyl CoA	acetyl group
Carboxylated biotin	carboxyl group
S-adenosylmethionine	methyl group
Uridine diphosphate glucose	glucose

In acetyl CoA and the other activated carriers in Table 3–2, the transferable group makes up only a small part of the molecule. The rest consists of a large organic portion that serves as a convenient "handle," facilitating the recognition of the carrier molecule by specific enzymes. As with acetyl CoA, this handle portion very often contains a nucleotide. This curious fact may be a relic from an early stage of cell evolution. It is thought that the main catalysts for early life forms on Earth were RNA molecules (or their close relatives) and that proteins were a later evolutionary addition. It is therefore tempting to speculate that many of the activated carriers that we find today originated in an earlier RNA world, where their nucleotide portions would have been useful for binding these carriers to RNA-based catalysts, or *ribozymes* (discussed in Chapter 7).

Activated carriers are usually generated in reactions coupled to ATP hydrolysis, as shown for biotin in **Figure 3–37**. Therefore, the energy that enables their groups to be used for biosynthesis ultimately comes from the catabolic reactions that generate ATP. Similar processes occur in the synthesis of the very large macromolecules—the nucleic acids, proteins, and polysaccharides—which we discuss next.

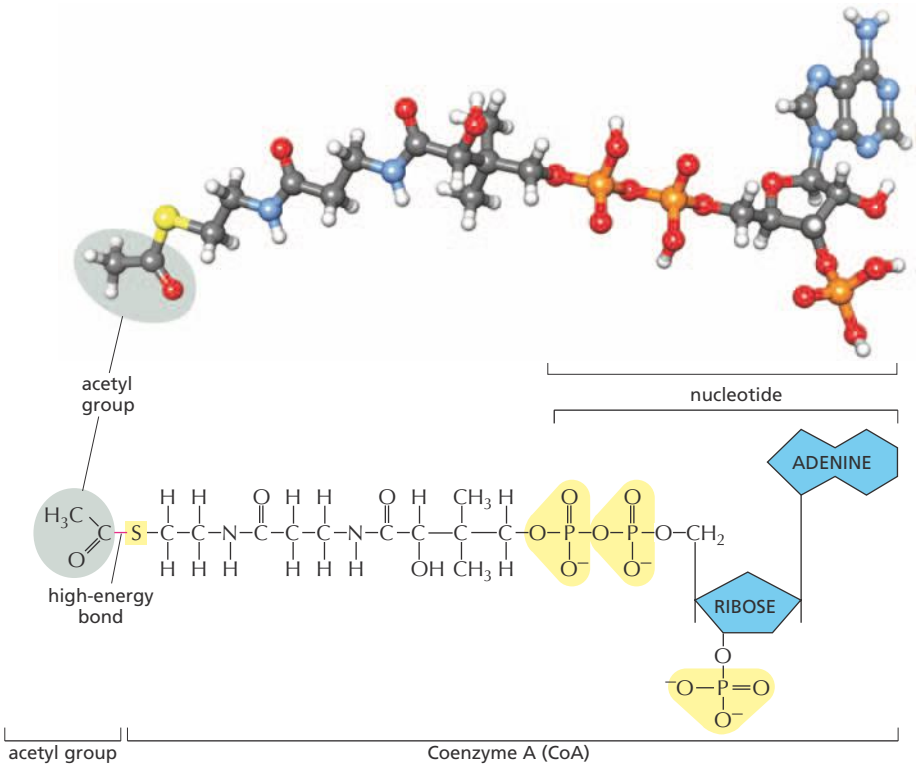


Figure 3–36 Acetyl coenzyme A (CoA) is another important activated carrier. A ball-and-stick model is shown above the structure of acetyl CoA. The sulfur atom (*yellow*) forms a thioester bond to acetate. Because the thioester bond is a high-energy linkage, it releases a large amount of free energy when it is hydrolyzed; thus the acetyl group carried by CoA can be readily transferred to other molecules.

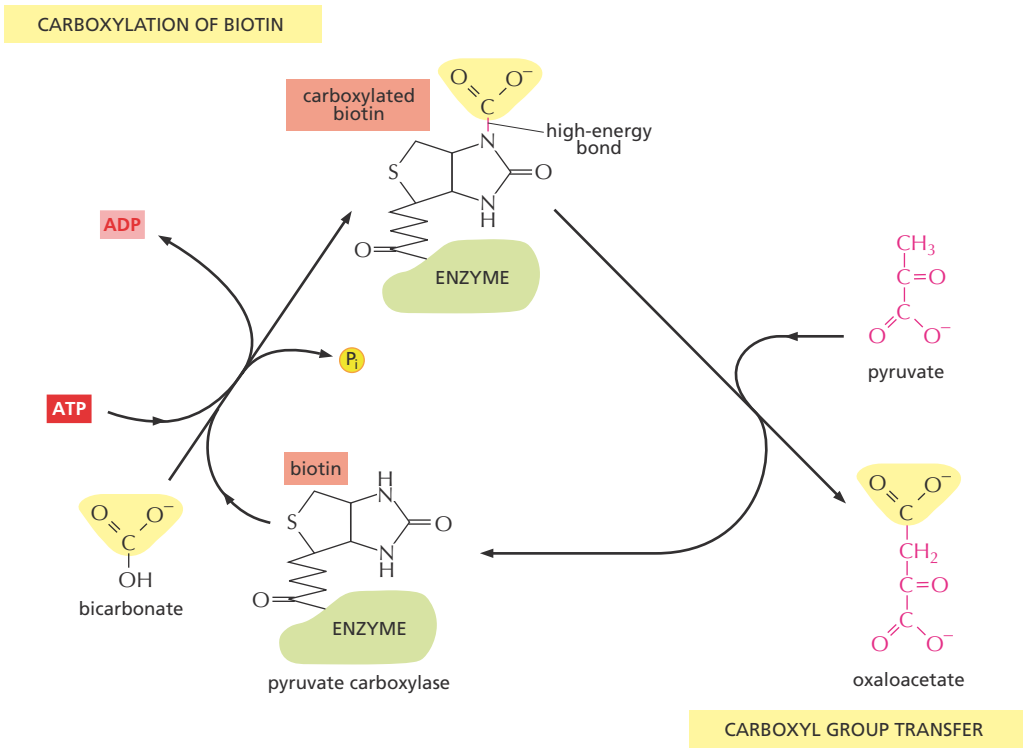


Figure 3–37 An activated carrier transfers a carboxyl group to a substrate. Biotin is a vitamin that is used by a number of enzymes, including *pyruvate carboxylase* shown here. Once it is carboxylated, biotin can transfer a carboxyl group to another molecule. Here, it transfers a carboxyl group to pyruvate, producing oxaloacetate, a molecule needed in the citric acid cycle (discussed in Chapter 13). Other enzymes use biotin to transfer carboxyl groups to other acceptor molecules. Note that the synthesis of carboxylated biotin requires energy derived from ATP hydrolysis—a general feature of many activated carriers.

The Synthesis of Biological Polymers Requires an Energy Input

The macromolecules of the cell constitute the vast majority of its dry mass—that is, the mass not due to water. These molecules are made from *subunits* (or monomers) that are linked together by bonds formed during an enzyme-catalyzed condensation reaction. The reverse reaction—the breakdown of polymers—occurs through enzyme-catalyzed hydrolysis reactions. These hydrolysis reactions are energetically favorable, whereas the corresponding biosynthetic reactions require an energy input and are more complex (**Figure 3–38**).

The nucleic acids (DNA and RNA), proteins, and polysaccharides are all polymers that are produced by the repeated addition of a subunit onto one end of a growing chain. The mode of synthesis of each of these macromolecules is outlined in **Figure 3–39**. As indicated, the condensation step in each case depends on energy provided by the hydrolysis of a nucleoside triphosphate. And yet, except for the nucleic acids, there are no phosphate groups left in the final product molecules. How, then, is the energy of ATP hydrolysis coupled to polymer synthesis?

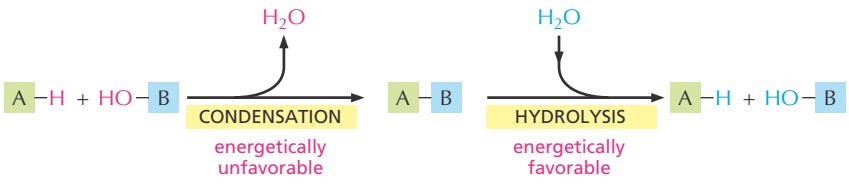


Figure 3–38 In cells, macromolecules are synthesized by condensation reactions and broken down by hydrolysis reactions. Condensation reactions are all energetically unfavorable, whereas hydrolysis reactions are all energetically favorable.

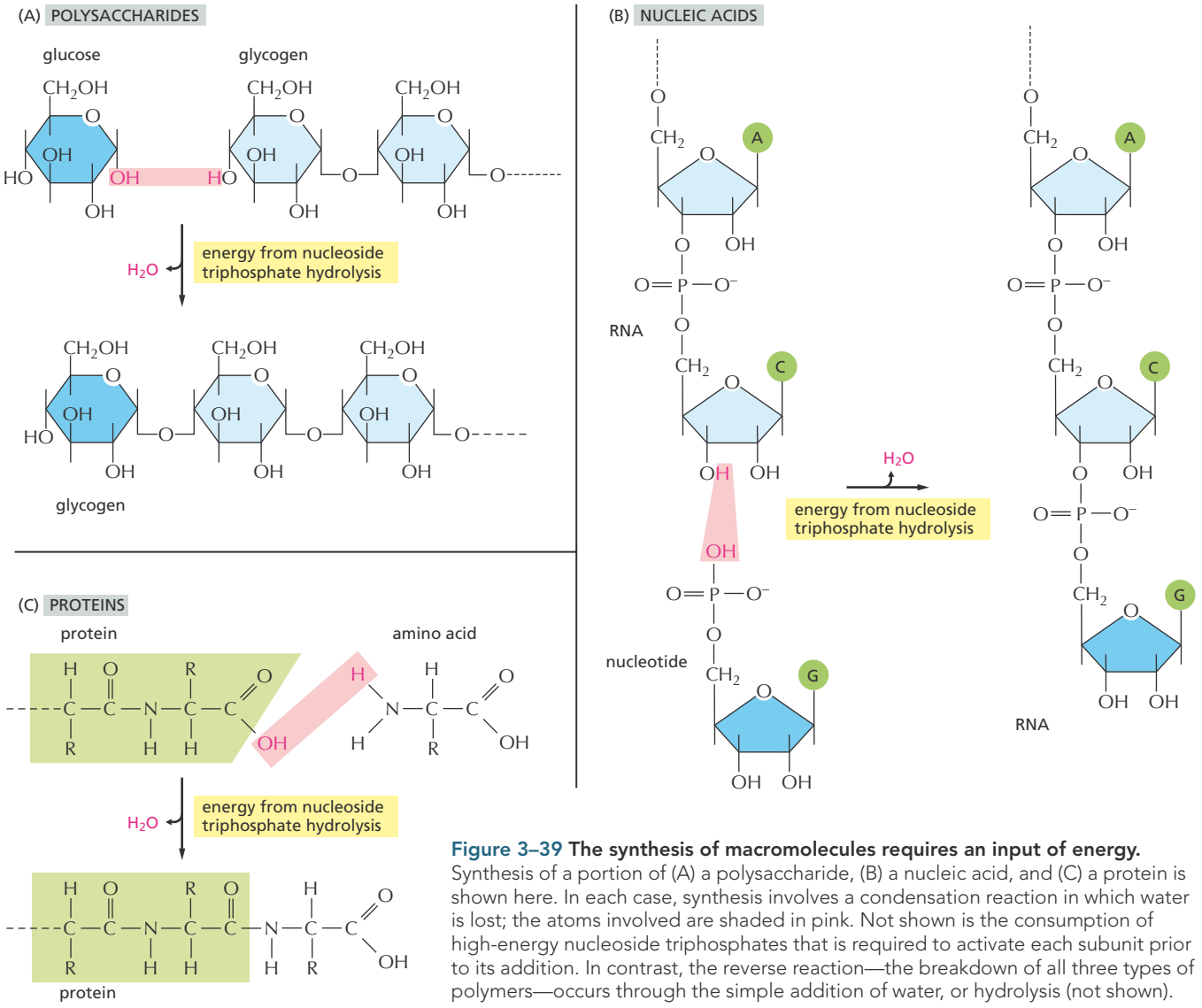


Figure 3–39 The synthesis of macromolecules requires an input of energy. Synthesis of a portion of (A) a polysaccharide, (B) a nucleic acid, and (C) a protein is shown here. In each case, synthesis involves a condensation reaction in which water is lost; the atoms involved are shaded in pink. Not shown is the consumption of high-energy nucleoside triphosphates that is required to activate each subunit prior to its addition. In contrast, the reverse reaction—the breakdown of all three types of polymers—occurs through the simple addition of water, or hydrolysis (not shown).

For each type of macromolecule, an enzyme-catalyzed pathway exists, which resembles that discussed previously for the synthesis of the amino acid glutamine (see Figure 3–33). The principle is exactly the same, in that the –OH group that will be removed in the condensation reaction is first activated by forming a high-energy linkage to a second molecule. The mechanisms used to link ATP hydrolysis to the synthesis of proteins and polysaccharides, however, are more complex than that used for glutamine synthesis. In the biosynthetic pathways leading to these macromolecules, a series of high-energy intermediates generates the final high-energy bond that is broken during the condensation step (as discussed in Chapter 7 for protein synthesis).

There are limits to what each activated carrier can do in driving biosynthesis. For example, the ΔG for the hydrolysis of ATP to ADP and inorganic phosphate (P_i) depends on the concentrations of all of the reactants, and under the usual conditions in a cell, is between –11 and –13 kcal/mole. In principle, this hydrolysis reaction can be used to drive an unfavorable reaction with a ΔG of, perhaps, +10 kcal/mole, provided that a suitable reaction path is available. For some biosynthetic reactions, however, even –13 kcal/mole may be insufficient. In these cases, the path of ATP

QUESTION 3–9

Which of the following reactions will occur only if coupled to a second, energetically favorable reaction?

A. glucose + O_2 → CO_2 + H_2O
B. CO_2 + H_2O → glucose + O_2
C. nucleoside triphosphates → DNA
D. nucleotide bases → nucleoside triphosphates
E. ADP + P_i → ATP

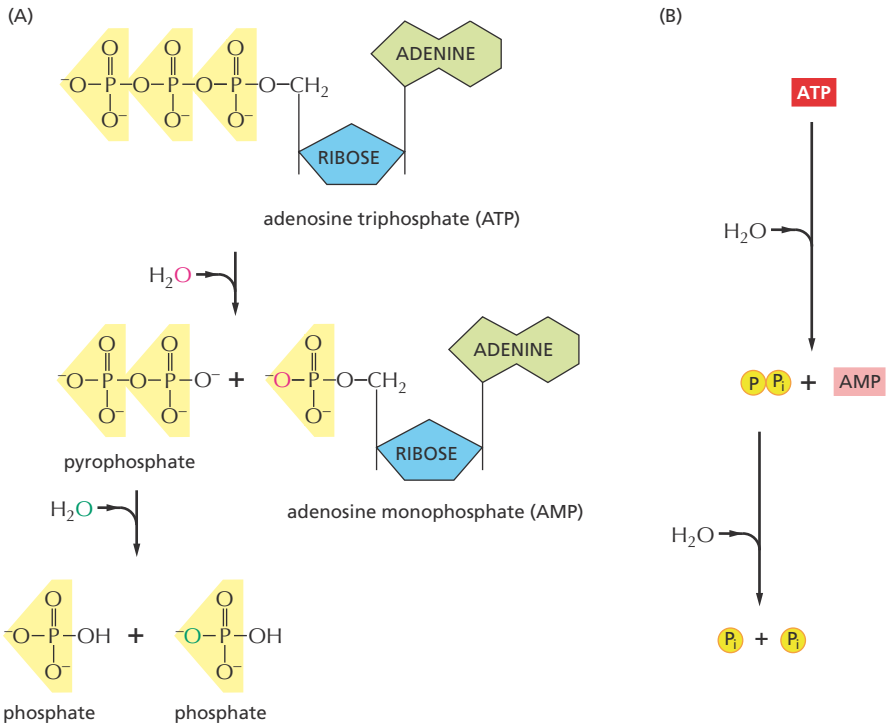


Figure 3–40 In an alternative route for the hydrolysis of ATP, pyrophosphate is first formed and then hydrolyzed in solution. This route releases about twice as much free energy as the reaction shown earlier in Figure 3–31. (A) In each of the two successive hydrolysis reactions, an oxygen atom from the participating water molecule is retained in the products, whereas the hydrogen atoms from water form free hydrogen ions, H^+. (B) The overall reaction shown in summary form.

hydrolysis can be altered so that it initially produces AMP and pyrophosphate (PP_i), which is itself then hydrolyzed in solution in a subsequent step (**Figure 3–40**). The whole process makes available a total ΔG of about –26 kcal/mole. The biosynthetic reaction involved in the synthesis of nucleic acids (polynucleotides) is driven in this way (**Figure 3–41**).

ATP will make many appearances throughout the book as a molecule that powers reactions in the cell. And in Chapters 13 and 14, we discuss how the cell uses the energy from food to generate ATP. In the next chapter, we learn more about the proteins that make such reactions possible.

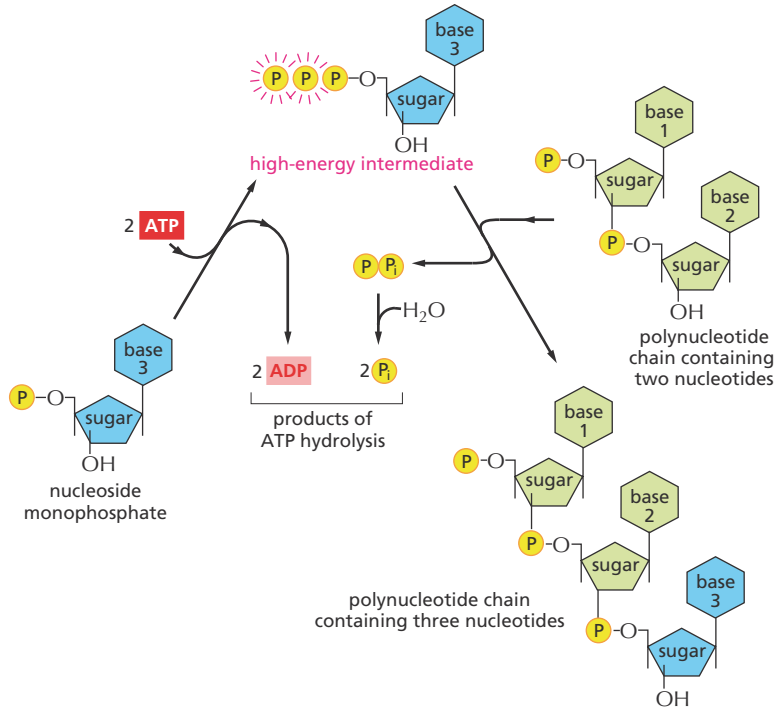


Figure 3–41 Synthesis of a polynucleotide, RNA or DNA, is a multistep process driven by ATP hydrolysis. In the first step, a nucleoside monophosphate is activated by the sequential transfer of the terminal phosphate groups from two ATP molecules. The high-energy intermediate formed—a nucleoside triphosphate—exists free in solution until it reacts with the growing end of an RNA or a DNA chain with release of pyrophosphate. Hydrolysis of the pyrophosphate to inorganic phosphate is highly favorable and helps to drive the overall reaction in the direction of polynucleotide synthesis.

ESSENTIAL CONCEPTS

- Living organisms are able to exist because of a continual input of energy. Part of this energy is used to carry out essential reactions that support cell metabolism, growth, movement, and reproduction; the remainder is lost in the form of heat.
- The ultimate source of energy for most living organisms is the sun. Plants, algae, and photosynthetic bacteria use solar energy to produce organic molecules from carbon dioxide. Animals obtain food by eating plants or by eating animals that feed on plants.
- Each of the many hundreds of chemical reactions that occur in a cell is specifically catalyzed by an enzyme. Large numbers of different enzymes work in sequence to form chains of reactions, called metabolic pathways, each performing a different function in the cell.
- Catabolic reactions release energy by breaking down organic molecules, including foods, through oxidative pathways. Anabolic reactions generate the many complex organic molecules needed by the cell, and they require an energy input. In animal cells, both the building blocks and the energy required for the anabolic reactions are obtained through catabolic reactions.
- Enzymes catalyze reactions by binding to particular substrate molecules in a way that lowers the activation energy required for making and breaking specific covalent bonds.
- The rate at which an enzyme catalyzes a reaction depends on how rapidly it finds its substrates and how quickly the product forms and then diffuses away. These rates vary widely from one enzyme to another.
- The only chemical reactions possible are those that increase the total amount of disorder in the universe. The free-energy change for a reaction, ΔG, measures this disorder, and it must be less than zero for a reaction to proceed spontaneously.
- The ΔG for a chemical reaction depends on the concentrations of the reacting molecules, and it may be calculated from these concentrations if the equilibrium constant (K) of the reaction (or the standard free-energy change, ΔG°, for the reactants) is known.
- Equilibrium constants govern all of the associations (and dissociations) that occur between macromolecules and small molecules in the cell. The larger the binding energy between two molecules, the larger the equilibrium constant and the more likely that these molecules will be found bound to each other.
- By creating a reaction pathway that couples an energetically favorable reaction to an energetically unfavorable one, enzymes can make otherwise impossible chemical transformations occur.
- A small set of activated carriers, particularly ATP, NADH, and NADPH, plays a central part in these coupled reactions in cells. ATP carries high-energy phosphate groups, whereas NADH and NADPH carry high-energy electrons.
- Food molecules provide the carbon skeletons for the formation of macromolecules. The covalent bonds of these larger molecules are produced by condensation reactions that are coupled to energetically favorable bond changes in activated carriers such as ATP and NADPH.

KEY TERMS

acetyl CoA
activated carrier
activation energy
ADP, ATP
anabolism
biosynthesis
catabolism
catalysis
catalyst
condensation reaction
coupled reaction
diffusion
entropy
enzyme
equilibrium
equilibrium constant, K
free energy, G
free-energy change, ΔG
hydrolysis
K_M
metabolism
Michaelis constant (K_M)
NAD^+, NADH
$NADP^+$, NADPH
oxidation
photosynthesis
reduction
respiration
standard free-energy change, $\Delta G°$
substrate
turnover number
V_{max}

QUESTIONS

QUESTION 3–10

Which of the following statements are correct? Explain your answers.

A. Some enzyme-catalyzed reactions cease completely if their enzyme is absent.

B. High-energy electrons (such as those found in the activated carriers NADH and NADPH) move faster around the atomic nucleus.

C. Hydrolysis of ATP to AMP can provide about twice as much energy as hydrolysis of ATP to ADP.

D. A partially oxidized carbon atom has a somewhat smaller diameter than a more reduced one.

E. Some activated carrier molecules can transfer both energy and a chemical group to a second molecule.

F. The rule that oxidations release energy, whereas reductions require energy input, applies to all chemical reactions, not just those that occur in living cells.

G. Cold-blooded animals have an energetic disadvantage because they release less heat to the environment than warm-blooded animals do. This slows their ability to make ordered macromolecules.

H. Linking the reaction X → Y to a second, energetically favorable reaction Y → Z will shift the equilibrium constant of the first reaction.

QUESTION 3–11

Consider a transition of X → Y. Assume that the only difference between X and Y is the presence of three hydrogen bonds in Y that are absent in X. What is the ratio of X to Y when the reaction is in equilibrium? Approximate your answer by using Table 3–1 (p. 98), with 1 kcal/mole as the energy of each hydrogen bond. If Y instead has six hydrogen bonds that distinguish it from X, how would that change the ratio?

QUESTION 3–12

Protein A binds to protein B to form a complex, AB. At equilibrium in a cell the concentrations of A, B, and AB are all at 1 μM.

A. Referring to Figure 3–19, calculate the equilibrium constant for the reaction A + B ⇌ AB.

B. What would the equilibrium constant be if A, B, and AB were each present in equilibrium at the much lower concentrations of 1 nM each?

C. How many extra hydrogen bonds would be needed to hold A and B together at this lower concentration so that a similar proportion of the molecules are found in the AB complex? (Remember that each hydrogen bond contributes about 1 kcal/mole.)

QUESTION 3–13

Discuss the following statement: "Whether the ΔG for a reaction is larger, smaller, or the same as $\Delta G°$ depends on the concentration of the compounds that participate in the reaction."

QUESTION 3–14

A. How many ATP molecules could maximally be generated from one molecule of glucose, if the complete oxidation of 1 mole of glucose to CO_2 and H_2O yields 686 kcal of free energy and the useful chemical energy available in the high-energy phosphate bond of 1 mole of ATP is 12 kcal?

B. As we will see in Chapter 14 (Table 14–1), respiration produces 30 moles of ATP from 1 mole of glucose. Compare this number with your answer in part (A). What is the overall efficiency of ATP production from glucose?

C. If the cells of your body oxidize 1 mole of glucose, by how much would the temperature of your body (assume that your body consists of 75 kg of water) increase if the heat were not dissipated into the environment? [Recall that a kilocalorie (kcal) is defined as that amount of energy that heats 1 kg of water by 1°C.]

D. What would the consequences be if the cells of your body could convert the energy in food substances with only 20% efficiency? Would your body—as it is presently constructed—work just fine, overheat, or freeze?

E. A resting human hydrolyzes about 40 kg of ATP every 24 hours. The oxidation of how much glucose would produce this amount of energy? (Hint: Look up the structure of ATP in Figure 2–24 to calculate its molecular weight; the atomic weights of H, C, N, O, and P are 1, 12, 14, 16, and 31, respectively.)

QUESTION 3–15

A prominent scientist claims to have isolated mutant cells that can convert 1 molecule of glucose into 57 molecules of ATP. Should this discovery be celebrated, or do you suppose that something might be wrong with it? Explain your answer.

QUESTION 3–16

In a simple reaction A $\rightleftharpoons$ A*, a molecule is interconvertible between two forms that differ in standard free energy $G°$ by 4.3 kcal/mole, with A* having the higher $G°$.

A. Use Table 3–1 (p. 98) to find how many more molecules will be in state A* compared with state A at equilibrium.

B. If an enzyme lowered the activation energy of the reaction by 2.8 kcal/mole, how would the ratio of A to A* change?

QUESTION 3–17

A reaction in a single-step biosynthetic pathway that converts a metabolite into a particularly vicious poison (metabolite $\rightleftharpoons$ poison) in a mushroom is energetically highly unfavorable. The reaction is normally driven by ATP hydrolysis. Assume that a mutation in the enzyme that catalyzes the reaction prevents it from utilizing ATP, but still allows it to catalyze the reaction.

A. Do you suppose it might be safe for you to eat a mushroom that bears this mutation? Base your answer on an estimation of how much less poison the mutant mushroom would produce, assuming the reaction is in equilibrium and most of the energy stored in ATP is used to drive the unfavorable reaction in nonmutant mushrooms.

B. Would your answer be different for another mutant mushroom whose enzyme couples the reaction to ATP hydrolysis but works 100 times more slowly?

QUESTION 3–18

Consider the effects of two enzymes, A and B. Enzyme A catalyzes the reaction

$$ATP + GDP \rightleftharpoons ADP + GTP$$

and enzyme B catalyzes the reaction

$$NADH + NADP^+ \rightleftharpoons NAD^+ + NADPH$$

Discuss whether the enzymes would be beneficial or detrimental to cells.

QUESTION 3–19

Discuss the following statement: "Enzymes and heat are alike in that both can speed up reactions that—although thermodynamically feasible—do not occur at an appreciable rate because they require a high activation energy. Diseases that seem to benefit from the careful application of heat—in the form of hot chicken soup, for example—are therefore likely to be due to the insufficient function of an enzyme."

QUESTION 3–20

The curve shown in Figure 3–24 is described by the Michaelis–Menten equation:

$$\text{rate } (v) = V_{max}\,[S]/([S] + K_M)$$

Can you convince yourself that the features qualitatively described in the text are accurately represented by this equation? In particular, how can the equation be simplified when the substrate concentration [S] is in one of the following ranges: (A) [S] is much smaller than the K_M, (B) [S] equals the K_M, and (C) [S] is much larger than the K_M?

QUESTION 3–21

The rate of a simple enzyme reaction is given by the standard Michaelis–Menten equation:

$$\text{rate} = V_{max}\,[S]/([S] + K_M)$$

If the V_{max} of an enzyme is 100 μmole/sec and the K_M is 1 mM, at what substrate concentration is the rate 50 μmole/sec? Plot a graph of rate versus substrate (S) concentration for [S] = 0 to 10 mM. Convert this to a plot of 1/rate versus 1/[S]. Why is the latter plot a straight line?

QUESTION 3–22

Select the correct options in the following and explain your choices. If [S] is much smaller than K_M, the active site of the enzyme is mostly occupied/unoccupied. If [S] is very much greater than K_M, the reaction rate is limited by the enzyme/substrate concentration.

QUESTION 3–23

A. The reaction rates of the reaction S → P catalyzed by enzyme E were determined under conditions such that only very little product was formed. The following data were measured:

Substrate concentration (μM)	Reaction rate (μmole/min)
0.08	0.15
0.12	0.21
0.54	0.7
1.23	1.1
1.82	1.3
2.72	1.5
4.94	1.7
10.00	1.8

Plot the above data as a graph. Use this graph to estimate the K_M and the V_{max} for this enzyme.

B. Recall from the How We Know essay (pp. 104–106) that to determine these values more precisely, a trick is generally used in which the Michaelis–Menten equation is transformed so that it is possible to plot the data as a straight line. A simple rearrangement yields

$$1/\text{rate} = (K_M/V_{max})\,(1/[S]) + 1/V_{max}$$

which is an equation of the form $y = ax + b$. Calculate 1/rate and 1/[S] for the data given in part (A) and then plot 1/rate versus 1/[S] as a new graph. Determine K_M and V_{max} from the intercept of the line with the axis, where 1/[S] = 0, combined with the slope of the line. Do your results agree with the estimates made from the first graph of the raw data?

C. It is stated in part (A) that only very little product was formed under the reaction conditions. Why is this important?

D. Assume the enzyme is regulated such that upon phosphorylation its K_M increases by a factor of 3 without changing its V_{max}. Is this an activation or inhibition? Plot the data you would expect for the phosphorylated enzyme in both the graph for (A) and the graph for (B).

Page left intentionally blank

CHAPTER **FOUR** 4

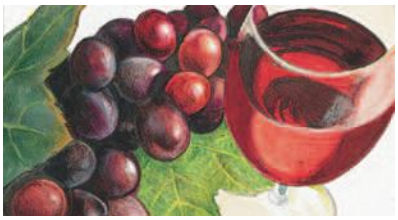
Protein Structure and Function

When we look at a cell in a microscope or analyze its electrical or biochemical activity, we are, in essence, observing the handiwork of proteins. **Proteins** are the main building blocks from which cells are assembled, and they constitute most of the cell's dry mass. In addition to providing the cell with shape and structure, proteins also execute nearly all its myriad functions. *Enzymes* promote intracellular chemical reactions by providing intricate molecular surfaces, contoured with particular bumps and crevices that can cradle or exclude specific molecules. Proteins embedded in the plasma membrane form the channels and pumps that control the passage of nutrients and other small molecules into and out of the cell. Other proteins carry messages from one cell to another, or act as signal integrators that relay information from the plasma membrane to the nucleus of individual cells. Some proteins act as motors that propel organelles through the cytoplasm, and others function as components of tiny molecular machines with precisely calibrated moving parts. Specialized proteins also act as antibodies, toxins, hormones, antifreeze molecules, elastic fibers, or luminescence generators. Before we can hope to understand how genes work, how muscles contract, how nerves conduct electricity, how embryos develop, or how our bodies function, we must understand proteins.

The multiplicity of functions carried out by proteins (**Panel 4–1**, p. 122) arises from the huge number of different shapes they adopt. We therefore begin our description of these remarkable macromolecules by discussing their three-dimensional structures and the properties that these structures confer. We next look at how proteins work: how enzymes catalyze chemical reactions, how some proteins act as molecular switches, and how others generate orderly movement. We then examine how cells

ENZYMES

function: Catalyze covalent bond breakage or formation.



examples: Living cells contain thousands of different enzymes, each of which catalyzes (speeds up) one particular reaction. Examples include: *tryptophan synthetase*—makes the amino acid tryptophan; *pepsin*—degrades dietary proteins in the stomach; *ribulose bisphosphate carboxylase*—helps convert carbon dioxide into sugars in plants; *DNA polymerase*—copies DNA; *protein kinase*—adds a phosphate group to a protein molecule.

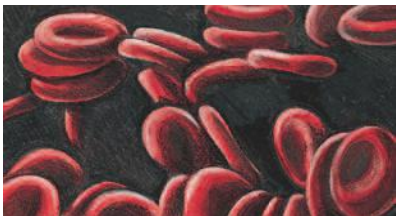
STRUCTURAL PROTEINS

function: Provide mechanical support to cells and tissues.

examples: Outside cells, *collagen* and *elastin* are common constituents of extracellular matrix and form fibers in tendons and ligaments. Inside cells, *tubulin* forms long, stiff microtubules, and *actin* forms filaments that underlie and support the plasma membrane; *keratin* forms fibers that reinforce epithelial cells and is the major protein in hair and horn.

TRANSPORT PROTEINS

function: Carry small molecules or ions.



examples: In the bloodstream, *serum albumin* carries lipids, *hemoglobin* carries oxygen, and *transferrin* carries iron. Many proteins embedded in cell membranes transport ions or small molecules across the membrane. For example, the bacterial protein *bacteriorhodopsin* is a light-activated proton pump that transports H^+ ions out of the cell; *glucose carriers* shuttle glucose into and out of cells; and a *Ca^{2+} pump* clears Ca^{2+} from a muscle cell's cytosol after the ions have triggered a contraction.

MOTOR PROTEINS

function: Generate movement in cells and tissues.

examples: *Myosin* in skeletal muscle cells provides the motive force for humans to move; *kinesin* interacts with microtubules to move organelles around the cell; *dynein* enables eukaryotic cilia and flagella to beat.

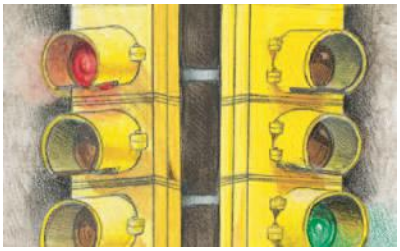
STORAGE PROTEINS

function: Store amino acids or ions.

examples: Iron is stored in the liver by binding to the small protein *ferritin*; *ovalbumin* in egg white is used as a source of amino acids for the developing bird embryo; *casein* in milk is a source of amino acids for baby mammals.

SIGNAL PROTEINS

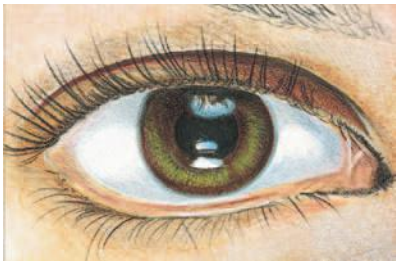
function: Carry extracellular signals from cell to cell.



examples: Many of the hormones and growth factors that coordinate physiological functions in animals are proteins; *insulin*, for example, is a small protein that controls glucose levels in the blood; *netrin* attracts growing nerve cell axons to specific locations in the developing spinal cord; *nerve growth factor* (*NGF*) stimulates some types of nerve cells to grow axons; *epidermal growth factor* (*EGF*) stimulates the growth and division of epithelial cells.

RECEPTOR PROTEINS

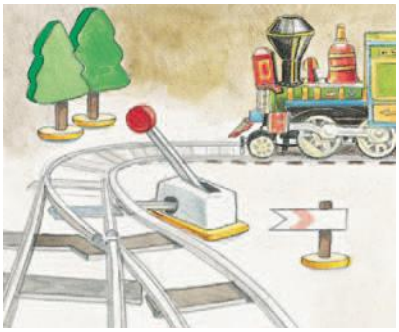
function: Detect signals and transmit them to the cell's response machinery.



examples: *Rhodopsin* in the retina detects light; the *acetylcholine receptor* in the membrane of a muscle cell is activated by acetylcholine released from a nerve ending; the *insulin receptor* allows a cell to respond to the hormone insulin by taking up glucose; the *adrenergic receptor* on heart muscle increases the rate of the heartbeat when it binds to adrenaline.

GENE REGULATORY PROTEINS

function: Bind to DNA to switch genes on or off.



examples: The *lactose repressor* in bacteria silences the genes for the enzymes that degrade the sugar lactose; many different *homeodomain proteins* act as genetic switches to control development in multicellular organisms, including humans.

SPECIAL-PURPOSE PROTEINS

function: Highly variable.

examples: Organisms make many proteins with highly specialized properties. These molecules illustrate the amazing range of functions that proteins can perform. The *antifreeze proteins* of Arctic and Antarctic fishes protect their blood against freezing; *green fluorescent protein* from jellyfish emits a green light; *monellin*, a protein found in an African plant, has an intensely sweet taste; mussels and other marine organisms secrete *glue proteins* that attach them firmly to rocks, even when immersed in seawater.

control the activity and location of the proteins they contain. Finally, we present a brief description of the techniques that biologists use to work with proteins, including methods for purifying them—from tissues or cultured cells—and for determining their structures.

THE SHAPE AND STRUCTURE OF PROTEINS

From a chemical point of view, proteins are by far the most structurally complex and functionally sophisticated molecules known. This is perhaps not surprising, considering that the structure and activity of each protein has been developed and fine-tuned over billions of years of evolution. We start by considering how the position of each amino acid in the long string of amino acids that forms a protein determines its three-dimensional shape, which is stabilized by noncovalent interactions between different parts of the molecule. Understanding the structure of a protein at the atomic level allows us to see how the precise shape of the protein determines its function.

The Shape of a Protein Is Specified by Its Amino Acid Sequence

Proteins, as you may recall from Chapter 2, are assembled mainly from a set of 20 different amino acids, each with different chemical properties. A protein molecule is made from a long chain of these amino acids, held together by covalent **peptide bonds** (Figure 4–1). Proteins are therefore referred to as **polypeptides**, and their amino acid chains are called **polypeptide chains**. In each type of protein, the amino acids are present in a unique order, called the **amino acid sequence**, which is exactly the same from one molecule of that protein to the next. One molecule of human insulin, for example, has the same amino acid sequence as every other molecule of human insulin. Many thousands of different proteins have been identified, each with its own distinct amino acid sequence.

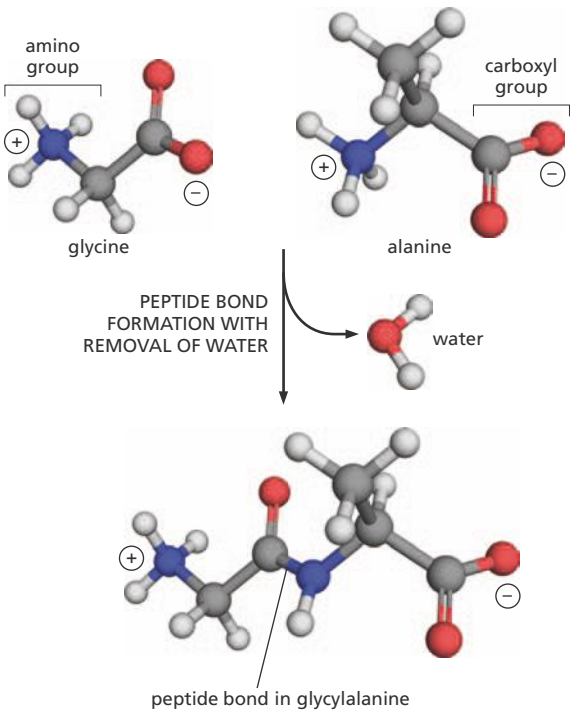


Figure 4–1 Amino acids are linked together by peptide bonds. A covalent peptide bond forms when the carbon atom of the carboxyl group of one amino acid (such as glycine) shares electrons with the nitrogen atom (*blue*) from the amino group of a second amino acid (such as alanine). Because a molecule of water is eliminated, peptide bond formation is classified as a condensation reaction (see Figure 2–29). In this diagram, carbon atoms are *gray*, nitrogen *blue*, oxygen *red*, and hydrogen *white*.

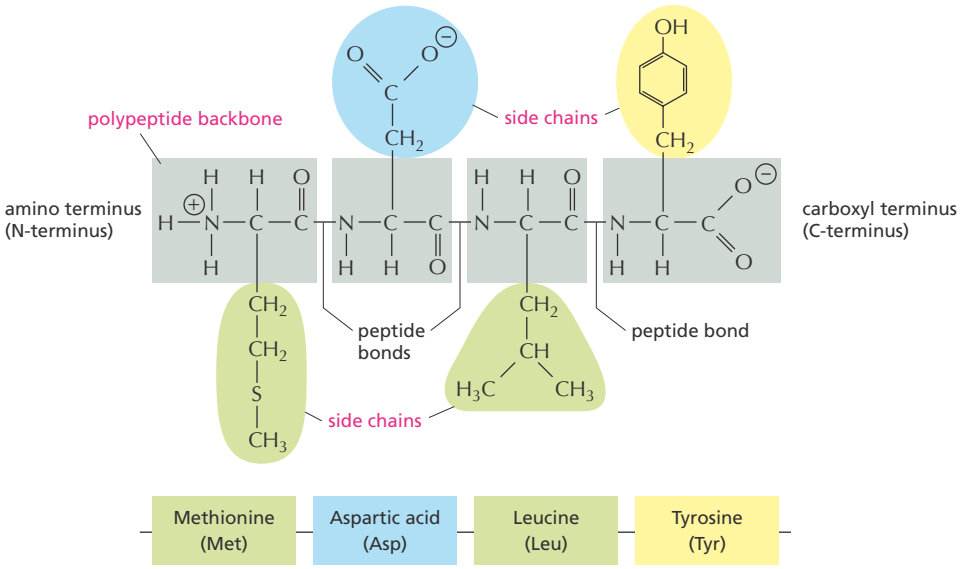


Figure 4–2 A protein is made of amino acids linked together into a polypeptide chain. The amino acids are linked by peptide bonds (see Figure 4–1) to form a polypeptide backbone of repeating structure (*gray boxes*), from which the side chain of each amino acid projects. The character and sequence of the chemically distinct side chains—for example, nonpolar (*green*), polar uncharged (*yellow*), and negative (*blue*) side chains—give each protein its distinct, individual properties. A small polypeptide of just four amino acids is shown here. Proteins are typically made up of chains of several hundred amino acids, whose sequence is always presented starting with the N-terminus reading from left to right.

Each polypeptide chain consists of a backbone that is adorned with a variety of chemical side chains. This **polypeptide backbone** is formed from a repeating sequence of the core atoms (–N–C–C–) found in every amino acid (see Figure 4–1). Because the two ends of each amino acid are chemically different—one sports an amino group (NH_3^+, also written NH_2) and the other a carboxyl group (COO^-, also written COOH)—each polypeptide chain has a directionality: the end carrying the amino group is called the amino terminus, or **N-terminus**, and the end carrying the free carboxyl group is the carboxyl terminus, or **C-terminus**.

Projecting from the polypeptide backbone are the amino acid **side chains**—the part of the amino acid that is not involved in forming peptide bonds (**Figure 4–2**). The side chains give each amino acid its unique properties: some are nonpolar and hydrophobic ("water-fearing"), some are negatively or positively charged, some can be chemically reactive, and so on. The atomic formula for each of the 20 amino acids in proteins is presented in Panel 2–5 (pp. 74–75), and a brief list of the 20 common amino acids, with their abbreviations, is provided in **Figure 4–3**.

AMINO ACID			SIDE CHAIN
Aspartic acid	Asp	D	negatively charged
Glutamic acid	Glu	E	negatively charged
Arginine	Arg	R	positively charged
Lysine	Lys	K	positively charged
Histidine	His	H	positively charged
Asparagine	Asn	N	uncharged polar
Glutamine	Gln	Q	uncharged polar
Serine	Ser	S	uncharged polar
Threonine	Thr	T	uncharged polar
Tyrosine	Tyr	Y	uncharged polar

POLAR AMINO ACIDS

AMINO ACID			SIDE CHAIN
Alanine	Ala	A	nonpolar
Glycine	Gly	G	nonpolar
Valine	Val	V	nonpolar
Leucine	Leu	L	nonpolar
Isoleucine	Ile	I	nonpolar
Proline	Pro	P	nonpolar
Phenylalanine	Phe	F	nonpolar
Methionine	Met	M	nonpolar
Tryptophan	Trp	W	nonpolar
Cysteine	Cys	C	nonpolar

NONPOLAR AMINO ACIDS

Figure 4–3 Twenty different amino acids are commonly found in proteins. Both three-letter and one-letter abbreviations are given, as well as the character of the side chain. There are equal numbers of polar (hydrophilic) and nonpolar (hydrophobic) side chains, and half of the polar side chains carry a positive or negative charge.

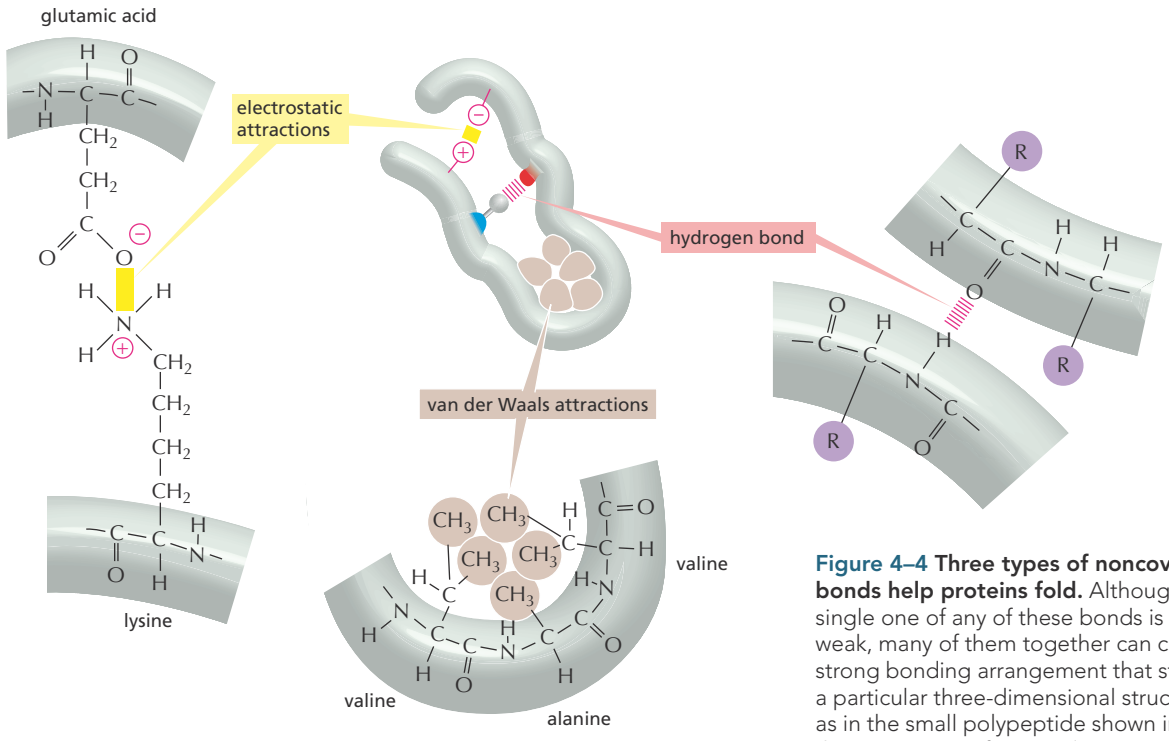


Figure 4–4 Three types of noncovalent bonds help proteins fold. Although a single one of any of these bonds is quite weak, many of them together can create a strong bonding arrangement that stabilizes a particular three-dimensional structure, as in the small polypeptide shown in the *center*. R is often used as a general designation for an amino acid side chain. Protein folding is also aided by hydrophobic forces, as shown in Figure 4–5.

Long polypeptide chains are very flexible, as many of the peptide bonds that link the carbon atoms in the polypeptide backbone allow free rotation of the atoms they join. Thus, proteins can in principle fold in an enormous number of ways. The shape of each of these folded chains, however, is constrained by many sets of weak noncovalent bonds that form within proteins. These bonds involve atoms in the polypeptide backbone, as well as atoms in the amino acid side chains. The noncovalent bonds that help proteins fold up and maintain their shape include hydrogen bonds, electrostatic attractions, and van der Waals attractions, which are described in Chapter 2 (see Panel 2–7, pp. 78–79). Because a noncovalent bond is much weaker than a covalent bond, it takes many noncovalent bonds to hold two regions of a polypeptide chain tightly together. The stability of each folded shape is largely influenced by the combined strength of large numbers of noncovalent bonds (**Figure 4–4**).

A fourth weak force, *hydrophobic interaction*, also has a central role in determining the shape of a protein. In an aqueous environment, hydrophobic molecules, including the nonpolar side chains of particular amino acids, tend to be forced together to minimize their disruptive effect on the hydrogen-bonded network of the surrounding water molecules (see Panel 2–2, pp. 68–69). Therefore, an important factor governing the folding of any protein is the distribution of its polar and nonpolar amino acids. The nonpolar (hydrophobic) side chains—which belong to amino acids such as phenylalanine, leucine, valine, and tryptophan (see Figure 4–3)—tend to cluster in the interior of the folded protein (just as hydrophobic oil droplets coalesce to form one large drop). Tucked away inside the folded protein, hydrophobic side chains can avoid contact with the aqueous cytosol that surrounds them inside a cell. In contrast, polar side chains—such as those belonging to arginine, glutamine, and histidine—tend to arrange themselves near the outside of the folded protein, where they can form hydrogen bonds with water and with other polar molecules (**Figure 4–5**). When polar amino acids are buried within the protein, they are usually hydrogen-bonded to other polar amino acids or to the polypeptide backbone (**Figure 4–6**).

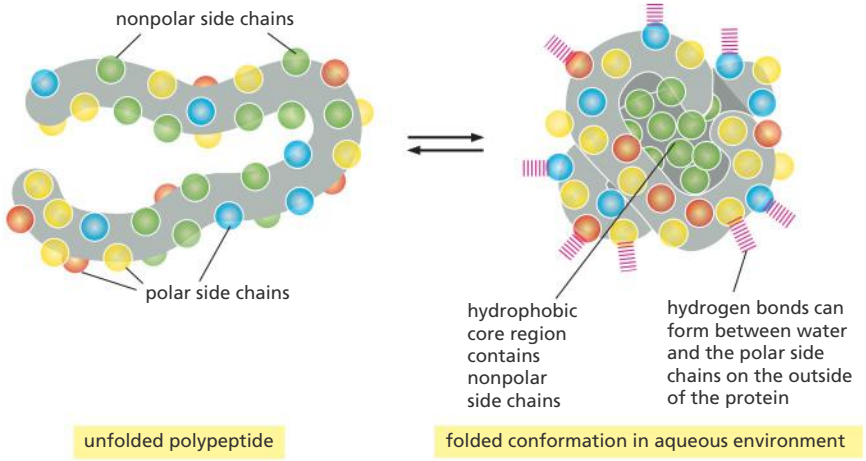


Figure 4–5 Hydrophobic forces help proteins fold into compact conformations. Polar amino acid side chains tend to be displayed on the outside of the folded protein, where they can interact with water; the nonpolar amino acid side chains are buried on the inside to form a highly packed hydrophobic core of atoms that are hidden from water.

Proteins Fold into a Conformation of Lowest Energy

Each type of protein has a particular three-dimensional structure, which is determined by the order of the amino acids in its polypeptide chain. The final folded structure, or **conformation**, adopted by any polypeptide chain is determined by energetic considerations: a protein generally folds into the shape in which its free energy (*G*) is minimized. The folding process is thus energetically favorable, as it releases heat and increases the disorder of the universe (see Panel 3–1, pp. 96–97).

Protein folding has been studied in the laboratory using highly purified proteins. A protein can be unfolded, or *denatured*, by treatment with solvents that disrupt the noncovalent interactions holding the folded chain together. This treatment converts the protein into a flexible polypeptide chain that has lost its natural shape. Under the right conditions, when the

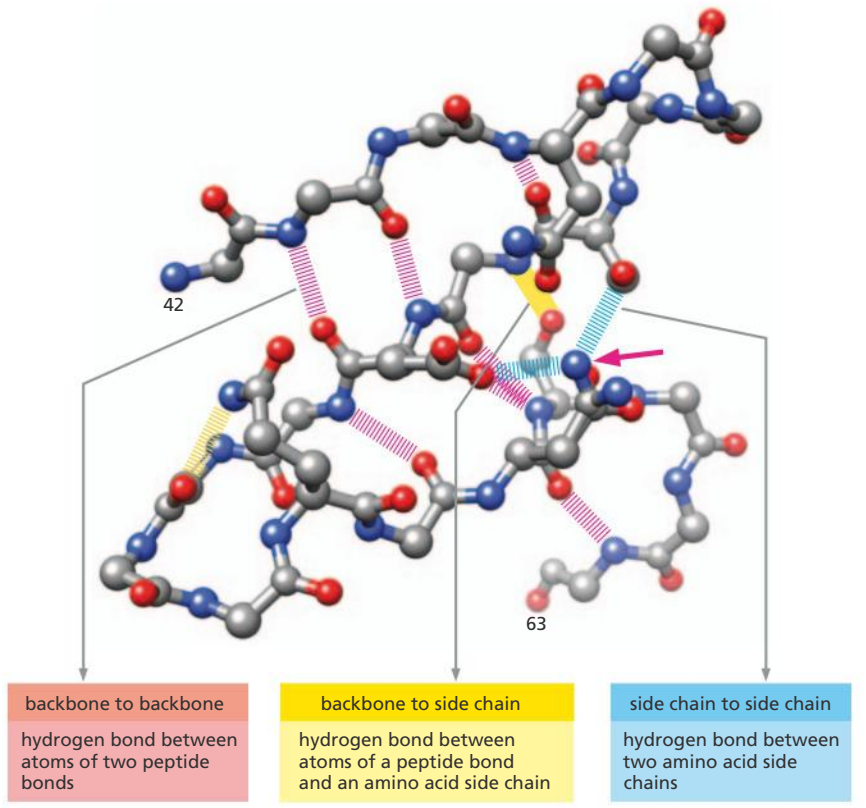


Figure 4–6 Hydrogen bonds within a protein molecule help stabilize its folded shape. Large numbers of hydrogen bonds form between adjacent regions of the folded polypeptide chain. The structure shown is a portion of the enzyme lysozyme. Hydrogen bonds between backbone atoms are shown in *red*; those between the backbone and a side chain are shown in *yellow*; and those between atoms of two side chains are shown in *blue*. Note that the same amino acid side chain can make multiple hydrogen bonds (*red* arrow). The atoms are colored as in Figure 4–1, although the hydrogen atoms are not shown. (After C.K. Mathews, K.E. van Holde, and K.G. Ahern, Biochemistry, 3rd ed. San Francisco: Benjamin Cummings, 2000.)

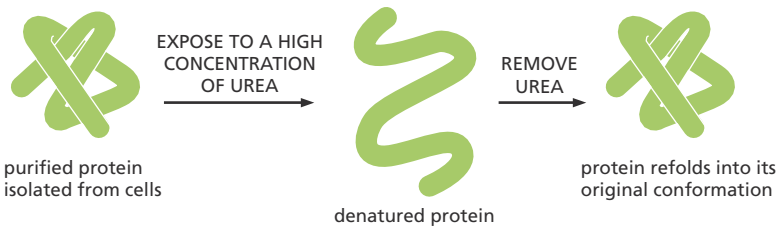


Figure 4–7 Denatured proteins can often recover their natural shapes. This type of experiment demonstrates that the conformation of a protein is determined solely by its amino acid sequence. Renaturation requires the correct conditions and works best for small proteins.

denaturing solvent is removed, the protein often refolds spontaneously into its original conformation—a process called *renaturation* (**Figure 4–7**). The fact that a denatured protein can, on its own, refold into the correct conformation indicates that all the information necessary to specify the three-dimensional shape of a protein is contained in its amino acid sequence.

Each protein normally folds into a single stable conformation. This conformation, however, often changes slightly when the protein interacts with other molecules in the cell. This change in shape is crucial to the function of the protein, as we discuss later.

When proteins fold incorrectly, they sometimes form aggregates that can damage cells and even whole tissues. Misfolded proteins are thought to contribute to a number of neurodegenerative disorders, such as Alzheimer's disease and Huntington's disease. Some infectious neurodegenerative diseases—including scrapie in sheep, bovine spongiform encephalopathy (BSE, or "mad cow" disease) in cattle, and Creutzfeldt–Jakob disease (CJD) in humans—are caused by misfolded proteins called prions. The misfolded prion form of a protein can convert the properly folded version of the protein in an infected brain into the abnormal conformation. This allows the misfolded prions, which tend to form aggregates, to spread rapidly from cell to cell, eventually causing the death of the affected animal or human (**Figure 4–8**). Prions are considered "infectious" because they can also spread from an affected individual to a normal individual via contaminated food, blood, or surgical instruments, for example.

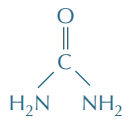
Although a protein chain can fold into its correct conformation without outside help, protein folding in a living cell is generally assisted by special proteins called *chaperone proteins*. Some of these chaperones bind to partly folded chains and help them to fold along the most energetically favorable pathway (**Figure 4–9**). Others form "isolation chambers" in which single polypeptide chains can fold without the risk of forming aggregates in the crowded conditions of the cytoplasm (**Figure 4–10**). In either case, the final three-dimensional shape of the protein is still specified by its amino acid sequence; chaperones merely make the folding process more efficient and reliable.

Proteins Come in a Wide Variety of Complicated Shapes

Proteins are the most structurally diverse macromolecules in the cell. Although they range in size from about 30 amino acids to more than

QUESTION 4–1

Urea used in the experiment shown in Figure 4–7 is a molecule that disrupts the hydrogen-bonded network of water molecules. Why might high concentrations of urea unfold proteins? The structure of urea is shown here.



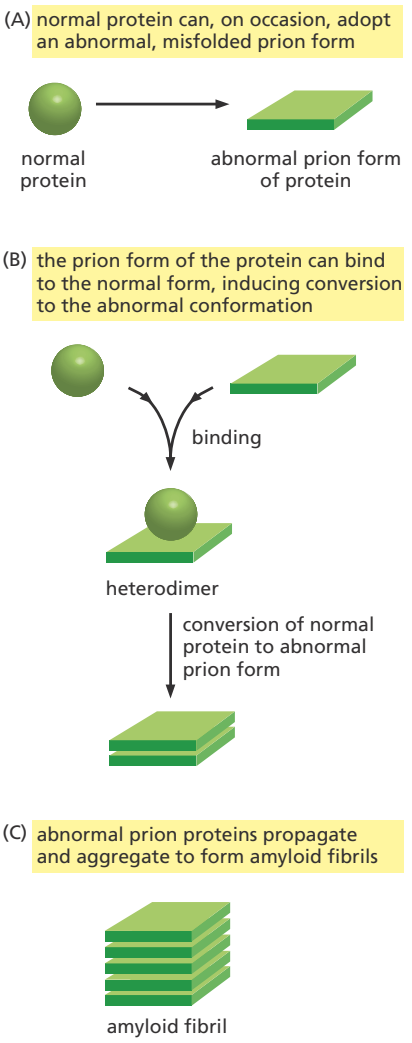


Figure 4–8 Prion diseases are caused by proteins whose misfolding is infectious. (A) The protein undergoes a rare conformational change to give an abnormally folded prion form. (B) The abnormal form causes the conversion of normal proteins in the host's brain into a misfolded prion form. (C) The prions aggregate into amyloid fibrils, which disrupt brain cell function, causing a neurodegenerative disorder, such as "mad cow" disease (see also Figure 4–18).

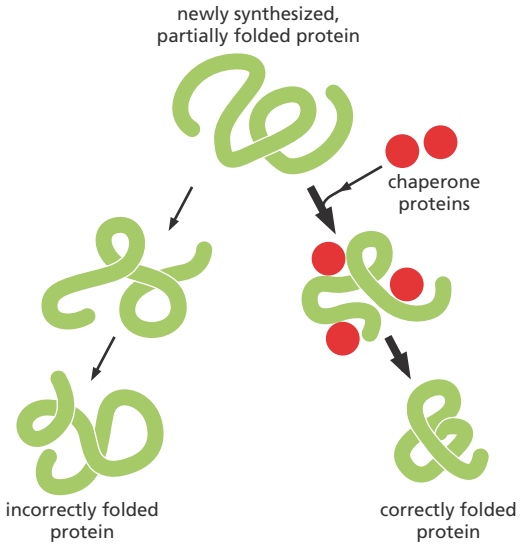


Figure 4–9 Chaperone proteins can guide the folding of a newly synthesized polypeptide chain. The chaperones bind to newly synthesized or partially folded chains and helping them to fold along the most energetically favorable pathway. Association of these chaperones with the target protein requires an input of energy from ATP hydrolysis.

10,000, the vast majority are between 50 and 2000 amino acids long. Proteins can be globular or fibrous, and they can form filaments, sheets, rings, or spheres (**Figure 4–11**). We will encounter many of these structures later in this chapter and throughout the book.

To date, the structures of about 100,000 different proteins have been determined. We discuss how scientists unravel these structures later in the chapter. Most proteins have a three-dimensional conformation so intricate and irregular that their structure would require an entire chapter to describe in detail. But we can get some sense of the intricacies of polypeptide structure by looking at the conformation of a relatively small protein, such as the bacterial transport protein *HPr*.

This small protein is only 88 amino acids long, and it serves as a carrier protein that facilitates the transport of sugar into bacterial cells. In **Figure 4–12**, we present HPr's three-dimensional structure in four different ways, each of which emphasizes different features of the protein. The backbone model (Figure 4–12A) shows the overall organization of the polypeptide chain and provides a straightforward way to compare the structures of related proteins. The ribbon model (Figure 4–12B) shows the polypeptide backbone in a way that emphasizes its various folds, which we describe in detail shortly. The wire model (Figure 4–12C) includes the positions of all the amino acid side chains; this view is especially useful

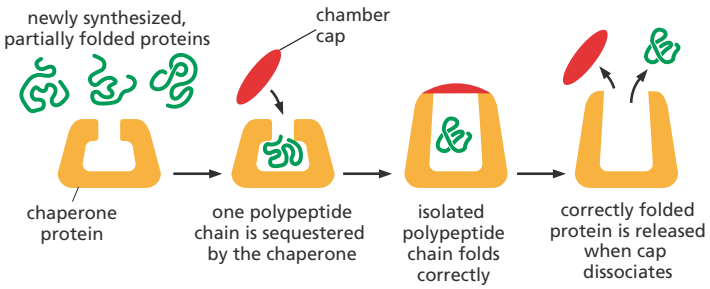


Figure 4–10 Other chaperone proteins act as isolation chambers that help a polypeptide fold. In this case, the barrel of the chaperone provides an enclosed chamber in which a newly synthesized polypeptide chain can fold without the risk of aggregating with other polypeptides in the crowded conditions of the cytoplasm. This system also requires an input of energy from ATP hydrolysis, mainly for the association and subsequent dissociation of the cap that closes off the chamber.

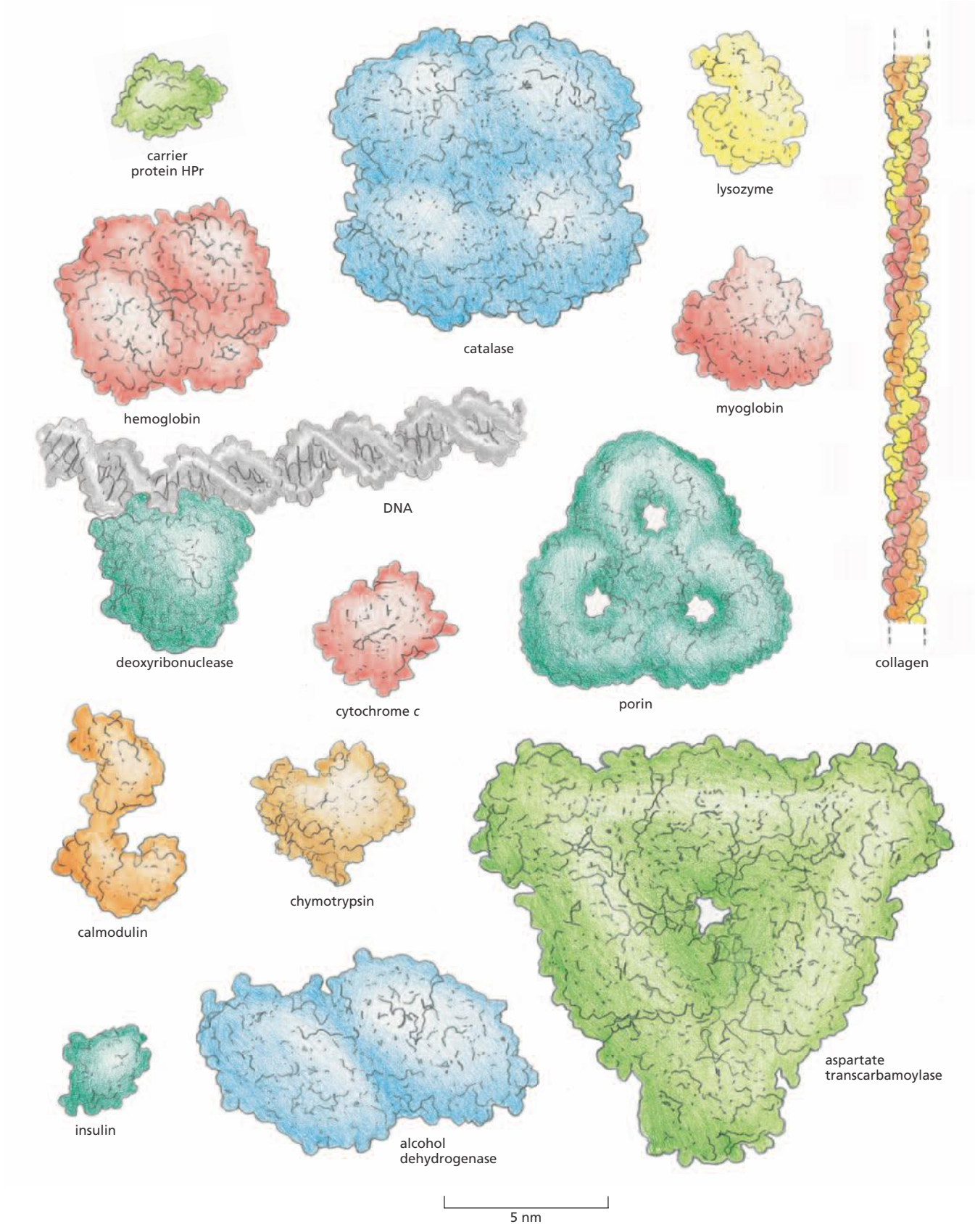


Figure 4–11 Proteins come in a variety of shapes and sizes. Each folded polypeptide is shown as a space-filling model, represented at the same scale. In the *top-left corner* is HPr, the small protein featured in detail in Figure 4–12. For comparison we also show a portion of a DNA molecule (*gray*) bound to the protein deoxyribonuclease. (After David S. Goodsell, Our Molecular Nature. New York: Springer-Verlag, 1996. With permission from Springer Science and Business Media.)

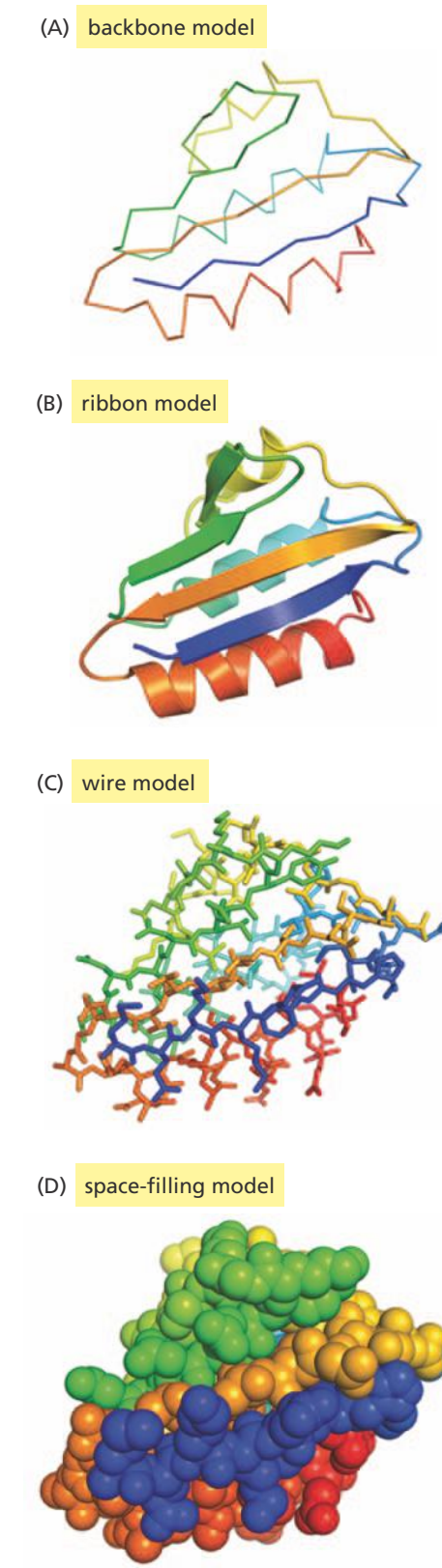


Figure 4–12 Protein conformation can be represented in a variety of ways. Shown here is the structure of the small bacterial transport protein HPr. The images are colored to make it easier to trace the path of the polypeptide chain. In these models, the region of polypeptide chain carrying the protein's N-terminus is *purple* and that near its C-terminus is *red*.

for predicting which amino acids might be involved in the protein's activity. Finally, the space-filling model (Figure 4–12D) provides a contour map of the protein surface, which reveals which amino acids are exposed on the surface and shows how the protein might look to a small molecule such as water or to another macromolecule in the cell.

The structures of larger proteins—or of multiprotein complexes—are even more complex. To visualize such detailed and complicated structures, scientists have developed various graphical and computer-based tools that generate a variety of images of a protein, only some of which are depicted in Figure 4–12. These images can be displayed on a computer screen and readily rotated and magnified to view all aspects of the structure (**Movie 4.1**).

When the three-dimensional structures of many different protein molecules are compared, it becomes clear that, although the overall conformation of each protein is unique, some regular folding patterns can be detected, as we discuss next.

The α Helix and the β Sheet Are Common Folding Patterns

More than 60 years ago, scientists studying hair and silk discovered two common folding patterns present in many different proteins. The first to be discovered, called the **α helix**, was found in the protein *α-keratin*, which is abundant in skin and its derivatives—such as hair, nails, and horns. Within a year of the discovery of the α helix, a second folded structure, called a **β sheet**, was found in the protein *fibroin*, the major constituent of silk. (Biologists often use Greek letters to name their discoveries, with the first example receiving the designation α, the second β, and so on.)

These two folding patterns are particularly common because they result from hydrogen bonds that form between the N–H and C=O groups in the polypeptide backbone (see Figure 4–6). Because the amino acid side chains are not involved in forming these hydrogen bonds, α helices and β sheets can be generated by many different amino acid sequences. In each case, the protein chain adopts a regular, repeating form. These structural features, and the shorthand cartoon symbols that are often used to represent them in models of protein structures, are presented in **Figure 4–13**.

Helices Form Readily in Biological Structures

The abundance of helices in proteins is, in a way, not surprising. A **helix** is a regular structure that resembles a spiral staircase. It is generated simply by placing many similar subunits next to one another, each in the same strictly repeated relationship to the one before. Because it is very rare for subunits to join up in a straight line, this arrangement will generally result in a helix (**Figure 4–14**). Depending on the twist of the staircase, a helix is said to be either right-handed or left-handed (Figure 4–14E). Handedness is not affected by turning the helix upside down, but it is reversed if the helix is reflected in a mirror.

An α helix is generated when a single polypeptide chain turns around itself to form a structurally rigid cylinder. A hydrogen bond is made between every fourth amino acid, linking the C=O of one peptide bond to

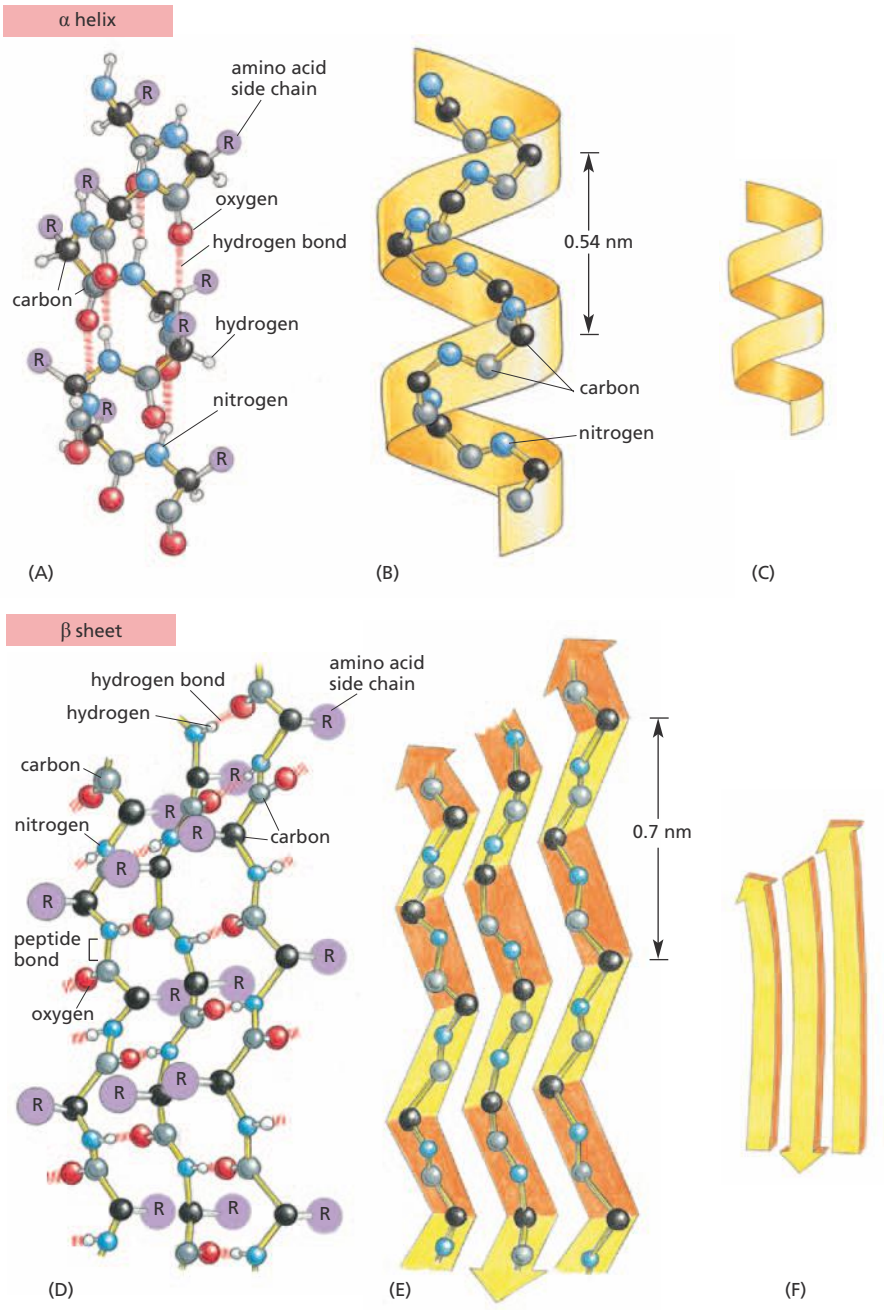


Figure 4–13 Polypeptide chains often fold into one of two orderly repeating forms known as an α helix and a β sheet. (A–C) In an α helix, the N–H of every peptide bond is hydrogen-bonded to the C=O of a neighboring peptide bond located four amino acids away in the same chain. (D–F) In a β sheet, several segments (strands) of an individual polypeptide chain are held together by hydrogen-bonding between peptide bonds in adjacent strands. The amino acid side chains in each strand project alternately above and below the plane of the sheet. In the example shown, the adjacent chains run in opposite directions, forming an *antiparallel β sheet*. (A) and (D) show all of the atoms in the polypeptide backbone, but the amino acid side chains are denoted by R. (B) and (E) show only the carbon (*black* and *gray*) and nitrogen (*blue*) backbone atoms, while (C) and (F) display the cartoon symbols that are used to represent the α helix and the β sheet in ribbon models of proteins (see Figure 4–12B).

the N–H of another (see Figure 4–13A). This gives rise to a regular right-handed helix with a complete turn every 3.6 amino acids (**Movie 4.2**).

Short regions of α helix are especially abundant in proteins that are embedded in cell membranes, such as transport proteins and receptors. We will see in Chapter 11 that those portions of a transmembrane protein that cross the lipid bilayer usually form an α helix that is composed largely of amino acids with nonpolar side chains. The polypeptide backbone, which is hydrophilic, is hydrogen-bonded to itself in the α helix, and it is shielded from the hydrophobic lipid environment of the membrane by its protruding nonpolar side chains (**Figure 4–15**).

Sometimes two (or three) α helices will wrap around one another to form a particularly stable structure known as a **coiled-coil**. This structure forms when the α helices have most of their nonpolar (hydrophobic) side chains on one side, so that they can twist around each other with

QUESTION 4–2

Remembering that the amino acid side chains projecting from each polypeptide backbone in a β sheet point alternately above and below the plane of the sheet (see Figure 4–13D), consider the following protein sequence: Leu-Lys-Val-Asp-Ile-Ser-Leu-Arg-Leu-Lys-Ile-Arg-Phe-Glu. Do you find anything remarkable about the arrangement of the amino acids in this sequence when incorporated into a β sheet? Can you make any predictions as to how the β sheet might be arranged in a protein? (Hint: consult the properties of the amino acids listed in Figure 4–3.)

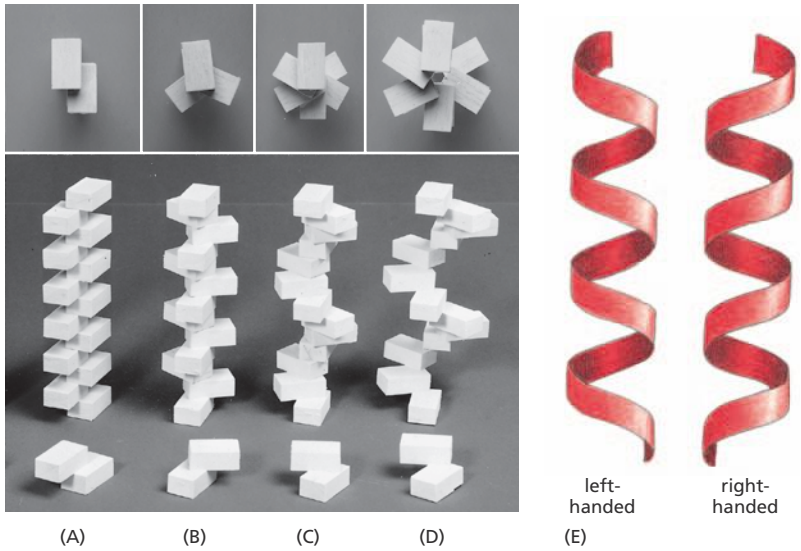


Figure 4–14 The helix is a common, regular, biological structure. A helix will form when a series of similar subunits bind to each other in a regular way. At the bottom, the interaction between two subunits is shown; behind them are the helices that result. These helices have two (A), three (B), or six (C and D) subunits per helical turn. At the top, the arrangement of subunits has been photographed from directly above the helix. Note that the helix in (D) has a wider path than that in (C), but the same number of subunits per turn. (E) A helix can be either right-handed or left-handed. As a reference, it is useful to remember that standard metal screws, which advance when turned clockwise, are right-handed. So to judge the handedness of a helix, imagine screwing it into a wall. Note that a helix preserves the same handedness when it is turned upside down.

these side chains facing inward—minimizing their contact with the aqueous cytosol (**Figure 4–16**). Long, rodlike coiled-coils form the structural framework for many elongated proteins. Examples include α-keratin, which forms the intracellular fibers that reinforce the outer layer of the skin, and myosin, the motor protein responsible for muscle contraction (discussed in Chapter 17).

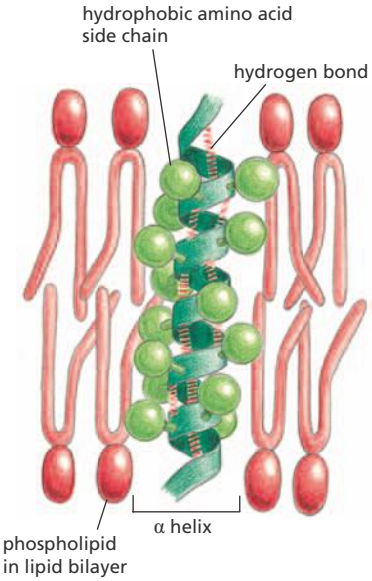


Figure 4–15 Many membrane-bound proteins cross the lipid bilayer as an α helix. The hydrophobic side chains of the amino acids forming the α helix contact the hydrophobic hydrocarbon tails of the phospholipid molecules, while the hydrophilic parts of the polypeptide backbone form hydrogen bonds with one another in the interior of the helix. About 20 amino acids are required to span a membrane in this way. Note that, despite the appearance of a space along the interior of the helix in this schematic diagram, the helix is not a channel: no ions or small molecules can pass through it.

β Sheets Form Rigid Structures at the Core of Many Proteins

A β sheet is made when hydrogen bonds form between segments of a polypeptide chain that lie side by side (see Figure 4–13D). When the neighboring segments run in the same orientation (say, from the N-terminus to the C-terminus), the structure is a *parallel β sheet*; when they run in opposite directions, the structure is an *antiparallel β sheet* (**Figure 4–17**). Both types of β sheet produce a very rigid, pleated structure, and they form the core of many proteins. Even the small bacterial protein HPr (see Figure 4–12) contains several β sheets.

β sheets have remarkable properties. They give silk fibers their extraordinary tensile strength. They also permit the formation of *amyloid fibers*—insoluble protein aggregates that include those associated with neurodegenerative disorders, such as Alzheimer's disease and prion diseases (see Figure 4–8). These structures, formed from abnormally folded proteins, are stabilized by β sheets that stack together tightly, with their amino acid side chains interdigitated like the teeth of a zipper (**Figure 4–18**). Although we tend to associate amyloid fibers with disease, many organisms take advantage of these stable structures to perform novel tasks. Infectious bacteria, for example, can use amyloid fibers to help form the biofilms that allow them to colonize host tissues. Other types of filamentous bacteria use amyloid fibers to extend filaments into the air, enabling the bacteria to disperse their spores far and wide.

Proteins Have Several Levels of Organization

A protein's structure does not end with α helices and β sheets; there are additional levels of organization. These levels are not independent but are built one upon the next to establish the three-dimensional structure of the entire protein. A protein's structure begins with its amino acid sequence, which is thus considered its **primary structure**. The next level of organization includes the α helices and β sheets that form within

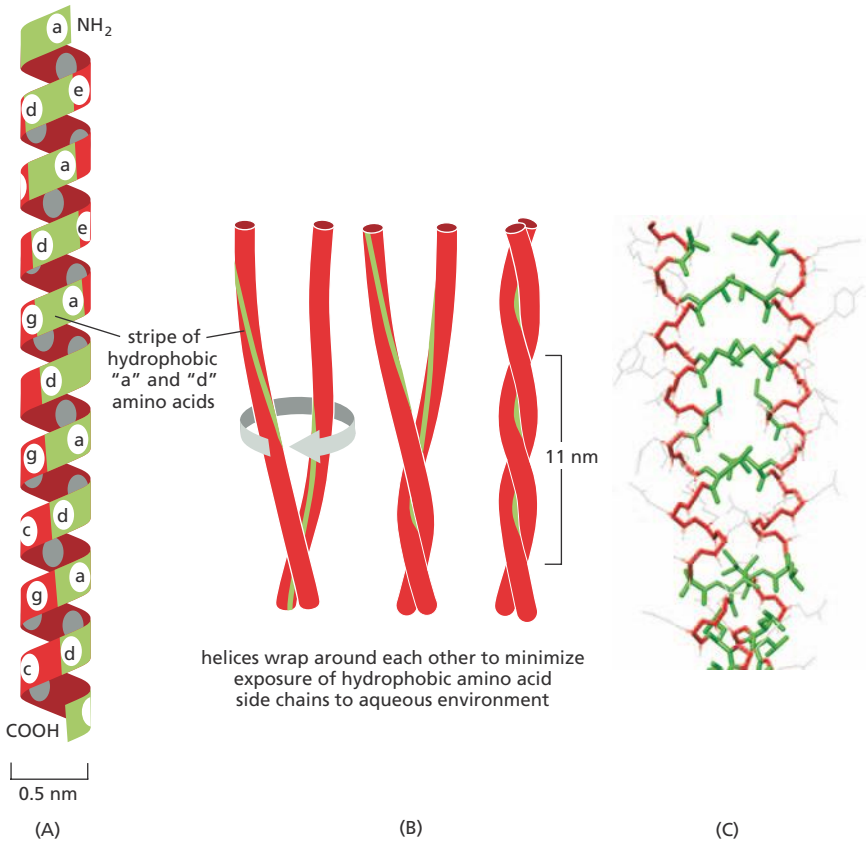


Figure 4–16 Intertwined α helices can form a stiff coiled-coil. In (A), a single α helix is shown, with successive amino acid side chains labeled in a sevenfold repeating sequence "abcdefg." Amino acids "a" and "d" in such a sequence lie close together on the cylinder surface, forming a stripe (shaded in *green*) that winds slowly around the α helix. Proteins that form coiled-coils typically have nonpolar amino acids at positions "a" and "d." Consequently, as shown in (B), the two α helices can wrap around each other, with the nonpolar side chains of one α helix interacting with the nonpolar side chains of the other, while the more hydrophilic amino acid side chains (shaded in *red*) are left exposed to the aqueous environment. (C) A portion of the atomic structure of a coiled-coil made of two α helices, as determined by X-ray crystallography. In this structure, atoms that form the backbone of the helices are shown in *red*; the interacting, nonpolar side chains are *green*, and the remaining side chains are *gray*. Coiled-coils can also form from three α helices (Movie 4.3).

certain segments of the polypeptide chain; these folds are elements of the protein's **secondary structure**. The full, three-dimensional conformation formed by an entire polypeptide chain—including the α helices, β sheets, random coils, and any other loops and folds that form between the N- and C-termini—is sometimes referred to as the **tertiary structure**. Finally, if the protein molecule is formed as a complex of more than one polypeptide chain, then the complete structure is designated its **quaternary structure**.

Studies of the conformation, function, and evolution of proteins have also revealed the importance of a level of organization distinct from the four just described. This organizational unit is the **protein domain**, which is defined as any segment of a polypeptide chain that can fold independently into a compact, stable structure. A protein domain usually contains between 40 and 350 amino acids—folded into α helices and β sheets and other elements of secondary structure—and it is the modular unit from which many larger proteins are constructed (Figure 4–19). The different domains of a protein are often associated with different functions. For example, the bacterial *catabolite activator protein* (*CAP*), illustrated in Figure 4–19, has two domains: the small domain binds to DNA, while the large domain binds cyclic AMP, a small intracellular signaling molecule. When the large domain binds cyclic AMP, it causes a conformational change in the protein that enables the small domain to bind to a specific DNA sequence and thereby promote the expression of an adjacent gene. To provide a sense of the many different domain structures observed in proteins, ribbon models of three different domains are shown in Figure 4–20.

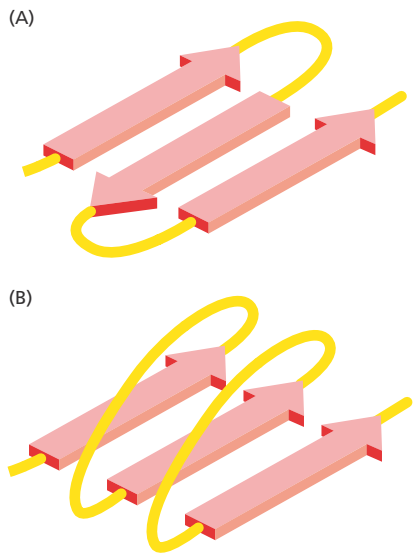


Figure 4–17 β sheets come in two varieties. (A) Antiparallel β sheet (see also Figure 4–13D). (B) Parallel β sheet. Both of these structures are common in proteins. By convention, the arrows point toward the C-terminus of the polypeptide chain (Movie 4.4).

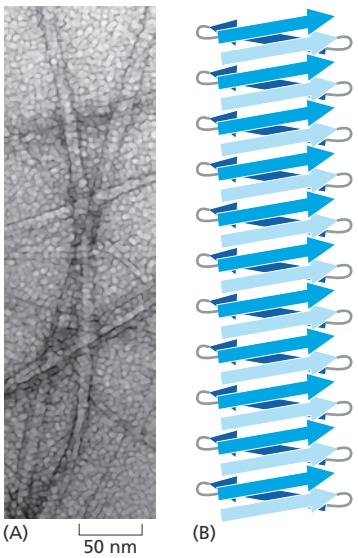


Figure 4–18 Stacking of β sheets allows some misfolded proteins to aggregate into amyloid fibers. (A) Electron micrograph shows an amyloid fiber formed from a segment of a yeast prion protein. (B) Schematic representation shows the stacking of β sheets that stabilize an individual amyloid fiber. (A, courtesy of David Eisenberg.)

Many Proteins Also Contain Unstructured Regions

Small protein molecules, such as the oxygen-carrying muscle protein myoglobin, contain only a single domain (see Figure 4–11). Larger proteins can contain as many as several dozen domains, which are usually connected by relatively unstructured lengths of polypeptide chain. Such regions of polypeptide chain lacking any definite structure, which continually bend and flex due to thermal buffeting, are abundant in cells. These **intrinsically disordered sequences** are often found as short stretches linking domains in otherwise highly ordered proteins. Other proteins, however, are almost entirely without secondary structure and exist as unfolded polypeptide chains in the cytosol.

Intrinsically disordered sequences remained undetected for many years. Their lack of folded structure makes them prime targets for the proteolytic enzymes that are released when cells are fractionated to isolate their molecular components (see Panel 4–3, pp. 164–165). Unstructured sequences also fail to form protein crystals and for this reason escape the attention of X-ray crystallographers (see How We Know, pp. 162–163). Indeed, the ubiquity of disordered sequences became appreciated only after bioinformatics methods were developed that could recognize them from their amino acid sequences. Present estimates suggest that a third of all eukaryotic proteins have long unstructured regions in their polypeptide chain (greater than 30 amino acids in length), while a substantial number of eukaryotic proteins are mostly disordered under normal conditions.

Unstructured sequences have a variety of important functions in cells. Being able to flex and bend, they can wrap around one or more target proteins like a scarf, binding with both high specificity and low affinity (**Figure 4–21**). By forming flexible tethers between the compact domains in a protein, they provide flexibility while increasing the frequency of encounters between the domains (Figure 4–21). They can help *scaffold proteins* bring together proteins in an intracellular signaling pathway, facilitating interactions (Figure 4–21). They also give proteins like elastin

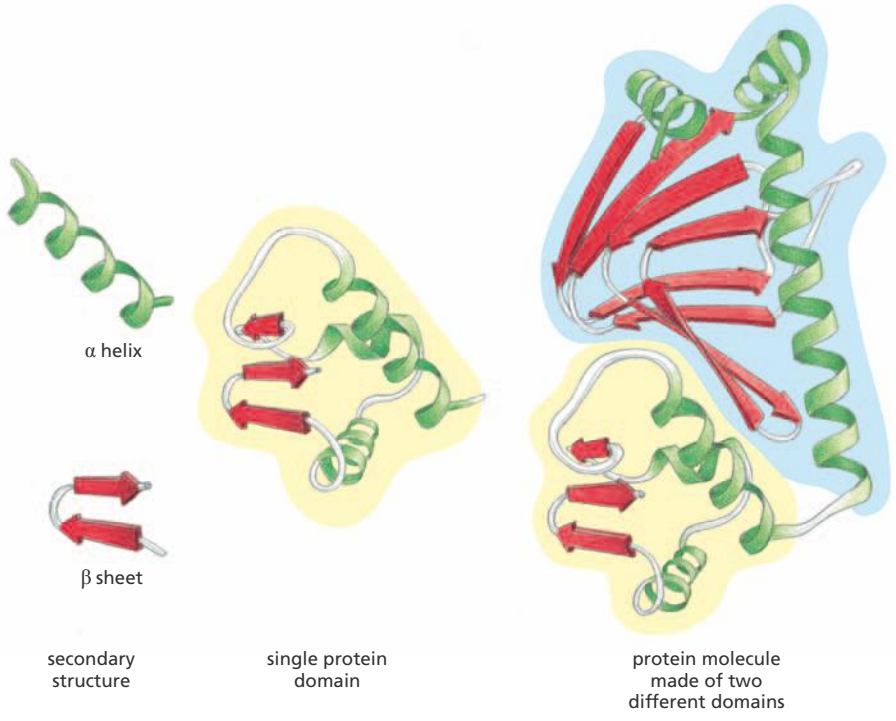


Figure 4–19 Many proteins are composed of separate functional domains. Elements of secondary structure such as α helices and β sheets pack together into stable, independently folding, globular elements called protein domains. A typical protein molecule is built from one or more domains, linked by a region of polypeptide chain that is often relatively unstructured. The ribbon diagram on the right represents the bacterial transcription regulator protein CAP, with one large domain (outlined in *blue*) and one small domain (outlined in *yellow*).

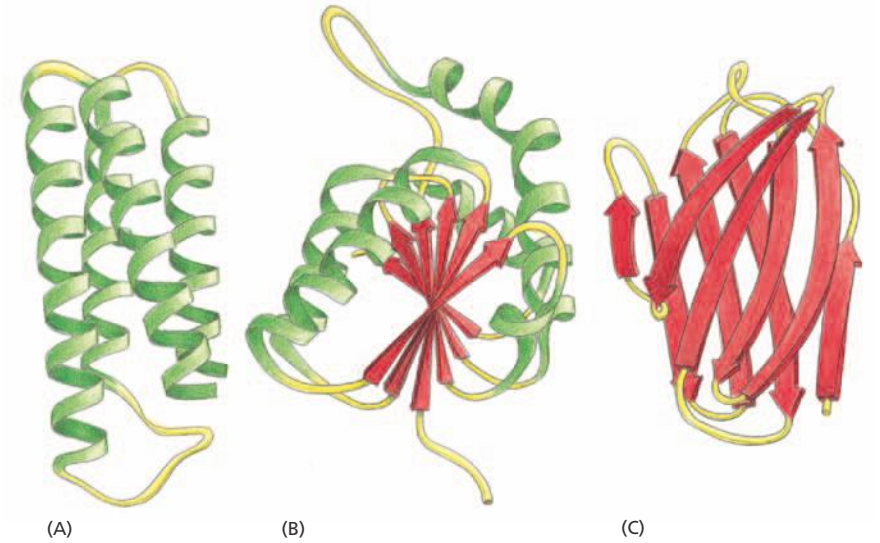


Figure 4–20 Ribbon models show three different protein domains. (A) Cytochrome b_{562} is a single-domain protein involved in electron transfer in *E. coli*. It is composed almost entirely of α helices. (B) The NAD-binding domain of the enzyme lactic dehydrogenase is composed of a mixture of α helices and β sheets. (C) An immunoglobulin domain of an antibody molecule is composed of a sandwich of two antiparallel β sheets. In these examples, the α helices are shown in *green*, while strands organized as β sheets are denoted by *red arrows*. The protruding loop regions (*yellow*) are often unstructured and can provide binding sites for other molecules. (Redrawn from originals courtesy of Jane Richardson.)

the ability to form rubberlike fibers, allowing our tendons and skin to recoil after being stretched. In addition to providing structural flexibility, unstructured sequences are also ideal substrates for the addition of chemical groups that control the way many proteins behave—a topic we discuss at length later in the chapter.

Few of the Many Possible Polypeptide Chains Will Be Useful

In theory, a vast number of different polypeptide chains could be made from 20 different amino acids. Because each amino acid is chemically distinct and could, in principle, occur at any position, a polypeptide chain four amino acids long has 20 × 20 × 20 × 20 = 160,000 different possible sequences. In other words, for a polypeptide that is *n* amino acids long, 20^n different chains are possible. For a typical protein length of 300 amino acids, more than 20^{300} (that's 10^{390}) different polypeptide chains could theoretically be made.

Of the unimaginably large collection of potential polypeptide sequences, only a miniscule fraction is actually present in cells. That's because many biological functions depend on proteins with stable, well-defined three-dimensional conformations. This requirement restricts the list of possible polypeptide sequences. Another constraint is that functional proteins

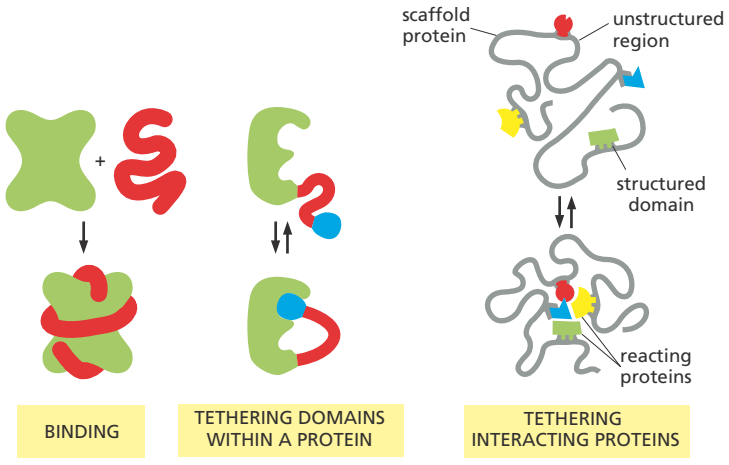


Figure 4–21 Unstructured regions of a polypeptide chain in proteins can peform many functions. A few of these functions are illustrated here.

QUESTION 4–3

Random mutations only very rarely result in changes in a protein that improve its usefulness for the cell, yet useful mutations are selected in evolution. Because these changes are so rare, for each useful mutation there are innumerable mutations that lead to either no improvement or inactive proteins. Why, then, do cells not contain millions of proteins that are of no use?

must be "well-behaved" and not engage in unwanted associations with other proteins in the cell—forming insoluble protein aggregates, for example. Many potential proteins would therefore have been eliminated by natural selection through the long trial-and-error process that underlies evolution (discussed in Chapter 9).

Thanks to this rigorous process of selection, the amino acid sequences of many present-day proteins have evolved to guarantee that the polypeptide will adopt a stable conformation—one that bestows upon the protein the exact chemical properties that will enable it to perform a particular function. Such proteins are so precisely built that a change in even a few atoms in one amino acid can sometimes disrupt the structure of a protein and thereby eliminate its function. In fact, the structures of many proteins—and their constituent domains—are so stable and effective that they have been conserved throughout evolution among many diverse organisms. The three-dimensional structures of the DNA-binding domains from the yeast α2 protein and the *Drosophila* Engrailed protein, for example, are almost completely superimposable, even though these organisms are separated by more than a billion years of evolution. Other proteins, however, have changed their structure and function over evolutionary time, as we now discuss.

Proteins Can Be Classified into Families

Once a protein had evolved a stable conformation with useful properties, its structure could be modified over time to enable it to perform new functions. We know that this occurred quite often during evolution, because many present-day proteins can be grouped into **protein families**, in which each family member has an amino acid sequence and a three-dimensional conformation that closely resemble those of the other family members.

Consider, for example, the *serine proteases*, a family of protein-cleaving (proteolytic) enzymes that includes the digestive enzymes chymotrypsin, trypsin, and elastase, as well as several proteases involved in blood clotting. When any two of these enzymes are compared, portions of their amino acid sequences are found to be nearly the same. The similarity of their three-dimensional conformations is even more striking: most of the detailed twists and turns in their polypeptide chains, which are several hundred amino acids long, are virtually identical (**Figure 4–22**). The various serine proteases nevertheless have distinct enzymatic activities, each cleaving different proteins or the peptide bonds between different types of amino acids.

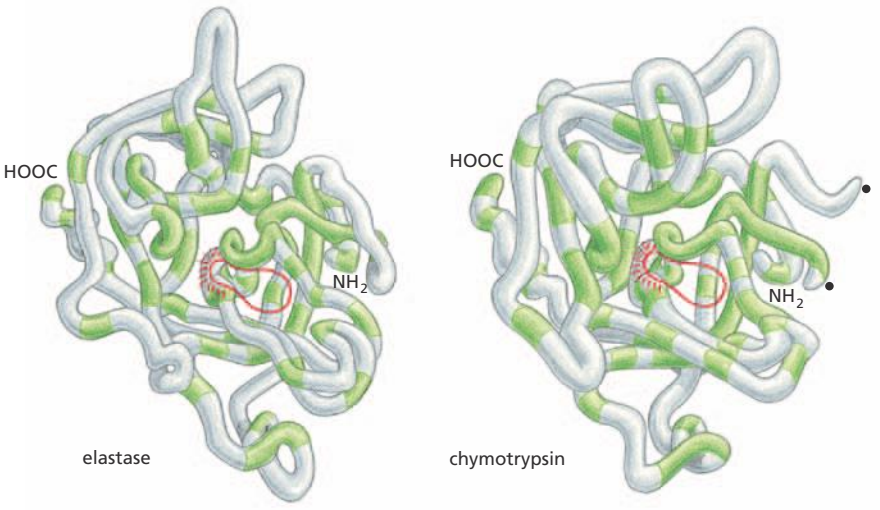


Figure 4–22 Serine proteases constitute a family of proteolytic enzymes. Backbone models of two serine proteases, elastase and chymotrypsin, are illustrated. Although only those amino acid sequences in the polypeptide chain shaded in *green* are the same in the two proteins, the two conformations are very similar nearly everywhere. Nonetheless, the two proteases prefer different substrates. The active site of each enzyme—where its substrates are bound and cleaved—is circled in *red*.

Serine proteases derive their name from the amino acid serine, which directly participates in the cleavage reaction. The two *black dots* on the *right side* of the chymotrypsin molecule mark the two ends created where the enzyme has cleaved its own backbone.

Large Protein Molecules Often Contain More Than One Polypeptide Chain

The same type of weak noncovalent bonds that enable a polypeptide chain to fold into a specific conformation also allow proteins to bind to each other to produce larger structures in the cell. Any region on a protein's surface that interacts with another molecule through sets of noncovalent bonds is termed a *binding site*. A protein can contain binding sites for a variety of molecules, large and small. If a binding site recognizes the surface of a second protein, the tight binding of two folded polypeptide chains at this site will create a larger protein, whose quaternary structure has a precisely defined geometry. Each polypeptide chain in such a protein is called a **subunit**, and each subunit may contain more than one domain.

In the simplest case, two identical, folded polypeptide chains form a symmetrical complex of two protein subunits (called a *dimer*) that is held together by interactions between two identical binding sites. The CAP protein in bacterial cells is a dimer (**Figure 4–23A**) formed from two identical copies of the protein subunit shown previously in Figure 4–19. Many other symmetrical protein complexes, formed from multiple copies of the same polypeptide chain, are commonly found in cells. The enzyme *neuraminidase*, for example, consists of a ring of four identical protein subunits (**Figure 4–23B**).

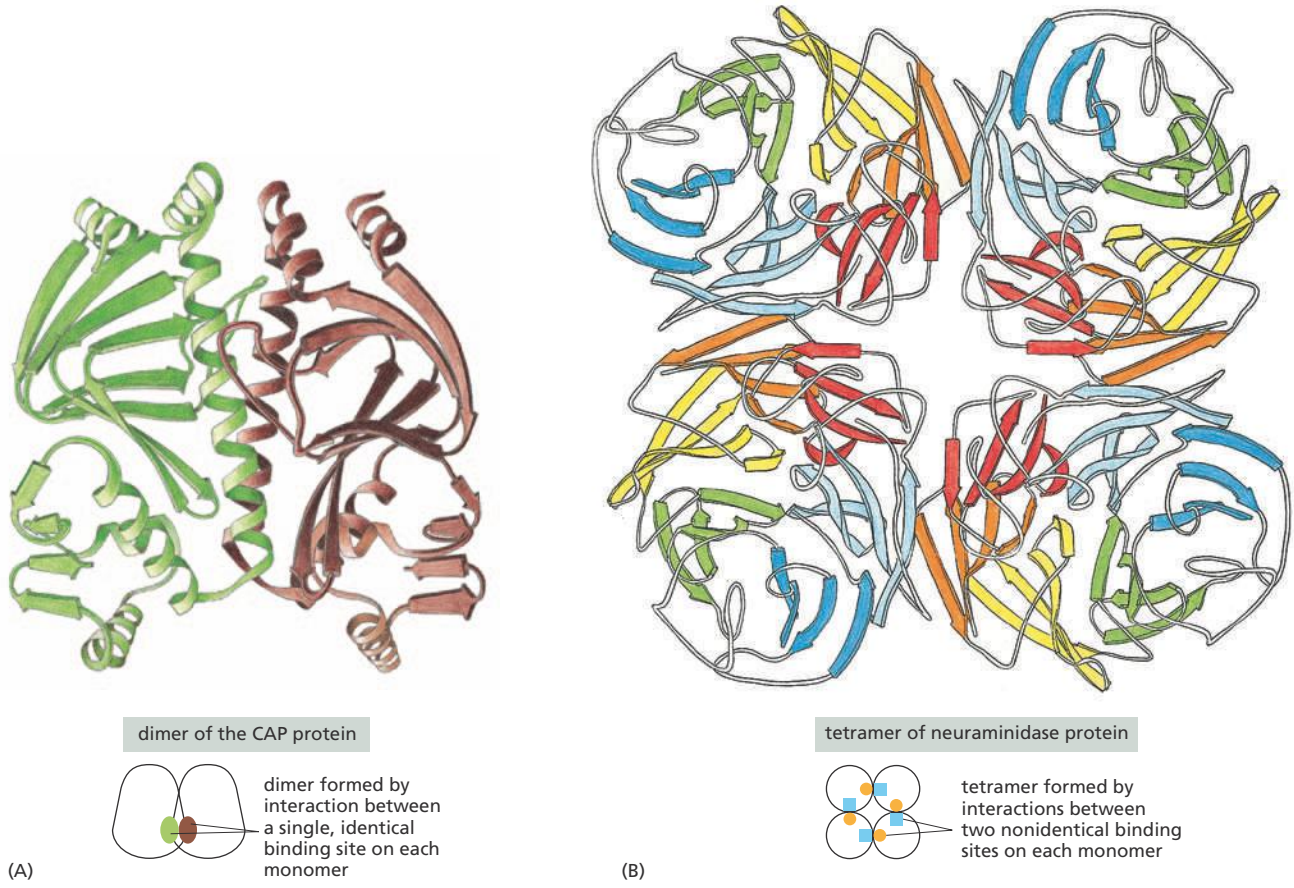


Figure 4–23 Many protein molecules contain multiple copies of the same protein subunit. (A) A symmetrical dimer. The CAP protein is a complex of two identical polypeptide chains (see also Figure 4–19). (B) A symmetrical homotetramer. The enzyme neuraminidase exists as a ring of four identical polypeptide chains. For both (A) and (B), a small schematic below the structure emphasizes how the repeated use of the same binding interaction forms the structure. In (A), the use of the same binding site on each monomer (represented by *brown* and *green ovals*) causes the formation of a symmetrical dimer. In (B), a pair of nonidentical binding sites (represented by *orange circles* and *blue squares*) causes the formation of a symmetrical tetramer.

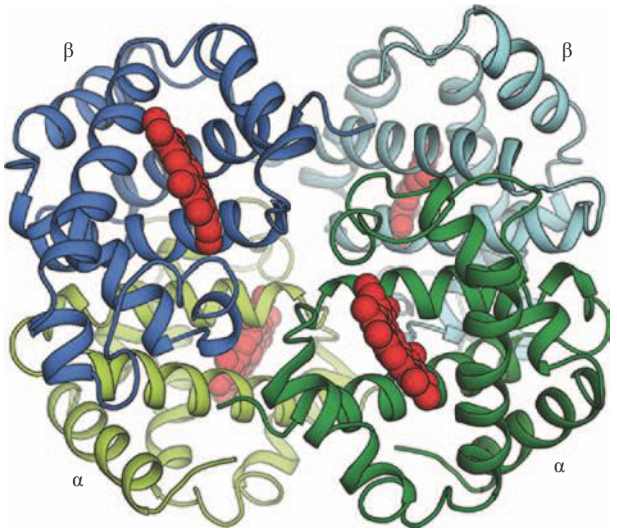


Figure 4–24 Some proteins are formed as a symmetrical assembly of two different subunits. Hemoglobin, an oxygen-carrying protein abundant in red blood cells, contains two copies of α-globin (*green*) and two copies of β-globin (*blue*). Each of these four polypeptide chains contains a heme molecule (*red*), where oxygen (O_2) is bound. Thus, each molecule of hemoglobin in the blood carries four molecules of oxygen.

Other proteins contain two or more different polypeptide chains. *Hemoglobin*, the protein that carries oxygen in red blood cells, is a particularly well-studied example. The protein contains two identical α-globin subunits and two identical β-globin subunits, symmetrically arranged (**Figure 4–24**). Many proteins contain multiple subunits, and they can be very large (**Movie 4.5**).

Proteins Can Assemble into Filaments, Sheets, or Spheres

Proteins can form even larger assemblies than those discussed so far. Most simply, a chain of identical protein molecules can be formed if the binding site on one protein molecule is complementary to another region on the surface of another protein molecule of the same type. Because each protein molecule is bound to its neighbor in an identical way (see Figure 4–14), the molecules will often be arranged in a helix that can be extended indefinitely in either direction (**Figure 4–25**). This type of arrangement can produce an extended protein filament. An actin filament, for example, is a long, helical structure formed from many molecules of the protein actin (**Figure 4–26**). Actin is extremely abundant in eukaryotic cells, where it forms one of the major filament systems of the cytoskeleton (discussed in Chapter 17). Other sets of identical proteins associate to form tubes, as in the microtubules of the cytoskeleton (**Figure 4–27**), or cagelike spherical shells, as in the protein coats of virus particles (**Figure 4–28**).

Many large structures, such as viruses and ribosomes, are built from a mixture of one or more types of protein plus RNA or DNA molecules. These structures can be isolated in pure form and dissociated into their constituent macromolecules. It is often possible to mix the isolated components back together and watch them reassemble spontaneously into the original structure. This demonstrates that all the information needed for assembly of the complicated structure is contained in the macromolecules themselves. Experiments of this type show that much of the

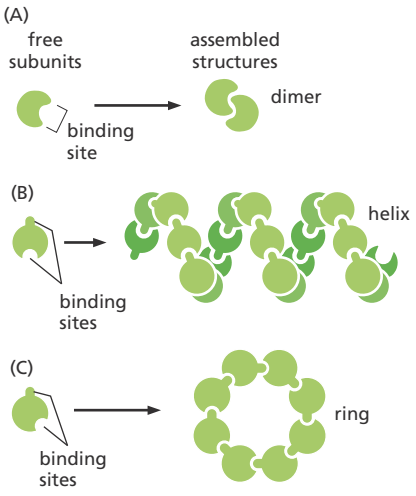


Figure 4–25 Identical protein subunits can assemble into complex structures. (A) A protein with just one binding site can form a dimer with another identical protein. (B) Identical proteins with two different binding sites will often form a long, helical filament. (C) If the two binding sites are disposed appropriately in relation to each other, the protein subunits will form a closed ring instead of a helix (see also Figure 4–23B).

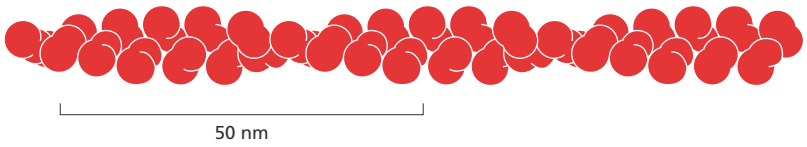


Figure 4–26 An actin filament is composed of identical protein subunits. The helical array of actin molecules in a filament often contains thousands of molecules and extends for micrometers in the cell.

structure of a cell is self-organizing: if the required proteins are produced in the right amounts, the appropriate structures will form automatically.

Some Types of Proteins Have Elongated Fibrous Shapes

Most of the proteins we have discussed so far are **globular proteins**, in which the polypeptide chain folds up into a compact shape like a ball with an irregular surface. Enzymes, for example, tend to be globular proteins: even though many are large and complicated, with multiple subunits, most have a quaternary structure with an overall rounded shape (see Figure 4–11). In contrast, other proteins have roles in the cell that require them to span a large distance. These proteins generally have a relatively simple, elongated three-dimensional structure and are commonly referred to as **fibrous proteins**.

One large class of intracellular fibrous proteins resembles α-keratin, which we met earlier when we introduced the α-helix. Keratin filaments are extremely stable: long-lived structures such as hair, horns, and nails are composed mainly of this protein. An α-keratin molecule is a dimer of two identical subunits, with the long α helices of each subunit forming a coiled-coil (see Figure 4–16). These coiled-coil regions are capped at either end by globular domains containing binding sites that allow them to assemble into ropelike *intermediate filaments*—a component of the cytoskeleton that gives cells mechanical strength (discussed in Chapter 17).

Fibrous proteins are especially abundant outside the cell, where they form the gel-like *extracellular matrix* that helps bind cells together to form tissues. These proteins are secreted by the cells into their surroundings, where they often assemble into sheets or long fibrils. *Collagen* is the most abundant of these fibrous extracellular proteins in animal tissues. A collagen molecule consists of three long polypeptide chains, each containing the nonpolar amino acid glycine at every third position. This regular structure allows the chains to wind around one another to generate a long, regular, triple helix with glycine at its core (**Figure 4–29A**). Many such

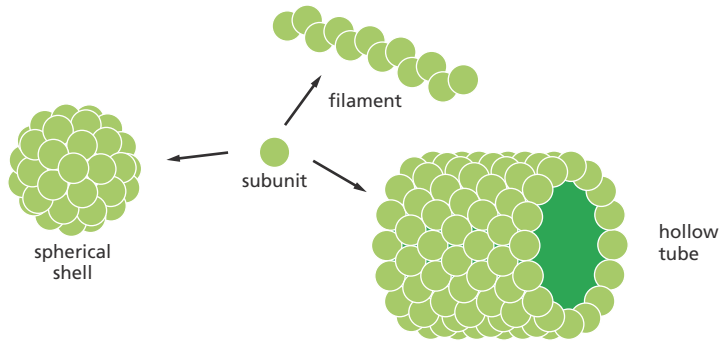


Figure 4–27 A single type of protein subunit can pack together to form a filament, a hollow tube, or a spherical shell. Actin subunits, for example, form actin filaments (see Figure 4–26), whereas tubulin subunits form hollow microtubules, and some virus proteins form a spherical shell (capsid) that encloses the viral genome (see Figure 4–28).

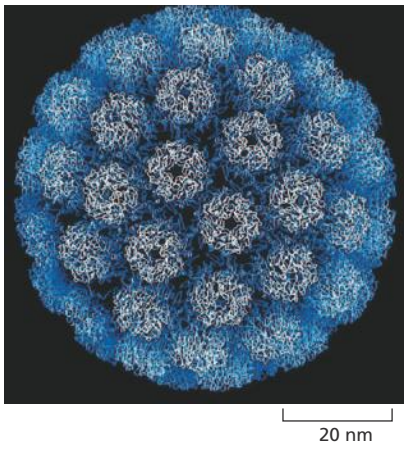


Figure 4–28 Many viral capsids are more or less spherical protein assemblies. They are formed from many copies of a small set of protein subunits. The nucleic acid of the virus (DNA or RNA) is packaged inside. The structure of the simian virus SV40, shown here, was determined by X-ray crystallography and is known in atomic detail. (Courtesy of Robert Grant, Stephan Crainic, and James M. Hogle.)

collagen molecules bind to one another side-by-side and end-to-end to create long overlapping arrays called *collagen fibrils*, which are extremely strong and help hold tissues together, as described in Chapter 20.

In complete contrast to collagen is another fibrous protein in the extracellular matrix, *elastin*. Elastin molecules are formed from relatively loose and unstructured polypeptide chains that are covalently cross-linked into a rubberlike elastic meshwork. The resulting *elastic fibers* enable skin and other tissues, such as arteries and lungs, to stretch and recoil without tearing. As illustrated in **Figure 4–29B**, the elasticity is due to the ability of the individual protein molecules to uncoil reversibly whenever they are stretched.

Extracellular Proteins Are Often Stabilized by Covalent Cross-Linkages

Many protein molecules are either attached to the outside of a cell's plasma membrane or secreted as part of the extracellular matrix, which exposes them to extracellular conditions. To help maintain their structures, the polypeptide chains in such proteins are often stabilized by covalent cross-linkages. These linkages can either tie together two amino acids in the same polypeptide chain or join together many polypeptide chains in a large protein complex—as for the collagen fibrils and elastic fibers just described.

The most common covalent cross-links in proteins are sulfur–sulfur bonds. These **disulfide bonds** (also called *S–S bonds*) are formed before a protein is secreted by an enzyme in endoplasmic reticulum that links together two –SH groups from cysteine side chains that are adjacent in the folded protein (**Figure 4–30**). Disulfide bonds do not change a protein's conformation, but instead act as a sort of "atomic staple" to reinforce the protein's most favored conformation. For example, lysozyme—an

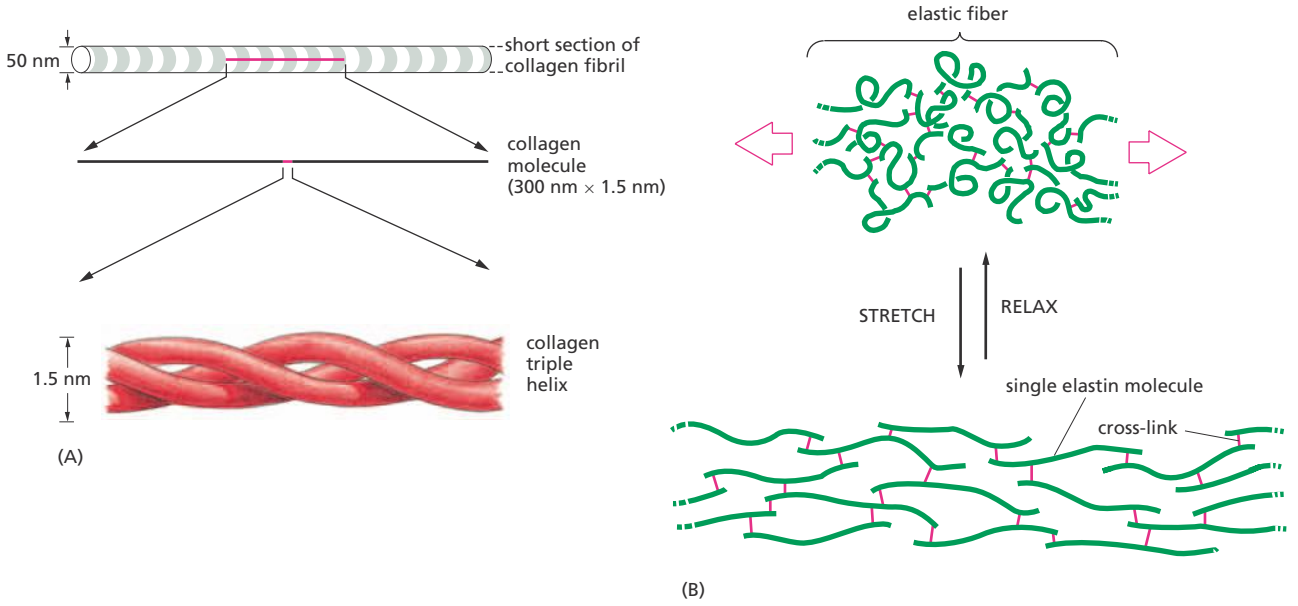


Figure 4–29 Collagen and elastin are abundant extracellular fibrous proteins. (A) A collagen molecule is a triple helix formed by three extended protein chains that wrap around one another. Many rodlike collagen molecules are cross-linked together in the extracellular space to form collagen fibrils (*top*), which have the tensile strength of steel. The striping on the collagen fibril is caused by the regular repeating arrangement of the collagen molecules within the fibril. (B) Elastin molecules are cross-linked together by covalent bonds (*red*) to form rubberlike, elastic fibers. Each elastin polypeptide chain uncoils into a more extended conformation when the fiber is stretched, and recoils spontaneously as soon as the stretching force is relaxed.

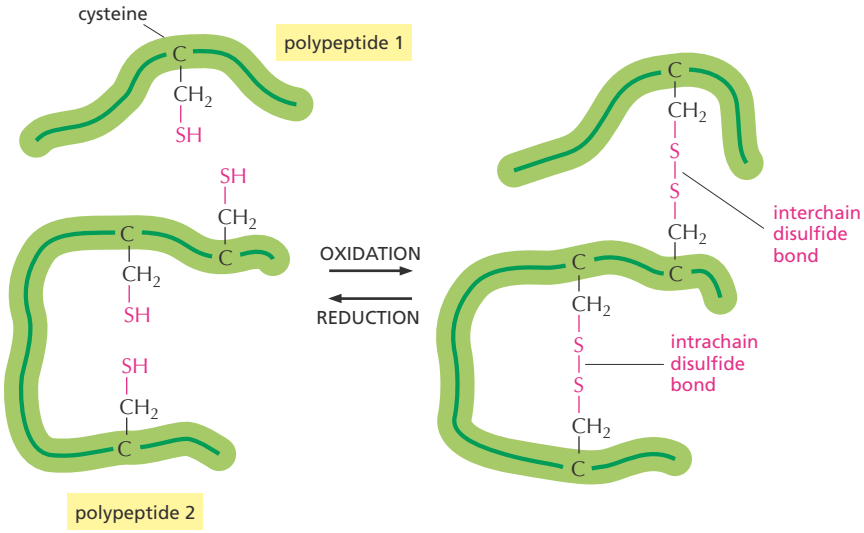


Figure 4–30 Disulfide bonds help stabilize a favored protein conformation. This diagram illustrates how covalent disulfide bonds form between adjacent cysteine side chains by the oxidation of their –SH groups. As indicated, these cross-links can join either two parts of the same polypeptide chain or two different polypeptide chains. Because the energy required to break one covalent bond is much larger than the energy required to break even a whole set of noncovalent bonds (see Table 2–1, p. 48), a disulfide bond can have a major stabilizing effect on a protein's folded structure (**Movie 4.6**).

enzyme in tears, saliva, and other secretions that can disrupt bacterial cell walls—retains its antibacterial activity for a long time because it is stabilized by such disulfide cross-links.

Disulfide bonds generally do not form in the cell cytosol, where a high concentration of reducing agents converts such bonds back to cysteine –SH groups. Apparently, proteins do not require this type of structural reinforcement in the relatively mild conditions in the cytosol.

HOW PROTEINS WORK

As we have just seen, proteins are made from an enormous variety of amino acid sequences and can fold into a unique shape. The surface topography of a protein's side chains endows each protein with a unique function, based on its chemical properties. The union of structure, chemistry, and function gives proteins the extraordinary ability to orchestrate the large number of dynamic processes that occur in cells.

Thus, for proteins, form and function are inextricably linked. But the fundamental question remains: How do proteins actually work? In this section, we will see that the activity of proteins depends on their ability to bind specifically to other molecules, allowing them to act as catalysts, structural supports, tiny motors, and so on. The examples we review here by no means exhaust the vast functional repertoire of proteins. However, the specialized functions of the proteins you will encounter elsewhere in this book are based on the same principles.

All Proteins Bind to Other Molecules

The biological properties of a protein molecule depend on its physical interaction with other molecules. Antibodies attach to viruses or bacteria as part of the body's defenses; the enzyme hexokinase binds glucose and ATP to catalyze a reaction between them; actin molecules bind to one another to assemble into long filaments; and so on. Indeed, all proteins stick, or bind, to other molecules in a specific manner. In some cases, this binding is very tight; in others, it is weak and short-lived. As we saw in Chapter 3, the affinity of an enzyme for its substrate is reflected in its K_M: the lower the K_M, the tighter the binding.

Regardless of its strength, the binding of a protein to other biological molecules always shows great *specificity*: each protein molecule can bind to just one or a few molecules out of the many thousands of different

QUESTION 4–4

Hair is composed largely of fibers of the protein keratin. Individual keratin fibers are covalently cross-linked to one another by many disulfide (S–S) bonds. If curly hair is treated with mild reducing agents that break a few of the cross-links, pulled straight, and then oxidized again, it remains straight. Draw a diagram that illustrates the three different stages of this chemical and mechanical process at the level of the keratin filaments, focusing on the disulfide bonds. What do you think would happen if hair were treated with strong reducing agents that break all the disulfide bonds?

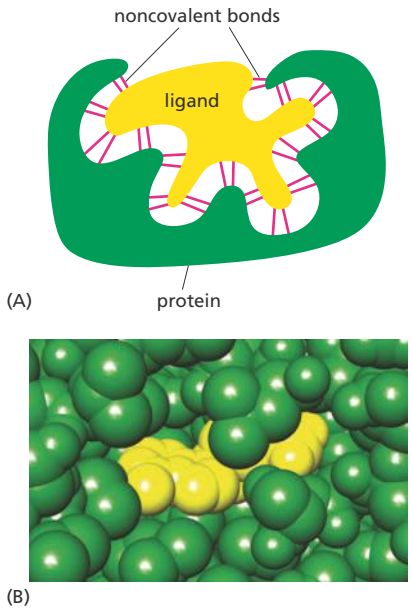


Figure 4–31 The binding of a protein to another molecule is highly selective. Many weak interactions are needed to enable a protein to bind tightly to a second molecule (a ligand). The ligand must therefore fit precisely into the protein's binding site, like a hand into a glove, so that a large number of noncovalent interactions can be formed between the protein and the ligand. (A) Schematic drawing shows the binding of a hypothetical protein and ligand; (B) space-filling model.

molecules it encounters. Any substance that is bound by a protein—whether it is an ion, a small organic molecule, or a macromolecule—is referred to as a **ligand** for that protein (from the Latin *ligare*, "to bind").

The ability of a protein to bind selectively and with high affinity to a ligand is due to the formation of a set of weak, noncovalent interactions—hydrogen bonds, electrostatic attractions, and van der Waals attractions—plus favorable hydrophobic forces (see Panel 2–7, pp. 78–79). Each individual noncovalent interaction is weak, so that effective binding requires many such bonds to be formed simultaneously. This is possible only if the surface contours of the ligand molecule fit very closely to the protein, matching it like a hand in a glove (**Figure 4–31**).

When molecules have poorly matching surfaces, few noncovalent interactions occur, and the two molecules dissociate as rapidly as they come together. This is what prevents incorrect and unwanted associations from forming between mismatched molecules. At the other extreme, when many noncovalent interactions are formed, the association can persist for a very long time. Strong binding between molecules occurs in cells whenever a biological function requires that the molecules remain tightly associated for a long time—for example, when a group of macromolecules come together to form a functional subcellular structure such as a ribosome.

The region of a protein that associates with a ligand, known as its **binding site**, usually consists of a cavity in the protein surface formed by a particular arrangement of amino acid side chains. These side chains can belong to amino acids that are widely separated on the linear polypeptide chain, but are brought together when the protein folds (**Figure 4–32**). Other regions on the surface often provide binding sites for different ligands that regulate the protein's activity, as we discuss later. Yet other parts of the protein may be required to attract or attach the protein to a particular location in the cell—for example, the hydrophobic α helix of a

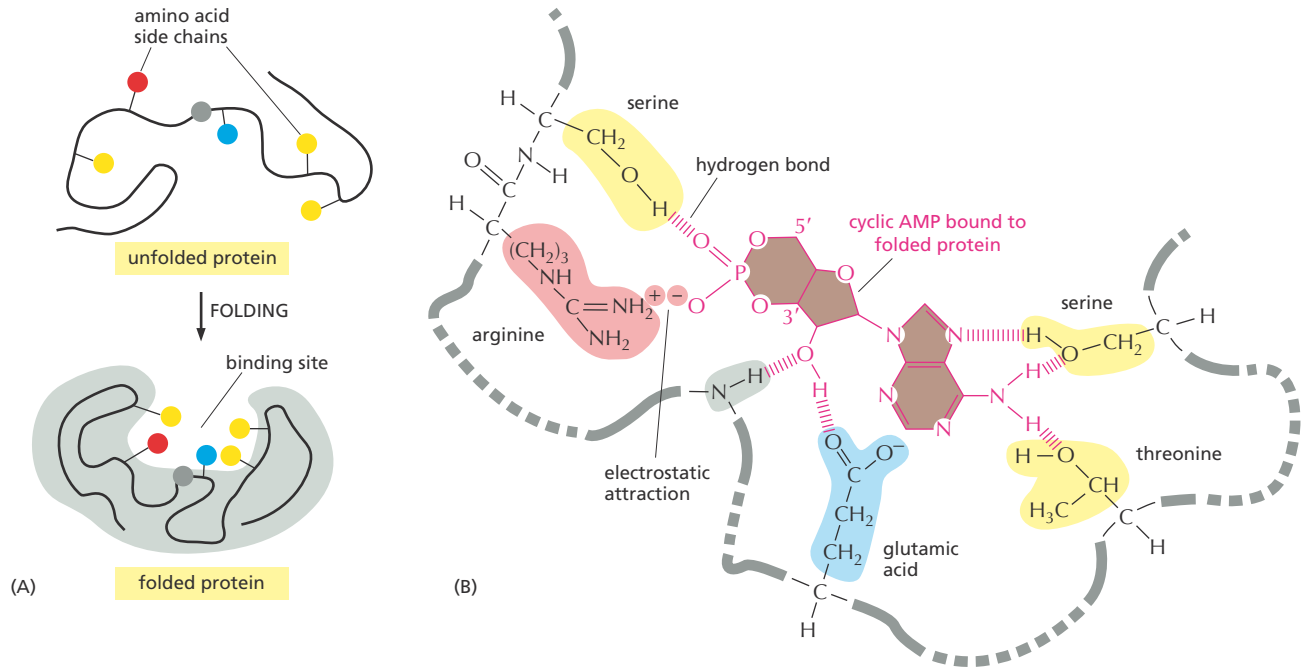


Figure 4–32 Binding sites allow proteins to interact with specific ligands. (A) The folding of the polypeptide chain typically creates a crevice or cavity on the folded protein's surface, where specific amino acid side chains are brought together in such a way that they can form a set of noncovalent bonds only with certain ligands. (B) Close-up view of an actual binding site showing the hydrogen bonds and an electrostatic interaction formed between a protein and its ligand (in this example, the bound ligand is cyclic AMP, shown in *dark brown*).

membrane-spanning protein allows it to be inserted into the lipid bilayer of a cell membrane (discussed in Chapter 11).

Although the atoms buried in the interior of a protein have no direct contact with the ligand, they provide an essential scaffold that gives the surface its contours and chemical properties. Even tiny changes to the amino acids in the interior of a protein can change the protein's three-dimensional shape and destroy its function.

There Are Billions of Different Antibodies, Each with a Different Binding Site

All proteins must bind to particular ligands to carry out their various functions. For antibodies, the universe of possible ligands is limitless. Each of us has the capacity to produce a huge variety of antibodies, among which there will be one that is capable of recognizing and binding tightly to almost any molecule imaginable.

Antibodies are immunoglobulin proteins produced by the immune system in response to foreign molecules, especially those on the surface of an invading microorganism. Each antibody binds to a particular target molecule extremely tightly, either inactivating the target directly or marking it for destruction. An antibody recognizes its target molecule—called an **antigen**—with remarkable specificity, and, because there are potentially billions of different antigens that a person might encounter, we have to be able to produce billions of different antibodies.

Antibodies are Y-shaped molecules with two identical antigen-binding sites, each of which is complementary to a small portion of the surface of the antigen molecule. A detailed examination of the antigen-binding sites of antibodies reveals that they are formed from several loops of polypeptide chain that protrude from the ends of a pair of closely juxtaposed protein domains (**Figure 4–33**). The amino acid sequence in these

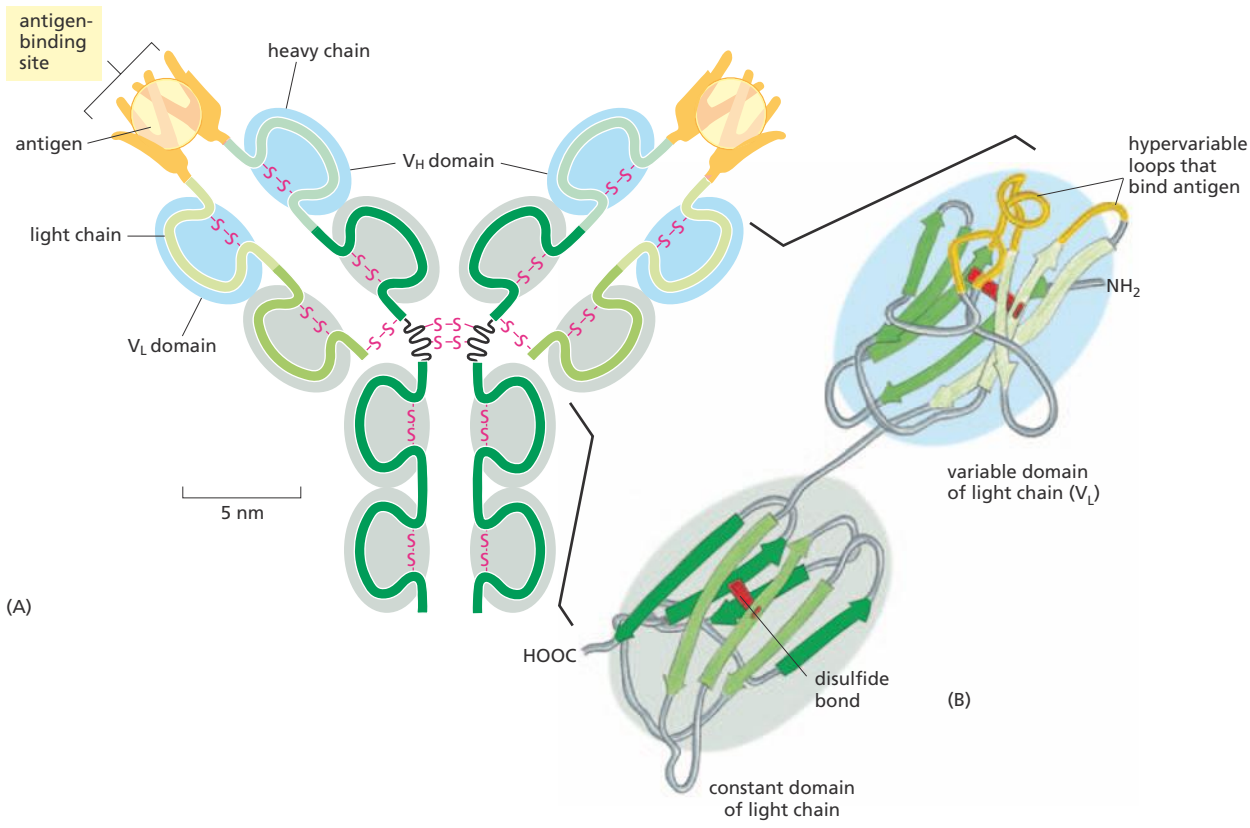


Figure 4–33 An antibody is Y-shaped and has two identical antigen-binding sites, one on each arm of the Y. (A) Schematic drawing of a typical antibody molecule. The protein is composed of four polypeptide chains (two identical heavy chains and two identical and smaller light chains), held together by disulfide bonds (*red*). Each chain is made up of several similar domains, here shaded either *blue* or *gray*. The antigen-binding site is formed where a heavy-chain variable domain (V_H) and a light-chain variable domain (V_L) come close together. These are the domains that differ most in their amino acid sequence in different antibodies—hence their name. (B) Ribbon drawing of a single light chain showing that the most variable parts of the polypeptide chain (*orange*) extend as loops at one end of the variable domain (V_L) to form half of one antigen-binding site of the antibody molecule shown in (A). Note that both the constant and variable domains are composed of a sandwich of two antiparallel β sheets (see also Figure 4–20C), connected by a disulfide bond (*red*).

loops can vary greatly without altering the basic structure of the antibody. An enormous diversity of antigen-binding sites can be generated by changing only the length and amino acid sequence of the loops, which is how the wide variety of different antibodies is formed (**Movie 4.7**).

With their unique combination of specificity and diversity, antibodies are not only indispensable for fighting off infections, they are also invaluable in the laboratory, where they can be used to identify, purify, and study other molecules (**Panel 4–2**, pp. 146–147).

Enzymes Are Powerful and Highly Specific Catalysts

For many proteins, binding to another molecule is their main function. An actin molecule, for example, need only associate with other actin molecules to form a filament. There are proteins, however, for which ligand binding is simply a necessary first step in their function. This is the case for the large and very important class of proteins called **enzymes**. These remarkable molecules are responsible for nearly all of the chemical transformations that occur in cells. Enzymes bind to one or more ligands, called **substrates**, and convert them into chemically modified products, doing this over and over again with amazing rapidity. As we saw in Chapter 3, they speed up reactions, often by a factor of a million or more, without themselves being changed—that is, enzymes act as *catalysts* that permit cells to make or break covalent bonds at will. This catalysis of organized sets of chemical reactions by enzymes creates and maintains the cell, making life possible.

Enzymes can be grouped into functional classes based on the chemical reactions they catalyze (**Table 4–1**). Each type of enzyme is highly specific, catalyzing only a single type of reaction. Thus, *hexokinase* adds a phosphate group to D-glucose but not to its optical isomer L-glucose; the blood-clotting enzyme *thrombin* cuts one type of blood-clotting protein between a particular arginine and its adjacent glycine and nowhere

QUESTION 4–5

Use drawings to explain how an enzyme (such as hexokinase, mentioned in the text) can distinguish its normal substrate (here D-glucose) from the optical isomer L-glucose, which is not a substrate. (Hint: remembering that a carbon atom forms four single bonds that are tetrahedrally arranged and that the optical isomers are mirror images of each other around such a bond, draw the substrate as a simple tetrahedron with four different corners and then draw its mirror image. Using this drawing, indicate why only one optical isomer might bind to a schematic active site of an enzyme.)

TABLE 4–1 SOME COMMON FUNCTIONAL CLASSES OF ENZYMES

Enzyme Class	Biochemical Function
Hydrolase	General term for enzymes that catalyze a hydrolytic cleavage reaction
Nuclease	Breaks down nucleic acids by hydrolyzing bonds between nucleotides
Protease	Breaks down proteins by hydrolyzing peptide bonds between amino acids
Ligase	Joins two molecules together; DNA ligase joins two DNA strands together end-to-end
Isomerase	Catalyzes the rearrangement of bonds within a single molecule
Polymerase	Catalyzes polymerization reactions such as the synthesis of DNA and RNA
Kinase	Catalyzes the addition of phosphate groups to molecules. Protein kinases are an important group of kinases that attach phosphate groups to proteins
Phosphatase	Catalyzes the hydrolytic removal of a phosphate group from a molecule
Oxido-reductase	General name for enzymes that catalyze reactions in which one molecule is oxidized while the other is reduced. Enzymes of this type are often called oxidases, reductases, or dehydrogenases
ATPase	Hydrolyzes ATP. Many proteins have an energy-harnessing ATPase activity as part of their function, including motor proteins such as myosin (discussed in Chapter 17) and membrane transport proteins such as the sodium pump (discussed in Chapter 12)

Enzyme names typically end in "-ase," with the exception of some enzymes, such as pepsin, trypsin, thrombin, lysozyme, and so on, which were discovered and named before the convention became generally accepted at the end of the nineteenth century. The name of an enzyme usually indicates the nature of the reaction catalyzed. For example, citrate synthase catalyzes the synthesis of citrate by a reaction between acetyl CoA and oxaloacetate.

else. As discussed in detail in Chapter 3, enzymes often work in tandem, with the product of one enzyme becoming the substrate for the next. The result is an elaborate network of *metabolic pathways* that provides the cell with energy and generates the many large and small molecules that the cell needs.

Lysozyme Illustrates How an Enzyme Works

To explain how enzymes catalyze chemical reactions, we will use the example of **lysozyme**—an enzyme that acts as a natural antibiotic in egg white, saliva, tears, and other secretions. Lysozyme severs the polysaccharide chains that form the cell walls of bacteria. Because the bacterial cell is under pressure due to intracellular osmotic forces, cutting even a small number of polysaccharide chains causes the cell wall to rupture and the bacterium to burst, or lyse. Lysozyme is a relatively small and stable protein, which can be isolated easily in large quantities. For these reasons it has been intensively studied, and it was the first enzyme whose structure was worked out in atomic detail by X-ray crystallography.

The reaction catalyzed by lysozyme is a hydrolysis: the enzyme adds a molecule of water to a single bond between two adjacent sugar groups in the polysaccharide chain, thereby causing the bond to break. The reaction is energetically favorable because the free energy of the severed polysaccharide chain is lower than the free energy of the intact chain. However, the pure polysaccharide can sit for years in water without being hydrolyzed to any detectable degree. This is because there is an energy barrier to such reactions, called the *activation energy* (discussed in Chapter 3, pp. 91–93). For a colliding water molecule to break a bond linking two sugars, the polysaccharide molecule has to be distorted into a particular shape—the **transition state**—in which the atoms around the bond have an altered geometry and electron distribution. To distort the polysaccharide in this way requires a large input of energy from random molecular collisions. In aqueous solution at room temperature, the energy of such collisions almost never exceeds the activation energy; therefore, hydrolysis occurs extremely slowly, if at all.

This is where the enzyme comes in. Like all enzymes, lysozyme has a binding site on its surface, termed an **active site**, that cradles the contours of its substrate molecule. Here, the catalysis of the chemical reaction occurs. Because its substrate is a polymer, lysozyme's active site is a long groove that holds six linked sugars in the polysaccharide chain at the same time. As soon as the enzyme–substrate complex forms, the enzyme cuts the polysaccharide by catalyzing the addition of a water molecule to one of its sugar–sugar bonds. The severed chain is then quickly released, freeing the enzyme for further cycles of cleavage (**Figure 4–34**).

The chemistry that underlies the binding of lysozyme to its substrate is the same as that for antibody binding to its antigen: the formation of

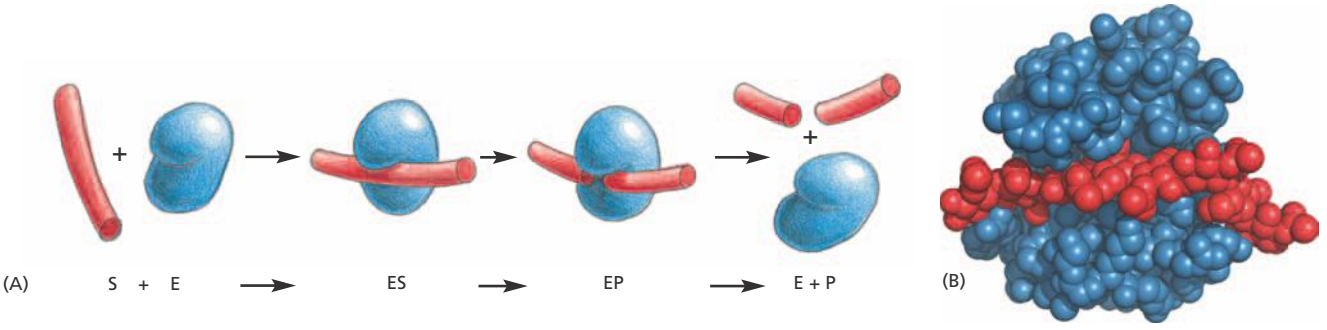
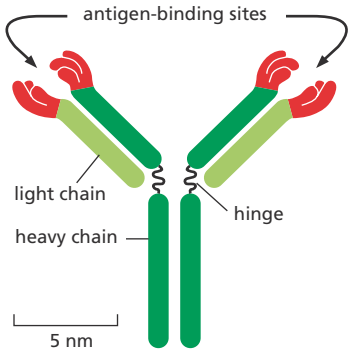


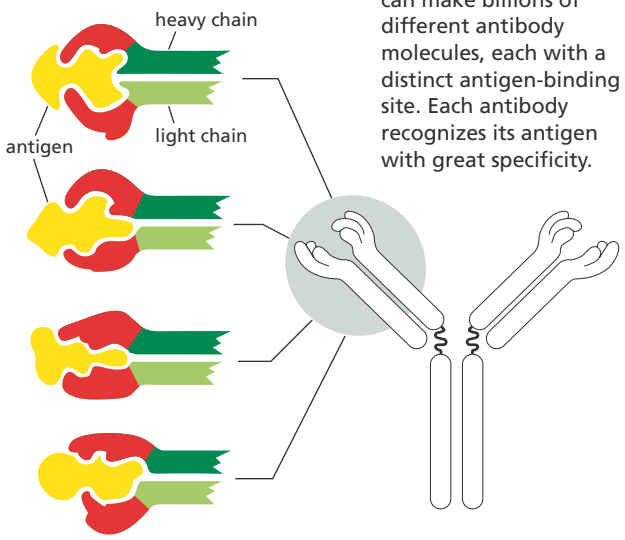
Figure 4–34 Lysozyme cleaves a polysaccharide chain. (A) Schematic view of the enzyme lysozyme (E), which catalyzes the cutting of a polysaccharide substrate molecule (S). The enzyme first binds to the polysaccharide to form an enzyme–substrate complex (ES), then it catalyzes the cleavage of a specific covalent bond in the backbone of the polysaccharide. The resulting enzyme–product complex (EP) rapidly dissociates, releasing the products (P) and leaving the enzyme free to act on another substrate molecule. (B) A space-filling model of lysozyme bound to a short length of polysaccharide chain prior to cleavage. (B, courtesy of Richard J. Feldmann.)

THE ANTIBODY MOLECULE



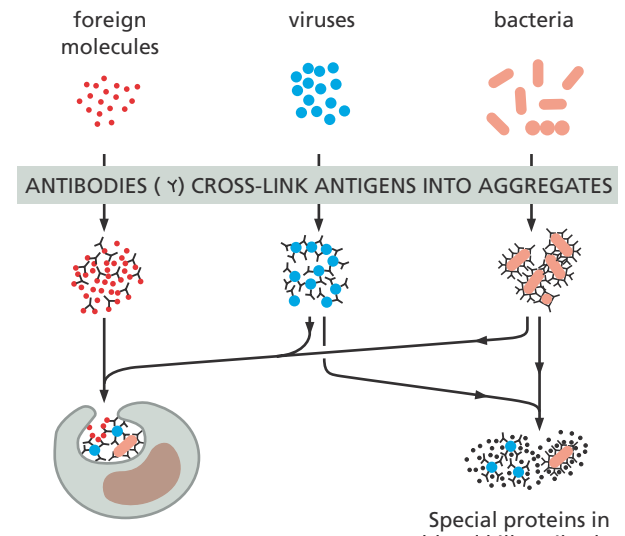
Antibodies are proteins that bind very tightly to their targets (antigens). They are produced in vertebrates as a defense against infection. Each antibody molecule is made of two identical light chains and two identical heavy chains, so the two antigen-binding sites are identical.

ANTIBODY SPECIFICITY



An individual human can make billions of different antibody molecules, each with a distinct antigen-binding site. Each antibody recognizes its antigen with great specificity.

ANTIBODIES DEFEND US AGAINST INFECTION

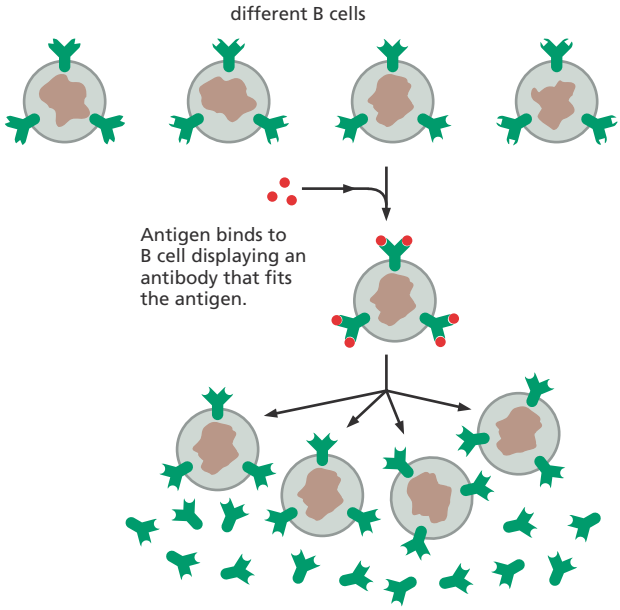


Antibody–antigen aggregates are ingested by phagocytic cells.

Special proteins in blood kill antibody-coated bacteria or viruses.

B CELLS PRODUCE ANTIBODIES

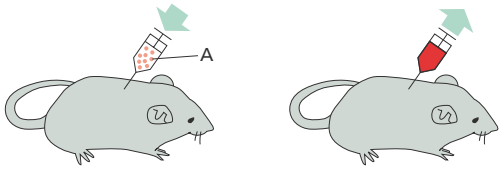
Antibodies are made by a class of white blood cells called B lymphocytes, or B cells. Each resting B cell carries a different membrane-bound antibody molecule on its surface that serves as a receptor for recognizing a specific antigen. When antigen binds to this receptor, the B cell is stimulated to divide and to secrete large amounts of the same antibody in a soluble form.



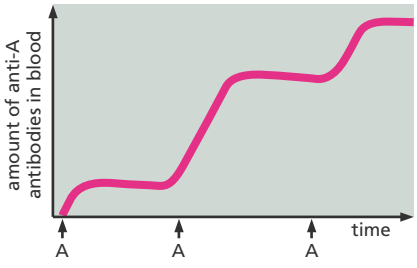
The B cell is stimulated both to proliferate and to make and secrete more of same antibody.

RAISING ANTIBODIES IN ANIMALS

Antibodies can be made in the laboratory by injecting an animal (usually a mouse, rabbit, sheep, or goat) with antigen A.



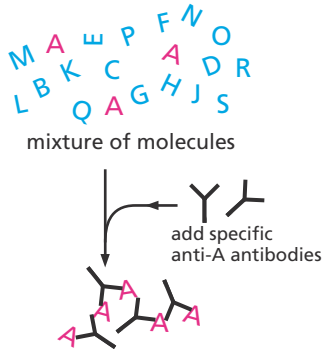
Repeated injections of the same antigen at intervals of several weeks stimulate specific B cells to secrete large amounts of anti-A antibodies into the bloodstream.



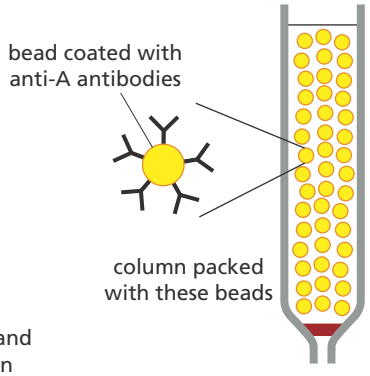
Because many different B cells are stimulated by antigen A, the blood will contain a variety of anti-A antibodies, each of which binds A in a slightly different way.

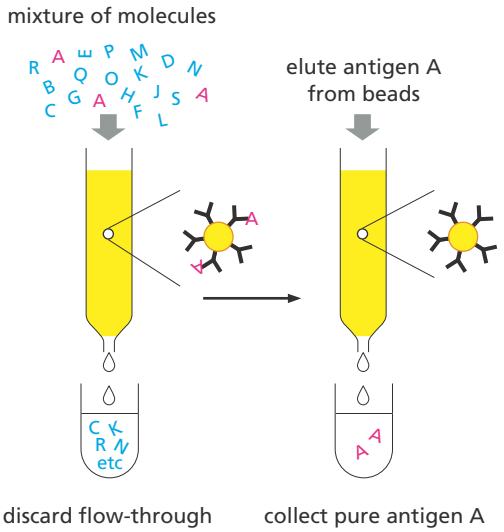
USING ANTIBODIES TO PURIFY MOLECULES

IMMUNOPRECIPITATION



IMMUNOAFFINITY COLUMN CHROMATOGRAPHY



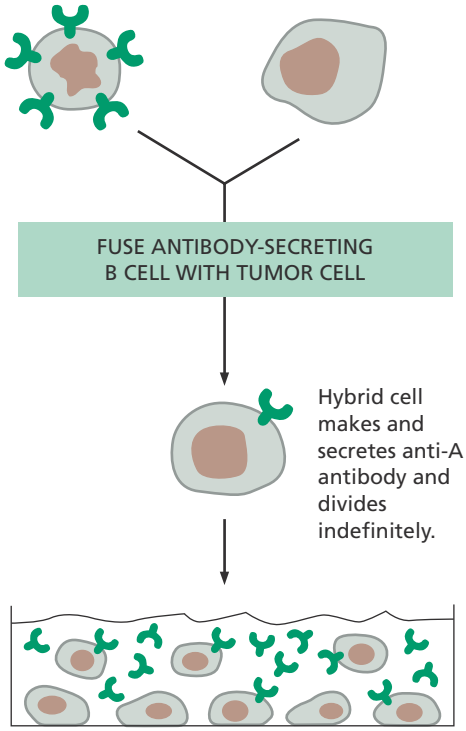


MONOCLONAL ANTIBODIES

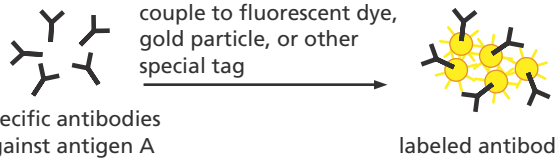
Large quantities of a single type of antibody molecule can be obtained by fusing a B cell (taken from an animal injected with antigen A) with a tumor cell. The resulting hybrid cell divides indefinitely and secretes anti-A antibodies of a single (monoclonal) type.

B cell from animal injected with antigen A makes anti-A antibody but does not divide forever.

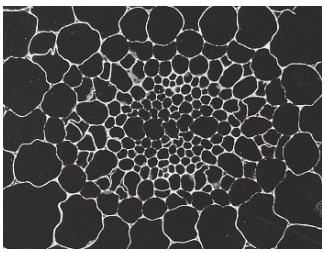
Tumor cells in culture divide indefinitely but do not make antibody.



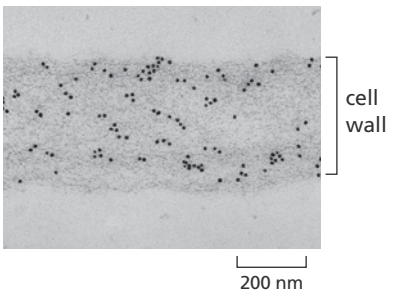
USING ANTIBODIES AS MOLECULAR TAGS



MICROSCOPIC DETECTION

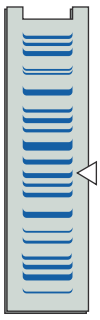


Fluorescent antibody binds to antigen A in tissue and is detected in a fluorescence microscope. The antigen here is pectin in the cell walls of a slice of plant tissue.

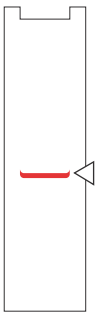


Gold-labeled antibody binds to antigen A in tissue and is detected in an electron microscope. The antigen is pectin in the cell wall of a single plant cell.

BIOCHEMICAL DETECTION

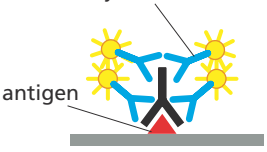


Antigen A is separated from other molecules by electrophoresis.



Incubation with the labeled antibodies that bind to antigen A allows the position of the antigen to be determined.

Labeled second antibody (*blue*) binds to first antibody (*black*).



Note: In all cases, the sensitivity can be greatly increased by using multiple layers of antibodies. This "sandwich" method enables smaller numbers of antigen molecules to be detected.

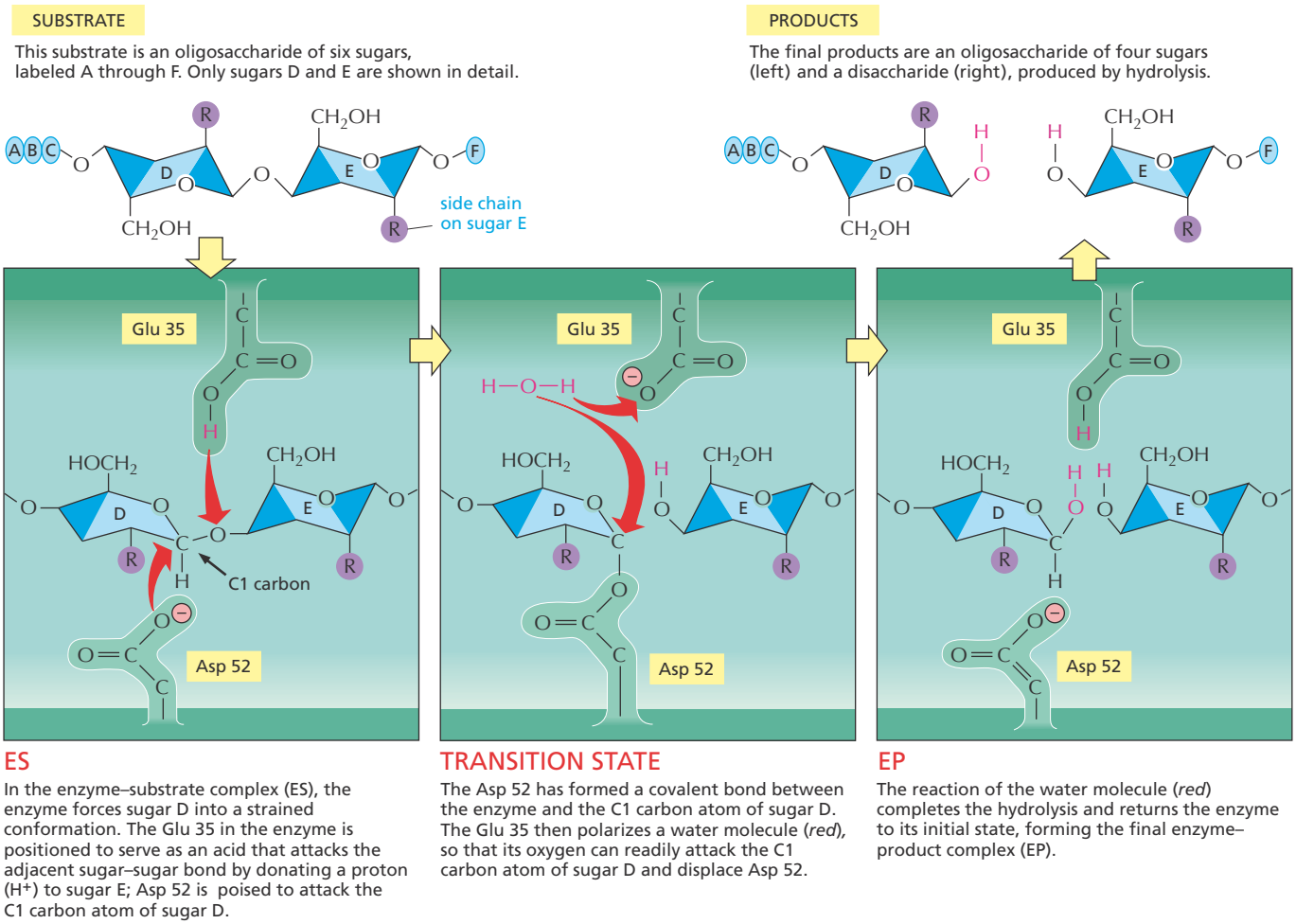
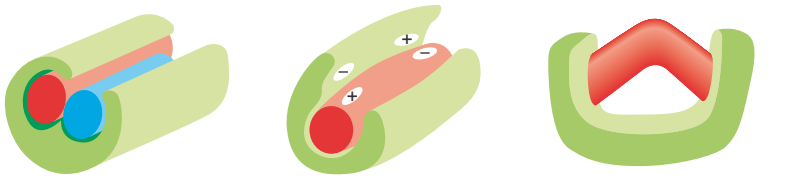


Figure 4–35 Enzymes bind to, and chemically alter, substrate molecules. In the active site of lysozyme, a covalent bond in a polysaccharide molecule is bent and then broken. The top row shows the free substrate and the free products. The three lower panels depict sequential events at the enzyme active site, during which a sugar–sugar covalent bond is broken. Note the change in the conformation of sugar D in the enzyme–substrate complex compared with the free substrate. This conformation favors the formation of the transition state shown in the middle panel, greatly lowering the activation energy required for the reaction. The reaction, and the structure of lysozyme bound to its product, are shown in **Movie 4.8** and **Movie 4.9**. (Based on D.J. Vocadlo et al., *Nature* 412:835–838, 2001.)

multiple noncovalent bonds. However, lysozyme holds its polysaccharide substrate in such a way that one of the two sugars involved in the bond to be broken is distorted from its normal, most stable conformation. The bond to be broken is held close to two specific amino acids with acidic side chains—a glutamic acid and an aspartic acid—located within the active site of the enzyme. Conditions are thereby created in the microenvironment of the lysozyme active site that greatly reduce the activation energy necessary for the hydrolysis to take place (**Figure 4–35**). The overall chemical reaction, from the initial binding of the polysaccharide on the surface of the enzyme to the final release of the severed chains, occurs many millions of times faster than it would in the absence of enzyme.

Other enzymes use similar mechanisms to lower the activation energies and speed up the reactions they catalyze. In reactions involving two or more substrates, the active site also acts like a template or mold that brings the reactants together in the proper orientation for the reaction to occur (**Figure 4–36A**). As we saw for lysozyme, the active site of an enzyme contains precisely positioned chemical groups that speed up the reaction by altering the distribution of electrons in the substrates (**Figure 4–36B**). Binding to the enzyme also changes the shape of the substrate, bending bonds so as to drive the bound molecule toward a particular transition state (**Figure 4–36C**). Finally, like lysozyme, many enzymes participate intimately in the reaction by briefly forming a covalent bond between the substrate and an amino acid side chain in the active site. Subsequent steps in the reaction restore the side chain to its original state, so the enzyme remains unchanged after the reaction and can go on to catalyze many more reactions.



(A) enzyme binds to two substrate molecules and orients them precisely to encourage a reaction to occur between them

(B) binding of substrate to enzyme rearranges electrons in the substrate, creating partial negative and positive charges that favor a reaction

(C) enzyme strains the bound substrate molecule, forcing it toward a transition state to favor a reaction

Figure 4–36 Enzymes can encourage a reaction in several ways. (A) Holding reacting substrates together in a precise alignment. (B) Rearranging the distribution of charge in a reaction intermediate. (C) Altering bond angles in the substrate to increase the rate of a particular reaction.

Many Drugs Inhibit Enzymes

Many of the drugs we take to treat or prevent illness work by blocking the activity of a particular enzyme. Cholesterol-lowering *statins* inhibit HMG-CoA reductase, an enzyme involved in the synthesis of cholesterol by the liver. *Methotrexate* kills some types of cancer cells by shutting down dihydrofolate reductase, an enzyme that produces a compound required for DNA synthesis during cell division. Because cancer cells have lost important intracellular control systems, some of them are unusually sensitive to treatments that interrupt chromosome replication, making them susceptible to methotrexate.

Pharmaceutical companies often develop drugs by first using automated methods to screen massive libraries of compounds to find chemicals that are able to inhibit the activity of an enzyme of interest. They can then chemically modify the most promising compounds to make them even more effective, enhancing their binding affinity and specificity for the target enzyme. As we discuss in Chapter 20, the anticancer drug Gleevec® was designed to specifically inhibit an enzyme whose aberrant behavior is required for the growth of a type of cancer called chronic myeloid leukemia. The drug binds tightly in the substrate-binding pocket of the enzyme, blocking its activity (see Figure 20–56).

Tightly Bound Small Molecules Add Extra Functions to Proteins

Although the order of amino acids in proteins gives these macromolecules their shape and functional versatility, sometimes the amino acids by themselves are not enough for a protein to do its job. Just as we use tools to enhance and extend the capabilities of our hands, so proteins often employ small, nonprotein molecules to perform functions that would be difficult or impossible using amino acids alone. Thus, the photoreceptor protein *rhodopsin*, which is the light-sensitive protein made by the rod cells in the retina, detects light by means of a small molecule, *retinal*, which is attached to the protein by a covalent bond to a lysine side chain (**Figure 4–37A**). Retinal changes its shape when it absorbs a photon of light, and this change is amplified by rhodopsin to trigger a cascade of reactions that eventually leads to an electrical signal being carried to the brain.

Another example of a protein that contains a nonprotein portion essential for its function is *hemoglobin* (see Figure 4–24). A molecule of hemoglobin carries four noncovalently bound *heme* groups, ring-shaped molecules each with a single central iron atom (**Figure 4–37B**). Heme gives hemoglobin (and blood) its red color. By binding reversibly to dissolved oxygen gas through its iron atom, heme enables hemoglobin to pick up oxygen in the lungs and release it in tissues that need it.

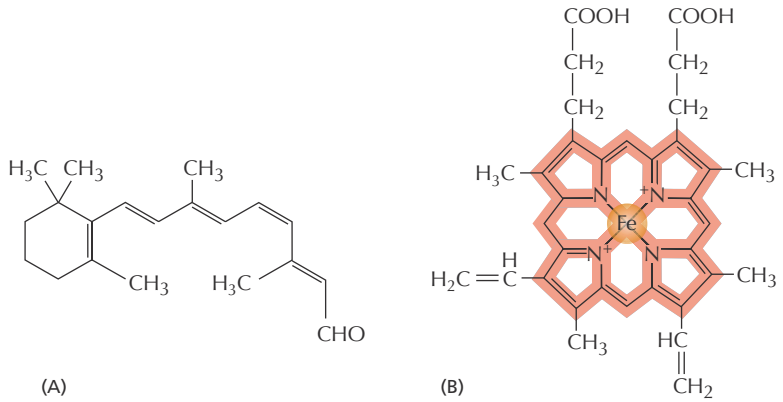


Figure 4–37 Retinal and heme are required for the function of certain proteins. (A) The structure of retinal, the light-sensitive molecule covalently attached to the rhodopsin protein in our eyes. (B) The structure of a heme group, shown with the carbon-containing heme ring colored *red* and the iron atom at its center in *orange*. A heme group is tightly, but noncovalently, bound to each of the four polypeptide chains in hemoglobin, the oxygen-carrying protein whose structure was shown in Figure 4–24.

When these small molecules are attached to their protein, they become an integral part of the protein molecule itself. We discuss in Chapter 11 how proteins can be anchored to cell membranes through covalently attached lipid molecules, and how proteins that are either secreted from the cell or bound to its surface can be modified by the covalent addition of sugars and oligosaccharides.

Enzymes, too, make use of nonprotein molecules: they frequently have a small molecule or metal atom associated with their active site that assists with their catalytic function. *Carboxypeptidase*, an enzyme that cuts polypeptide chains, carries a tightly bound zinc ion in its active site. During the cleavage of a peptide bond by carboxypeptidase, the zinc ion forms a transient bond with one of the substrate atoms, thereby assisting the hydrolysis reaction. In other enzymes, a small organic molecule serves a similar purpose. *Biotin*, for example, is found in enzymes that transfer a carboxyl group ($-COO^-$) from one molecule to another (see Figure 3–37). Biotin participates in these reactions by forming a transient covalent bond to the $-COO^-$ group to be transferred, thereby forming an activated carrier (see Table 3–2, p. 112). This small molecule is better suited for this function than any of the amino acids used to make proteins. Because biotin cannot be synthesized by humans, it must be provided in the diet; thus biotin is classified as a *vitamin*. Other vitamins are similarly needed to make small molecules that are essential components of our proteins; vitamin A, for example, is needed in the diet to make retinal, the light-sensitive part of rhodopsin just discussed.

HOW PROTEINS ARE CONTROLLED

So far, we have examined how proteins do their jobs: how binding to other proteins or small molecules allows them to perform their specific functions. But inside the cell, most proteins and enzymes do not work continuously, or at full speed. Instead, their activities are regulated in a coordinated fashion so the cell can maintain itself in an optimal state, producing only those molecules it requires to thrive under the current conditions. By coordinating when—and how vigorously—proteins function, the cell ensures that it does not deplete its energy reserves by accumulating molecules it does not need or waste its stockpiles of critical substrates. We now consider how cells control the activity of their enzymes and other proteins.

The regulation of protein activity occurs at many levels. At one level, the cell controls the amount of the protein it contains. It can do so by regulating the expression of the gene that encodes that protein (discussed in Chapter 8), and by regulating the rate at which the protein is degraded

(discussed in Chapter 7). At another level, the cell controls enzymatic activities by confining sets of enzymes to particular subcellular compartments, often—but not always—enclosed by distinct membranes (discussed in Chapters 14 and 15). But the most rapid and general mechanism used to adjust the activity of a protein occurs at the level of the protein itself. Although proteins can be switched on or off in various ways, as we see next, all of these mechanisms cause the protein to alter its shape, and therefore its function.

The Catalytic Activities of Enzymes Are Often Regulated by Other Molecules

A living cell contains thousands of different enzymes, many of which are operating at the same time in the same small volume of the cytosol. By their catalytic action, enzymes generate a complex web of metabolic pathways, each composed of chains of chemical reactions in which the product of one enzyme becomes the substrate of the next. In this maze of pathways, there are many branch points where different enzymes compete for the same substrate. The system is so complex that elaborate controls are required to regulate when and how rapidly each reaction occurs.

A common type of control occurs when a molecule other than a substrate specifically binds to an enzyme at a special *regulatory site*, altering the rate at which the enzyme converts its substrate to product. In **feedback inhibition**, for example, an enzyme acting early in a reaction pathway is inhibited by a late product of that pathway. Thus, whenever large quantities of the final product begin to accumulate, the product binds to an earlier enzyme and slows down its catalytic action, limiting further entry of substrates into that reaction pathway (**Figure 4–38**). Where pathways branch or intersect, there are usually multiple points of control by different final products, each of which works to regulate its own synthesis (**Figure 4–39**). Feedback inhibition can work almost instantaneously and is rapidly reversed when product levels fall.

Feedback inhibition is a *negative regulation*: it prevents an enzyme from acting. Enzymes can also be subject to *positive regulation*, in which the enzyme's activity is stimulated by a regulatory molecule rather than being suppressed. Positive regulation occurs when a product in one branch of the metabolic maze stimulates the activity of an enzyme in another pathway.

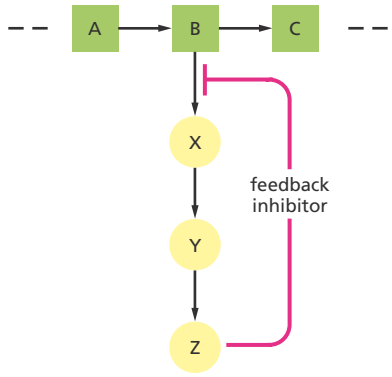


Figure 4–38 Feedback inhibition regulates the flow through biosynthetic pathways. B is the first metabolite in a pathway that gives the end product Z. Z inhibits the first enzyme that is specific to its own synthesis and thereby limits its own concentration in the cell. This form of negative regulation is called feedback inhibition.

Allosteric Enzymes Have Two or More Binding Sites That Influence One Another

One feature of feedback inhibition was initially puzzling to those who discovered it. Unlike what one expects to see for a competitive inhibitor (see Figure 3–29), the regulatory molecule often has a shape that is totally different from the shape of the enzyme's preferred substrate. Indeed, when this form of regulation was discovered in the 1960s, it was termed *allostery* (from the Greek *allo*, "other," and *stere*, "solid" or "shape"). As more was learned about feedback inhibition, researchers realized that many enzymes must have at least two different binding sites on their surface: the active site that recognizes the substrates and one or more sites that recognize regulatory molecules. And that these sites must somehow "communicate" to allow the catalytic events at the active site to be influenced by the binding of the regulatory molecule at its separate site.

The interaction between sites that are located in different regions on a protein molecule is now known to depend on *conformational changes* in the protein: binding of a ligand to one of the sites causes a shift in the protein's structure from one folded shape to a slightly different folded shape,

QUESTION 4–6

Consider the drawing in Figure 4–38. What will happen if, instead of the indicated feedback,
A. Feedback inhibition from Z affects the step B → C only?
B. Feedback inhibition from Z affects the step Y → Z only?
C. Z is a positive regulator of the step B → X?
D. Z is a positive regulator of the step B → C?
For each case, discuss how useful these regulatory schemes would be for a cell.

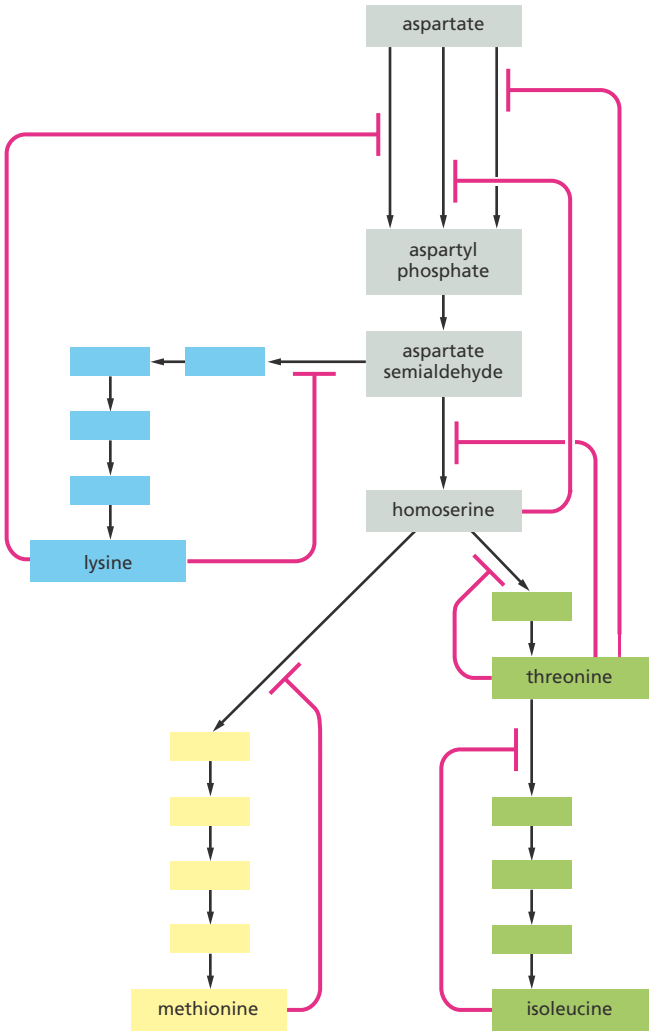


Figure 4–39 Feedback inhibition at multiple points regulates connected metabolic pathways. The biosynthetic pathways for four different amino acids in bacteria are shown, starting from the amino acid aspartate. The *red* lines indicate points at which products feed back to inhibit enzymes and the blank boxes represent intermediates in each pathway. In this example, each amino acid controls the first enzyme specific to its own synthesis, thereby limiting its own concentrations and avoiding a wasteful buildup of intermediates. Some of the products also separately inhibit the initial set of reactions common to all the syntheses. Three different enzymes catalyze the initial reaction from aspartate to aspartyl phosphate, and each of these enzymes is inhibited by a different product.

which alters the binding of a ligand to a second site. Many enzymes have two conformations that differ in activity, each stabilized by the binding of different ligands. During feedback inhibition, for example, the binding of an inhibitor at a regulatory site on the protein causes the protein to shift to a conformation in which its active site—located elsewhere in the protein—becomes less accommodating to the substrate molecule (**Figure 4–40**).

Many—if not most—protein molecules are **allosteric**: they can adopt two or more slightly different conformations, and their activity can be regulated by a shift from one to another. This is true not only for enzymes but also for many other proteins as well. The chemistry involved here is extremely simple in concept: because each protein conformation will have somewhat different contours on its surface, the protein's binding sites for ligands will be altered when the protein changes shape. Each ligand will stabilize the conformation that it binds to most strongly, and at high enough concentrations a ligand will tend to "switch" the population of proteins to the conformation that it favors (**Figure 4–41**).

Phosphorylation Can Control Protein Activity by Causing a Conformational Change

Enzymes are regulated solely by the binding of small molecules. Another method that eukaryotic cells use with great frequency to regulate protein

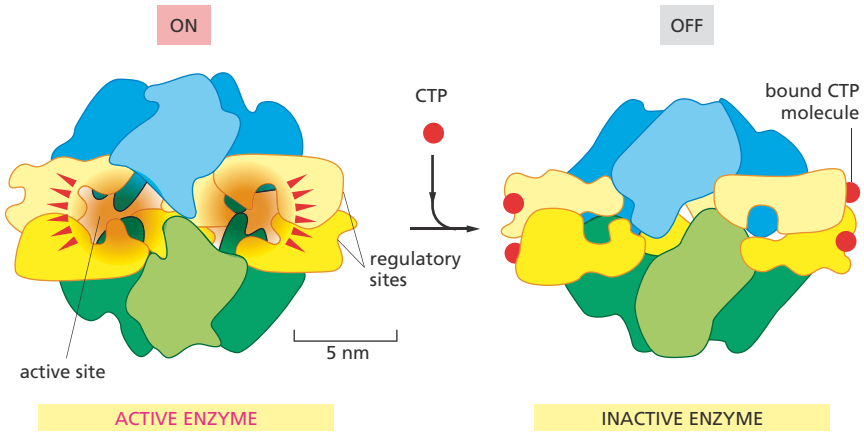


Figure 4–40 Feedback inhibition triggers a conformational change in an enzyme. The enzyme shown, aspartate transcarbamoylase from *E. coli*, was used in early studies of allosteric regulation. This large multisubunit enzyme (see Figure 4–11) catalyzes an important reaction that begins the synthesis of the pyrimidine ring of C, U, and T nucleotides (see Panel 2–6, p. 76–77). One of the final products of this pathway, cytosine triphosphate (CTP), binds to the enzyme to turn it off whenever CTP is plentiful. This diagram shows the conformational change that occurs when the enzyme is turned off by CTP binding to its four regulatory sites, which are distinct from the active site where the substrate binds. Note that the aspartate transcarbamoylase shown in Figure 4–11 is seen from the top. This figure depicts the enzyme as seen from the side.

activity involves attaching a phosphate group covalently to one or more of the protein's amino acid side chains. Because each phosphate group carries two negative charges, the enzyme-catalyzed addition of a phosphate group can cause a major conformational change in a protein by, for example, attracting a cluster of positively charged amino acid side chains from somewhere else in the same protein. This conformational change can, in turn, affect the binding of ligands elsewhere on the protein surface, thereby altering the protein's activity. Removal of the phosphate group by a second enzyme will return the protein to its original conformation and restore its initial activity.

This reversible **protein phosphorylation** controls the activity of many types of proteins in eukaryotic cells; indeed, it is used so extensively that more than one-third of the 10,000 or so proteins in a typical mammalian cell are phosphorylated at any one time. The addition and removal of phosphate groups from specific proteins often occur in response to signals that specify some change in a cell's state. For example, the complicated series of events that takes place as a eukaryotic cell divides is timed largely in this way (discussed in Chapter 18). And many of the intracellular signaling pathways activated by extracellular signals such as hormones depend on a network of protein phosphorylation events (discussed in Chapter 16).

Protein phosphorylation involves the enzyme-catalyzed transfer of the terminal phosphate group of ATP to the hydroxyl group on a serine, threonine, or tyrosine side chain of the protein. This reaction is catalyzed

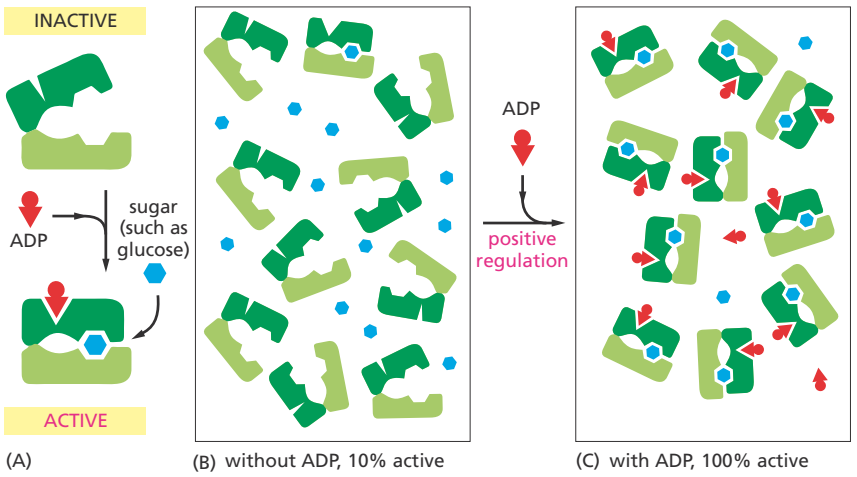


Figure 4–41 The equilibrium between two conformations of a protein is affected by the binding of a regulatory ligand. (A) Schematic diagram of a hypothetical, allosterically regulated enzyme for which a rise in the concentration of ADP molecules (*red* wedges) increases the rate at which the enzyme catalyzes the oxidation of sugar molecules (*blue* hexagons). (B) With no ADP present, only a small fraction of the enzyme molecules spontaneously adopt the active (closed) conformation; most are in the inactive (open) conformation. (C) Because ADP can bind to the protein only in its closed, active conformation, an increase in ADP concentration locks nearly all of the enzyme molecules in the active form. Such an enzyme could be used, for example, to sense when ADP is building up in the cell—which is usually a sign that ATP is decreasing. In this way, the increase in ADP would increase the oxidation of sugars to provide more energy for the synthesis of ATP from ADP—an example of positive regulation.

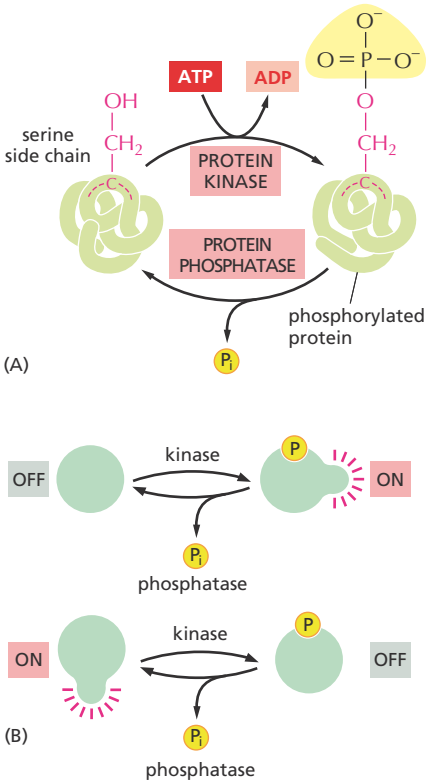


Figure 4–42 Protein phosphorylation is a very common mechanism for regulating protein activity. Many thousands of proteins in a typical eukaryotic cell are modified by the covalent addition of one or more phosphate groups. (A) The general reaction, shown here, entails transfer of a phosphate group from ATP to an amino acid side chain of the target protein by a protein kinase. Removal of the phosphate group is catalyzed by a second enzyme, a protein phosphatase. In this example, the phosphate is added to a serine side chain; in other cases, the phosphate is instead linked to the –OH group of a threonine or tyrosine side chain. (B) Phosphorylation can either increase or decrease the protein's activity, depending on the site of phosphorylation and the structure of the protein.

by a **protein kinase**. The reverse reaction—removal of the phosphate group, or *dephosphorylation*—is catalyzed by a **protein phosphatase** (**Figure 4–42A**). Phosphorylation can either stimulate protein activity or inhibit it, depending on the protein involved and the site of phosphorylation (**Figure 4–42B**). Cells contain hundreds of different protein kinases, each responsible for phosphorylating a different protein or set of proteins. Cells also contain a smaller set of different protein phosphatases; some of these are highly specific and remove phosphate groups from only one or a few proteins, whereas others act on a broad range of proteins. The state of phosphorylation of a protein at any moment in time, and thus its activity, will depend on the relative activities of the protein kinases and phosphatases that act on it.

For many proteins, a phosphate group is added to a particular side chain and then removed in a continuous cycle. Phosphorylation cycles of this kind allow proteins to switch rapidly from one state to another. The more rapidly the cycle is "turning," the faster the concentration of a phosphorylated protein can change in response to a sudden stimulus that increases its rate of phosphorylation. However, keeping the cycle turning costs energy, because one molecule of ATP is hydrolyzed with each turn of the cycle.

Covalent Modifications Also Control the Location and Interaction of Proteins

Phosphorylation can do more than control a protein's activity; it can create docking sites where other proteins can bind, thus promoting the assembly of proteins into larger complexes. For example, when extracellular signals stimulate a class of cell-surface, transmembrane proteins called *receptor tyrosine kinases*, they cause the receptor proteins to phosphorylate themselves on certain tyrosines. The phosphorylated tyrosines then serve as docking sites for the binding and activation of various intracellular signaling proteins, which pass along the message to the cell interior and change the behavior of the cell (see Figure 16–32).

Phosphorylation is not the only form of covalent modification that can affect a protein's activity or location. More than 100 types of covalent modifications can occur in the cell, each playing its own role in regulating protein function. Many proteins are modified by the addition of an acetyl group to a lysine side chain. And the addition of the fatty acid palmitate to a cysteine side chain drives a protein to associate with cell membranes. Attachment of ubiquitin, a 76-amino-acid polypeptide, can target a protein for degradation, as we discuss in Chapter 7. Each of these modifying groups is enzymatically added or removed depending on the needs of the cell.

A large number of proteins are modified on more than one amino acid side chain. The p53 protein, which plays a central part in controling how a cell responds to DNA damage and other stresses, can be modified at 20 sites (**Figure 4–43**). Because an enormous number of combinations of these 20 modifications is possible, the protein's behavior can in principle be altered in a huge number of ways.

The set of covalent modifications that a protein contains at any moment constitutes an important form of regulation. The attachment or removal of these modifying groups controls the behavior of a protein, changing its activity or stability, its binding partners, or its location inside the cell. In some cases, the modification alters the protein's conformation; in others, it serves as a docking site for other proteins to attach. This layer of control enables the cell to make optimal use of its proteins, and it allows the cell to respond rapidly to changes in its environment.

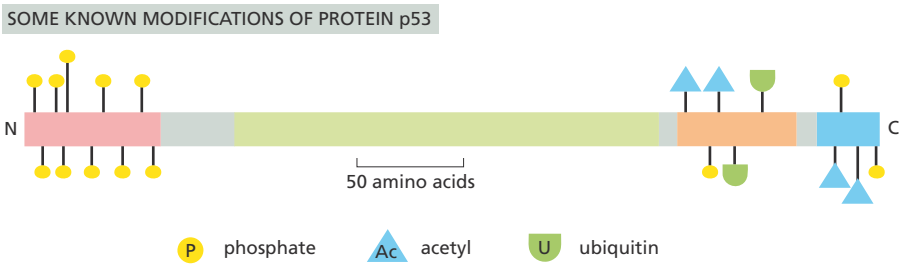


Figure 4–43 The modification of a protein at multiple sites can control the protein's behavior. This diagram shows some of the covalent modifications that control the activity and degradation of the protein p53, an important gene regulatory protein that regulates a cell's response to damage (discussed in Chapter 18). Not all of these modifications will be present at the same time. Colors along the body of the protein represent distinct protein domains, including one that binds to DNA (*green*) and one that activates gene transcription (*pink*). All of the modifications shown are located within relatively unstructured regions of the polypeptide chain.

GTP-Binding Proteins Are Also Regulated by the Cyclic Gain and Loss of a Phosphate Group

Eukaryotic cells have a second way to regulate protein activity by phosphate addition and removal. In this case, however, the phosphate is not enzymatically transferred from ATP to the protein. Instead, the phosphate is part of a guanine nucleotide—guanosine triphosphate (GTP)—that is bound tightly to various types of **GTP-binding proteins**. These proteins act as molecular switches: they are in their active conformation when GTP is bound, but they can hydrolyze this GTP to GDP, which releases a phosphate and flips the protein to an inactive conformation. As with protein phosphorylation, this process is reversible: the active conformation is regained by dissociation of the GDP, followed by the binding of a fresh molecule of GTP (**Figure 4–44**).

A large variety of such GTP-binding proteins function as molecular switches in cells. The dissociation of GDP and its replacement by GTP, which turns the switch on, is often stimulated in response to a signal received by the cell. The GTP-binding proteins in turn bind to other proteins to control their activities; their crucial role in intracellular signaling pathways is discussed in detail in Chapter 16.

ATP Hydrolysis Allows Motor Proteins to Produce Directed Movements in Cells

We have seen how conformational changes in proteins play a central part in enzyme regulation and cell signaling. But conformational changes also play another important role in the operation of the eukaryotic cell: they enable certain specialized proteins to drive directed movements of cells and their components. These **motor proteins** generate the forces responsible for muscle contraction and most other eukaryotic cell movements. They also power the intracellular movements of organelles and macromolecules. For example, they help move chromosomes to opposite ends of the cell during mitosis (discussed in Chapter 18), and they move organelles along cytoskeletal tracks (discussed in Chapter 17).

How are shape changes in proteins used to generate such orderly movements? If, for example, a protein is required to walk along a cytoskeletal fiber, it can move by undergoing a series of conformational changes. However, with nothing to drive these changes in an orderly sequence, the shape changes will be perfectly reversible. Thus the protein can only wander randomly back and forth (**Figure 4–45**).

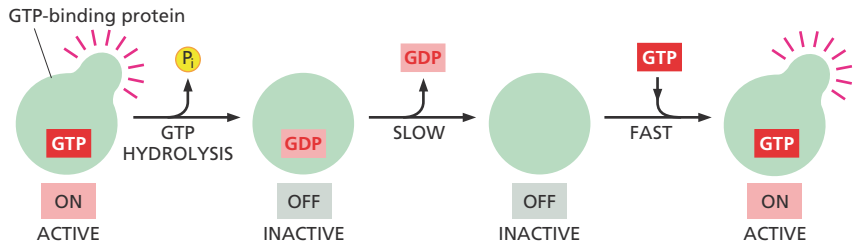


Figure 4–44 GTP-binding proteins function as molecular switches. A GTP-binding protein requires the presence of a tightly bound GTP molecule to be active (*switch ON*). The active protein can shut itself off by hydrolyzing its bound GTP to GDP and inorganic phosphate (P_i), which converts the protein to an inactive conformation (*switch OFF*). To reactivate the protein, the tightly bound GDP must dissociate, a slow step that can be greatly accelerated by specific signals; once the GDP dissociates, a molecule of GTP quickly replaces it, returning the protein to its active conformation.

QUESTION 4–7

Explain how phosphorylation and the binding of a nucleotide (such as ATP or GTP) can both be used to regulate protein activity. What do you suppose are the advantages of either form of regulation?

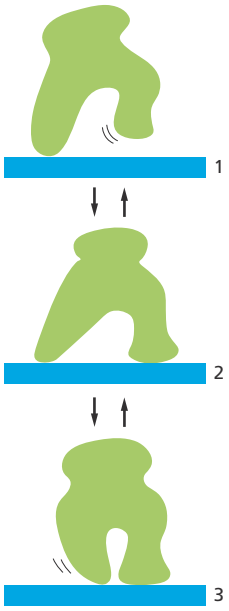


Figure 4–45 Changes in conformation can allow a protein to "walk" along a cytoskeletal filament. This protein's three different conformations allow it to wander randomly back and forth while bound to a filament. Without an input of energy to drive its movement in a single direction, the protein will only shuffle aimlessly, getting nowhere.

To make the conformational changes unidirectional—and force the entire cycle of movement to proceed in one direction—it is enough to make any one of the steps irreversible. For most proteins that are able to move in a single direction for long distances, this irreversibility is achieved by coupling one of the conformational changes to the hydrolysis of an ATP molecule bound to the protein—which is why motor proteins are also ATPases. A great deal of free energy is released when ATP is hydrolyzed, making it very unlikely that the protein will undergo a reverse shape change—as required for moving backward. (Such a reversal would require that the ATP hydrolysis be reversed, by adding a phosphate molecule to ADP to form ATP.) As a consequence, the protein moves steadily forward (**Figure 4–46**).

Many motor proteins generate directional movement by using the hydrolysis of a tightly bound ATP molecule to drive an orderly series of conformational changes. These movements can be rapid: the muscle motor protein *myosin* walks along actin filaments at about 6 μm/sec during muscle contraction (as discussed in Chapter 17).

Proteins Often Form Large Complexes That Function as Protein Machines

As one progresses from small, single-domain proteins to large proteins formed from many domains, the functions that the proteins can perform become more elaborate. The most complex tasks, however, are carried out by large protein assemblies formed from many protein molecules. Now that it is possible to reconstruct biological processes in cell-free systems in a test tube, it is clear that each central process in a cell—including DNA replication, gene transcription, protein synthesis, vesicle budding, and transmembrane signaling—is catalyzed by a highly coordinated, linked set of many proteins. In most such **protein machines**, the hydrolysis of bound nucleoside triphosphates (ATP or GTP) drives an ordered series of conformational changes in some of the individual protein subunits, enabling the ensemble of proteins to move coordinately. In this way, the appropriate enzymes can be positioned to carry out successive reactions in a series—as during the synthesis of proteins on a ribosome, for example (discussed in Chapter 7). Likewise, a large multiprotein complex moves rapidly along DNA to replicate the DNA double helix during cell division (discussed in Chapter 6). A simple mechanical analogy is illustrated in **Figure 4–47**.

Cells have evolved a large number of different protein machines suited to performing a variety of biological tasks. Cells employ protein machines for the same reason that humans have invented mechanical and electronic machines: for almost any job, manipulations that are spatially and temporally coordinated through linked processes are much more efficient than is the sequential use of individual tools.

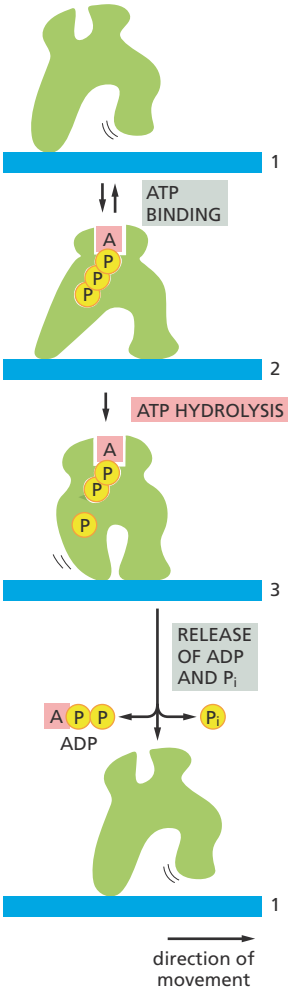


Figure 4–46 A schematic model of how a motor protein uses ATP hydrolysis to move in one direction along a cytoskeletal filament. An orderly transition among three conformations is driven by the hydrolysis of a bound ATP molecule and the release of the products: ADP and inorganic phosphate (P_i). Because these transitions are coupled to the hydrolysis of ATP, the entire cycle is essentially irreversible. Through repeated cycles, the protein moves continuously to the right along the filament. The movement of a single molecule of myosin has been captured by atomic force microscopy.

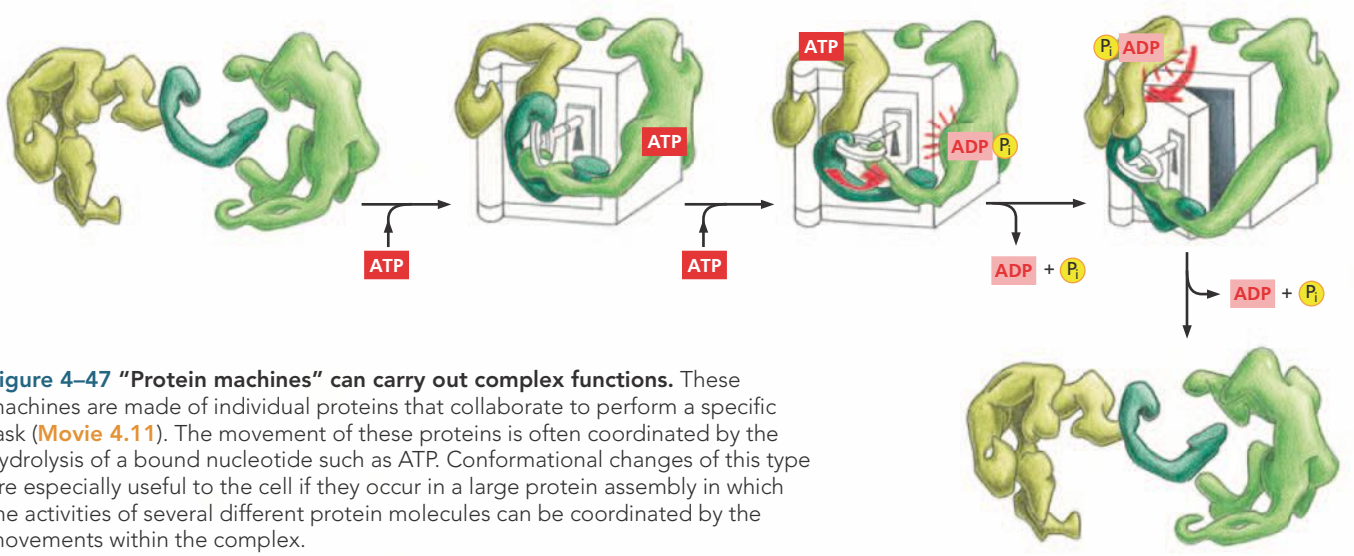


Figure 4–47 "Protein machines" can carry out complex functions. These machines are made of individual proteins that collaborate to perform a specific task (**Movie 4.11**). The movement of these proteins is often coordinated by the hydrolysis of a bound nucleotide such as ATP. Conformational changes of this type are especially useful to the cell if they occur in a large protein assembly in which the activities of several different protein molecules can be coordinated by the movements within the complex.

HOW PROTEINS ARE STUDIED

Understanding how a particular protein functions calls for detailed structural and biochemical analyses—both of which require large amounts of pure protein. But isolating a single type of protein from the thousands of other proteins present in a cell is a formidable task. For many years, proteins had to be purified directly from the source—the tissues in which they are most plentiful. That approach was inconvenient, entailing, for example, early-morning trips to the slaughterhouse. More important, the complexity of intact tissues and organs is a major disadvantage when trying to purify particular molecules, because a long series of chromatography steps is generally required. These procedures not only take weeks to perform, but they also yield only a few milligrams of pure protein.

Nowadays, proteins are more often isolated from cells that are grown in a laboratory (see, for example, Figure 1–38). Often these cells have been "tricked" into making large quantities of a given protein using the genetic engineering techniques that we describe in Chapter 10. Such engineered cells frequently allow large amounts of pure protein to be obtained in only a few days.

In this section, we outline how proteins are extracted and purified from cultured cells and other sources. We describe how these proteins are analyzed to determine their amino acid sequence and their three-dimensional structure. Finally, we discuss how technical advances are allowing proteins to be analyzed, cataloged, manipulated, and even designed from scratch.

Proteins Can be Purified from Cells or Tissues

Whether starting with a piece of liver, a dish of cultured cells, or a vat of bacterial, yeast, or animal cells that have been engineered to produce a protein of interest, the first step in any purification procedure is to break open the cells to release their contents. The resulting slurry is called a *cell homogenate* or *extract.* This physical disruption is followed by an initial fractionation procedure to separate out the class of molecules of interest—for example, all the soluble proteins in the cell (**Panel 4–3**, pp. 164–165).

With this collection of proteins in hand, the job is then to isolate the desired protein. The standard approach involves purifying the protein

QUESTION 4–8

Explain why the hypothetical enzymes in Figure 4–47 have a great advantage in opening the safe if they work together in a protein complex, as opposed to working individually in an unlinked, sequential manner.

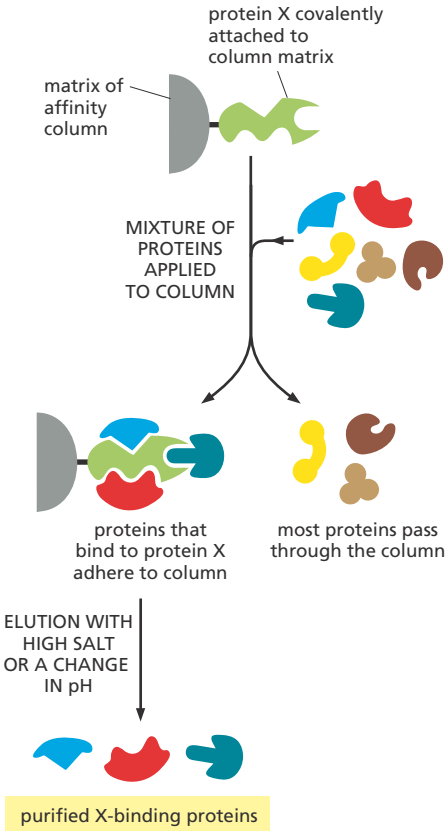


Figure 4–48 Affinity chromatography can be used to isolate the binding partners of a protein of interest. The purified protein of interest (protein X) is covalently attached to the matrix of a chromatography column. An extract containing a mixture of proteins is then loaded onto the column. Those proteins that associate with protein X inside the cell will usually bind to it on the column. Proteins not bound to the column pass right through, and the proteins that are bound tightly to protein X can then be released by changing the pH or ionic composition of the washing solution.

through a series of **chromatography** steps, which use different materials to separate the individual components of a complex mixture into portions, or *fractions*, based on the properties of the protein—such as size, shape, or electrical charge. After each separation step, the fractions are examined to determine which ones contain the protein of interest. These fractions are then pooled and subjected to additional chromatography steps until the desired protein is obtained in pure form.

The most efficient forms of protein chromatography separate polypeptides on the basis of their ability to bind to a particular molecule—a process called *affinity chromatography* (**Panel 4–4**, p. 166). If large amounts of antibodies that recognize the protein are available, for example, they can be attached to the matrix of a chromatography column and used to help extract the protein from a mixture (see Panel 4–2, pp. 146–147).

Affinity chromatography can also be used to isolate proteins that interact physically with the protein being studied. In this case, a purified protein of interest is attached tightly to the column matrix; the proteins that bind to it will remain in the column and can then be removed by changing the composition of the washing solution (**Figure 4–48**).

Proteins can also be separated by **electrophoresis**. In this technique, a mixture of proteins is loaded onto a polymer gel and subjected to an electric field; the polypeptides will then migrate through the gel at different speeds depending on their size and net charge (**Panel 4–5**, p. 167). If too many proteins are present in the sample, or if the proteins are very similar in their migration rate, they can be resolved further using two-dimensional gel electrophoresis (see Panel 4–5). These electrophoretic approaches yield a number of bands or spots that can be visualized by staining; each band or spot contains a different protein. Chromatography and electrophoresis—both developed more than 50 years ago but greatly improved since—have been instrumental in building an understanding of what proteins look like and how they behave (**Table 4–2**). Both techniques are still frequently used in laboratories.

Once a protein has been obtained in pure form, it can be used in biochemical assays to study the details of its activity. It can also be subjected to techniques that reveal its amino acid sequence and precise three-dimensional structure.

Determining a Protein's Structure Begins with Determining Its Amino Acid Sequence

The task of determining the amino acid sequence of a protein can be accomplished in several ways. For many years, sequencing a protein was done by directly analyzing the amino acids in the purified protein. First, the protein was broken down into smaller pieces using a selective protease; the enzyme trypsin, for example, cleaves polypeptide chains on the carboxyl side of a lysine or an arginine. Then the identities of the amino acids in each fragment were determined chemically. The first protein sequenced in this way was the hormone *insulin*, in 1955.

TABLE 4–2 HISTORICAL LANDMARKS IN OUR UNDERSTANDING OF PROTEINS

1838	The name "protein" (from the Greek *proteios*, "primary") was suggested by Berzelius for the complex nitrogen-rich substance found in the cells of all animals and plants.
1819–1904	Most of the 20 common amino acids found in proteins were discovered.
1864	Hoppe-Seyler crystallized, and named, the protein hemoglobin.
1894	Fischer proposed a lock-and-key analogy for enzyme–substrate interactions.
1897	Buchner and Buchner showed that cell-free extracts of yeast can break down sucrose to form carbon dioxide and ethanol, thereby laying the foundations of enzymology.
1926	Sumner crystallized urease in pure form, demonstrating that proteins could possess the catalytic activity of enzymes; Svedberg developed the first analytical ultracentrifuge and used it to estimate the correct molecular weight of hemoglobin.
1933	Tiselius introduced electrophoresis for separating proteins in solution.
1934	Bernal and Crowfoot presented the first detailed X-ray diffraction patterns of a protein, obtained from crystals of the enzyme pepsin.
1942	Martin and Synge developed chromatography, a technique now widely used to separate proteins.
1951	Pauling and Corey proposed the structure of a helical conformation of a chain of amino acids—the α helix—and the structure of the β sheet, both of which were later found in many proteins.
1955	Sanger determined the order of amino acids in insulin, the first protein whose amino acid sequence was determined.
1956	Ingram produced the first protein fingerprints, showing that the difference between sickle-cell hemoglobin and normal hemoglobin is due to a change in a single amino acid (**Movie 4.12**).
1960	Kendrew described the first detailed three-dimensional structure of a protein (sperm whale myoglobin) to a resolution of 0.2 nm, and Perutz proposed a lower-resolution structure for hemoglobin.
1963	Monod, Jacob, and Changeux recognized that many enzymes are regulated through allosteric changes in their conformation.
1966	Phillips described the three-dimensional structure of lysozyme by X-ray crystallography, the first enzyme to be analyzed in atomic detail.
1973	Nomura reconstituted a functional bacterial ribosome from purified components.
1975	Henderson and Unwin determined the first three-dimensional structure of a transmembrane protein (bacteriorhodopsin), using a computer-based reconstruction from electron micrographs.
1976	Neher and Sakmann developed patch-clamp recording to measure the activity of single ion-channel proteins.
1984	Wüthrich used nuclear magnetic resonance (NMR) spectroscopy to solve the three-dimensional structure of a soluble sperm protein.
1988	Tanaka and Fenn separately developed methods for the analysis of proteins and other biological macromolecules.
1996–2013	Mann, Aebersold, Yates, and others developed efficient methods for using mass spectrometry to identify proteins in complex mixtures, exploiting the availability of complete genome sequences.

A much faster way to determine the amino acid sequence of proteins that have been isolated from organisms for which the full genome sequence is known is a method called *mass spectrometry*. This technique determines the exact mass of every peptide fragment in a purified protein, which then allows the protein to be identified from a database that contains a list of every protein thought to be encoded by the genome of the organism in question. Such lists are computed by taking the genome sequence of the organism and applying the genetic code (discussed in Chapter 7).

To perform **mass spectrometry**, the peptides derived from digestion with trypsin are blasted with a laser. This treatment heats the peptides, causing them to become electrically charged (ionized) and ejected in the form of a gas. Accelerated by a powerful electric field, the peptide ions then fly toward a detector; the time it takes them to arrive is related to their mass and their charge. (The larger the peptide is, the more slowly it moves; the

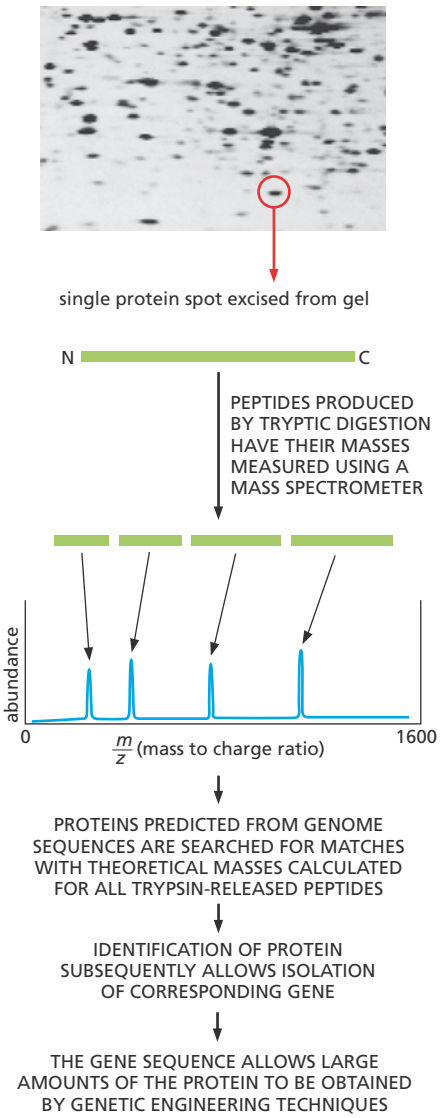


Figure 4–49 Mass spectrometry can be used to identify proteins by determining the precise masses of peptides derived from them. As indicated, this in turn allows the proteins to be produced in the large amounts needed for determining their three-dimensional structure. In this example, the protein of interest is excised from a polyacrylamide gel after two-dimensional electrophoresis (see Panel 4–5, p. 167) and then digested with trypsin. The peptide fragments are loaded into the mass spectrometer, and their exact masses are measured. Genome sequence databases are then searched to find the protein encoded by the organism in question whose profile matches this peptide fingerprint. Mixtures of proteins can also be analyzed in this way. (Image courtesy of Patrick O'Farrell.)

more highly charged it is, the faster it moves.) The set of exact masses of the protein fragments produced by trypsin cleavage then serves as a "fingerprint" that identifies the protein—and its corresponding gene—from publicly accessible databases (**Figure 4–49**).

This approach can even be applied to complex mixtures of proteins, for example, starting with an extract containing all the proteins made by yeast cells grown under a particular set of conditions. To obtain the increased resolution required to distinguish individual proteins, such mixtures are frequently analyzed using *tandem mass spectrometry*. In this case, after the peptides pass through the first mass spectrometer, they are broken into even smaller fragments and analyzed by a second mass spectrometer.

Although all the information required for a polypeptide chain to fold is contained in its amino acid sequence, we have not yet learned how to reliably predict a protein's detailed three-dimensional conformation—the spatial arrangement of its atoms—from its sequence alone. At present, the only way to discover the precise folding pattern of any protein is by experiment, using either **X-ray crystallography** or **nuclear magnetic resonance** (**NMR**) spectroscopy (**How We Know**, pp. 162–163).

Genetic Engineering Techniques Permit the Large-Scale Production, Design, and Analysis of Almost Any Protein

Advances in genetic engineering techniques now permit the production of large quantities of almost any desired protein. In addition to making life much easier for biochemists interested in purifying specific proteins, this ability to churn out huge quantities of a protein has given rise to an entire biotechnology industry (**Figure 4–50**). Bacteria, yeast, and cultured mammalian cells are now used to mass produce a variety of therapeutic proteins, such as insulin, human growth hormone, and even the fertility-enhancing drugs used to boost egg production in women undergoing *in vitro* fertilization. Preparing these proteins previously required the collection and processing of vast amounts of tissue and other biological products—including, in the case of the fertility drugs, the urine of postmenopausal nuns.

The same sorts of genetic engineering techniques can also be employed to produce new proteins and enzymes that contain novel structures or perform unusual tasks: metabolizing toxic wastes, synthesizing life-saving drugs, or operating under conditions that would destroy most biological catalysts (see Chapter 3 How We Know, pp. 104–106). Most of these synthetic catalysts are nowhere near as effective as naturally occurring enzymes in terms of their ability to speed the rate of selected chemical reactions. But, as we continue to learn more about how proteins and enzymes exploit their unique conformations to carry out their biological functions, our ability to make novel proteins with useful functions can only improve.

Of course, to be able to study—or benefit from—the activity of an engineered protein in a living organism, the DNA encoding that protein must somehow be introduced into cells. Again, thanks to genetic engineering techniques, we are able to do just that. We discuss these methods in great detail in Chapter 10.

The Relatedness of Proteins Aids the Prediction of Protein Structure and Function

Biochemists have made enormous progress in understanding the structure and function of proteins over the past 150 years (see Table 4–2, p. 159). These advances are the fruits of decades of painstaking research on isolated proteins, performed by individual scientists working tirelessly on single proteins or protein families, one by one, sometimes for their entire careers. In the future, however, more and more of these investigations of protein conformation and activity will likely take place on a larger scale.

Improvements in our ability to rapidly sequence whole genomes, and the development of methods such as mass spectrometry, have fueled our ability to determine the amino acid sequences of enormous numbers of proteins. Millions of unique protein sequences from thousands of different species have thereby been deposited into publicly available databases, and the collection is expected to double in size every two years. Comparing the amino acid sequences of all of these proteins reveals that the majority belong to protein families that share specific "sequence patterns"—stretches of amino acids that fold into distinct structural domains. In some of these families, the proteins contain only a single structural domain. In others, the proteins include multiple domains arranged in novel combinations (**Figure 4–51**).

Although the number of multidomain families is growing rapidly, the discovery of novel single domains appears to be leveling off. This plateau suggests that the vast majority of proteins may fold up into a limited number of structural domains—perhaps as few as 10,000 to 20,000. For many single-domain families, the structure of at least one family member is known. And knowing the structure of one family member allows us to say something about the structure of its relatives. By this account, we have some structural information for almost three-quarters of the proteins archived in databases (**Movie 4.13**).

A future goal is to acquire the ability to look at a protein's amino acid sequence and be able to deduce its structure and gain insight into its function. We are coming closer to being able to predict protein structure based on sequence information, but there is still a long way to go. Predicting how a protein will function, alone, as part of a complex, or as part of a network in the cell, is much more challenging. But, the closer we get to addressing these questions, the closer we should be to understanding the fundamental basis of life.

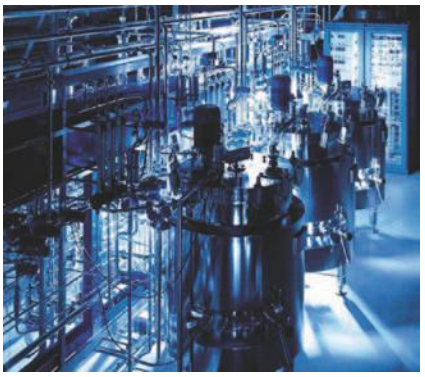
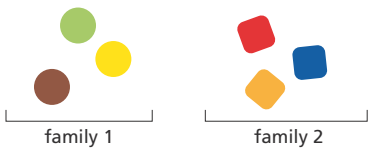
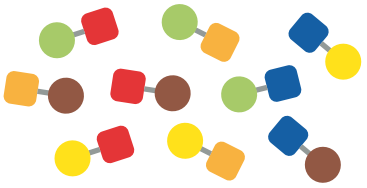


Figure 4–50 Biotechnology companies produce mass quantities of useful proteins. Shown in this photograph are the fermenters used to grow the cells needed for such large-scale protein production. (Courtesy of Bioengineering AG, Switzerland.)



(A) single-domain protein families



(B) a two-domain protein family

Figure 4–51 Most proteins belong to structurally related families. (A) More than two-thirds of all well-studied proteins contain a single structural domain. The members of these single-domain families can have different amino acid sequences but fold into a protein with a similar shape. (B) During evolution, structural domains have been combined in different ways to produce families of multidomain proteins. Almost all novelty in protein structure comes from the way these single domains are arranged. The number of multidomain families being added to the public databases is still rapidly increasing, unlike the number of novel single domains.

PROBING PROTEIN STRUCTURE

As you've no doubt already concluded in reading this chapter, for many proteins, their three-dimensional shape determines their function. So to learn more about how a protein works, it helps to know exactly what it looks like.

The problem is that most proteins are too small to be seen in any detail, even with a powerful electron microscope. To follow the path of an amino acid chain that is folded into a protein molecule, you need to be able to "see" its individual atoms. Scientists use two main methods to map the locations of atoms in a protein. The first involves the use of X-rays. Like light, X-rays are a form of electromagnetic radiation. But they have a wavelength that's much shorter: 0.1 nanometer (nm) as opposed to the 400–700 nm wavelength of visible light. That tiny wavelength—which is the approximate diameter of a hydrogen atom—allows scientists to probe the structure of very small objects at the atomic level.

A second method, called nuclear magnetic resonance (NMR) spectroscopy, takes advantage of the fact that—in many atoms—the nucleus is intrinsically magnetic. When exposed to a large magnet, these nuclei act like tiny bar magnets and align themselves with the magnetic field. If they are then excited with a blast of radio waves, the nuclei will wobble around their magnetic axes, and, as they relax back into the aligned position, they will give off a signal that can be used to reveal their relative positions in a protein.

Using these techniques, investigators have painstakingly pieced together many thousands of protein structures. With the help of computer graphics programs, they have been able to traverse the surfaces and climb inside these proteins, exploring the nooks where ATP likes to nestle, for example, or examining the loops and helices that proteins use to grab hold of a ligand or wrap around a segment of DNA. If the protein happens to belong to a virus or to a cancer cell, seeing its structure can provide clues to designing drugs that might thwart an infection or eliminate a tumor.

X-rays

To determine a protein's structure using X-ray crystallography, you first need to coax the protein into forming crystals: large, highly ordered arrays of the pure protein in which every molecule has the same conformation and is perfectly aligned with its neighbors. Growing high-quality protein crystals is still something of an art and is largely a matter of trial and error. Although robotic methods increase efficiency, it can still take years to find the right conditions—and some proteins resist crystallization altogether.

If you're lucky enough to get good crystals, you are ready for the X-ray analysis. When a narrow beam of X-rays is directed at a protein crystal, the atoms in the protein molecules scatter the incoming X-rays. These scattered waves either reinforce or cancel one another, producing a complex diffraction pattern that is collected by electronic detectors. The position and intensity of each spot in the diffraction pattern contains information about the position of the atoms in the protein crystal (**Figure 4–52**).

Because these patterns are so complex—even a small protein can generate 25,000 discrete spots—computers are used to interpret them and transform them by complex mathematical calculations into maps of the relative spatial positions of the atoms. By combining information obtained from such maps with the amino acid sequence of the protein, you can eventually generate an atomic model of the protein's structure. To determine whether the protein undergoes conformational changes in its structure when it binds a ligand that boosts its activity, you might subsequently try crystallizing it in the presence of its ligand. With crystals of sufficient quality, even small atomic movements can be detected by comparing the structures obtained in the presence and absence of stimulatory or inhibitory ligands.

Magnets

The trouble with X-ray crystallography is that you need crystals. And not all proteins like to form such orderly assemblies. Many have intrinsically disordered regions that wiggle around too much to stack neatly into a crystalline array. Others might not crystallize in the absence of the membranes in which they normally reside.

The other way to solve the structure of a protein does not require protein crystals. If the protein is small—say, 50,000 daltons or less—you can determine its structure by NMR spectroscopy. In this technique, a concentrated solution of pure protein is placed in a strong magnetic field and then bombarded with radio waves of different frequencies. Hydrogen nuclei, in particular, will generate an NMR signal that can be used to determine the distances between these atoms in different parts of the protein. This information is then used to build a model of how the hydrogens are arranged in space. Again, combined with the known amino acid sequence, an NMR spectrum can allow you to compute the three-dimensional structure of the protein (**Figure 4–53**). If the protein is larger than 50,000 daltons, you can try to break it up into its constituent functional domains and analyze each domain by NMR.

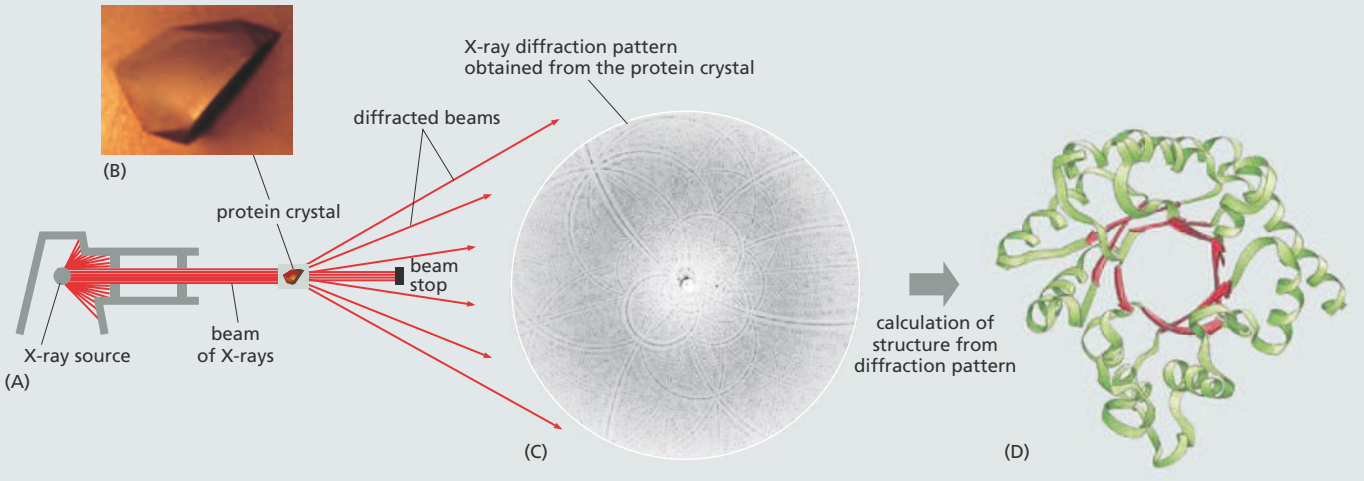


Figure 4–52 The structure of a protein can be determined by X-ray crystallography. Ribulose bisphosphate carboxylase is an enzyme that plays a central role in CO_2 fixation during photosynthesis. (A) X-ray diffraction apparatus; (B) photograph of crystal; (C) diffraction pattern; (D) three-dimensional structure determined from the pattern (α helices are shown in *green*, and β sheets in *red*). (B, courtesy of C. Branden; C, courtesy of J. Hajdu and I. Anderson; D, adapted from original provided by B. Furugren.)

Because determining the precise conformation of a protein is so time-consuming and costly—and the resulting insights so valuable—scientists routinely make their structures freely available by submitting the information to a publicly accessible database. Thanks to such databases, anyone interested in viewing the structure of, say, the ribosome—a complex macromolecular machine made of several RNAs and more than 50 proteins—can easily do so. In the future, improvements in X-ray crystallography and NMR spectroscopy should permit rapid analysis of many more proteins and protein machines. And once enough structures have been determined, it might become possible to generate algorithms for accurately predicting structure solely on the basis of a protein's amino acid sequence. After all, it is the sequence of the amino acids alone that determines how each protein folds up into its three-dimensional shape.

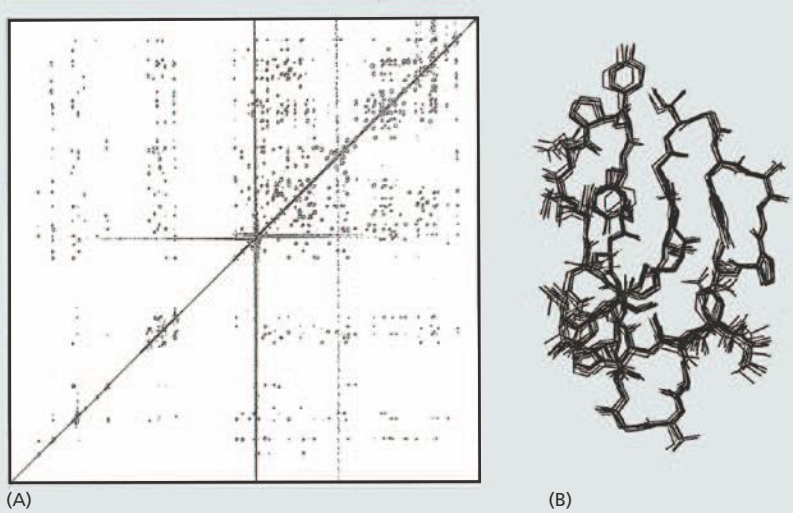


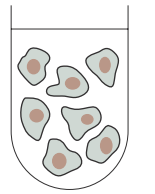
Figure 4–53 NMR spectroscopy can be used to determine the structure of small proteins or protein domains. (A) Two-dimensional NMR spectrum derived from the C-terminal domain of the enzyme cellulase, which breaks down cellulose. The spots represent interactions between neighboring hydrogen atoms. (B) The set of overlapping structures shown all satisfy the distance constraints equally well. (Courtesy of P. Kraulis.)

BREAKING CELLS AND TISSUES

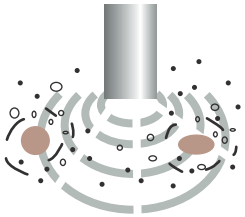
The first step in the purification of most proteins is to disrupt tissues and cells in a controlled fashion.

Using gentle mechanical procedures, called homogenization, the plasma membranes of cells can be ruptured so that the cell contents are released. Four commonly used procedures are shown here.

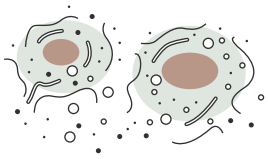
The resulting thick soup (called a homogenate or an extract) contains large and small molecules from the cytosol, such as enzymes, ribosomes, and metabolites, as well as all of the membrane-enclosed organelles.



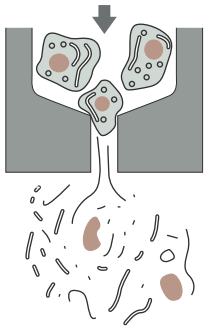
cell suspension or tissue



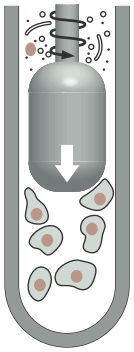
1 Break cells with high-frequency sound (ultrasound).



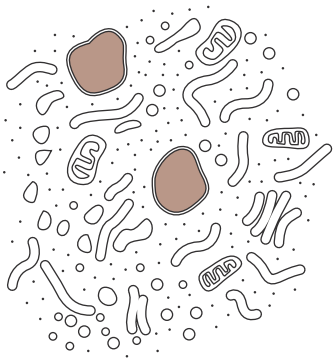
2 Use a mild detergent to make holes in the plasma membrane.



3 Force cells through a small hole using high pressure.

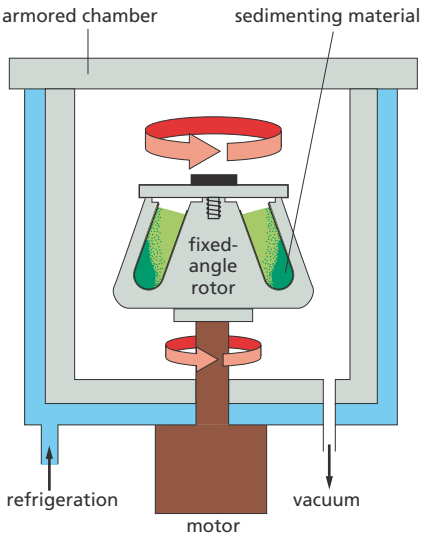


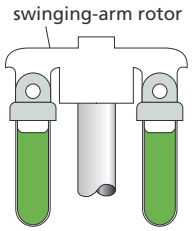
4 Shear cells between a close-fitting rotating plunger and the thick walls of a glass vessel.



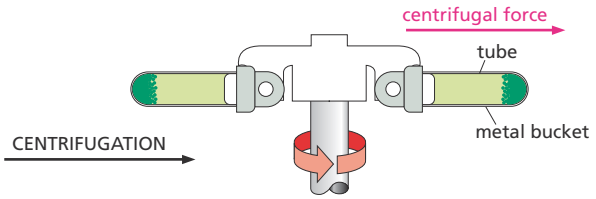
When carefully conducted, homogenization leaves most of the membrane-enclosed organelles largely intact.

THE CENTRIFUGE

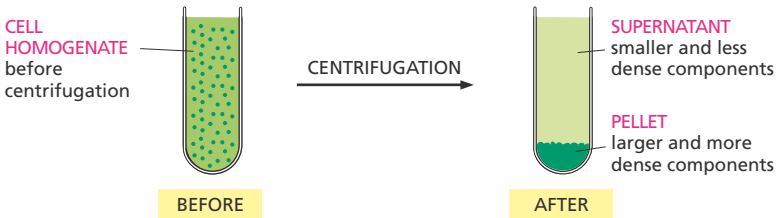




Many cell fractionations are done in a second type of rotor, a swinging-arm rotor.



The metal buckets that hold the tubes are free to swing outward as the rotor turns.



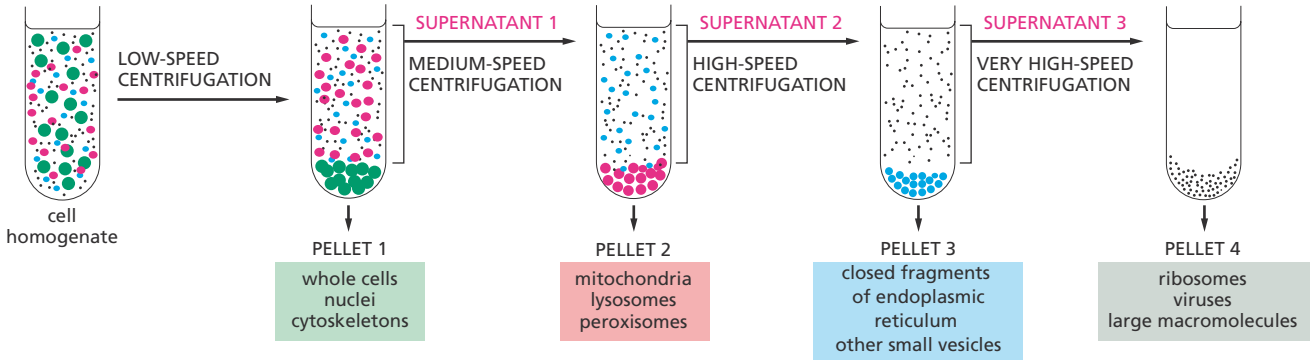
Centrifugation is the most widely used procedure to separate a homogenate into different parts, or fractions. The homogenate is placed in test tubes and rotated at high speed in a centrifuge or ultracentrifuge. Present-day ultracentrifuges rotate at speeds up to 100,000 revolutions per minute and produce enormous forces, as high as 600,000 times gravity.

Such speeds require centrifuge chambers to be refrigerated and have the air evacuated so that friction does not heat up the homogenate. The centrifuge is surrounded by thick armor plating, because an unbalanced rotor can shatter with an explosive release of energy. A fixed-angle rotor can hold larger volumes than a swinging-arm rotor, but the pellet forms less evenly, as shown.

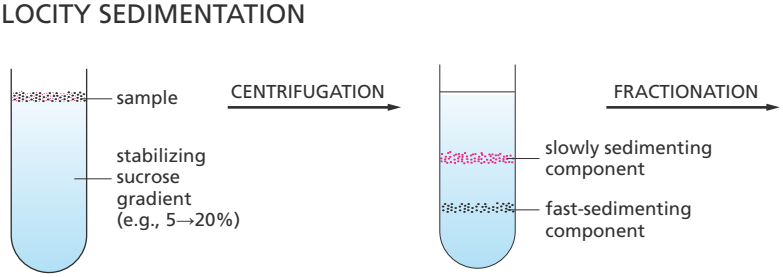
DIFFERENTIAL CENTRIFUGATION

Repeated centrifugation at progressively higher speeds will fractionate cell homogenates into their components.

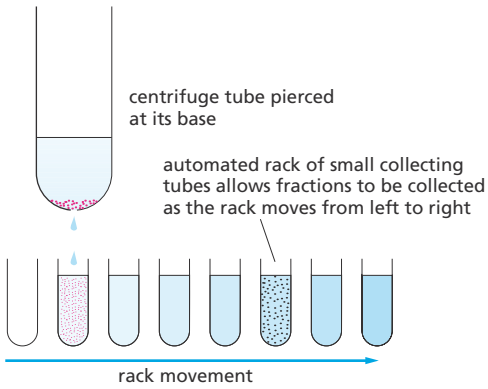
Centrifugation separates cell components on the basis of size and density. The larger and denser components experience the greatest centrifugal force and move most rapidly. They sediment to form a pellet at the bottom of the tube, while smaller, less dense components remain in suspension above, a portion called the supernatant.



VELOCITY SEDIMENTATION



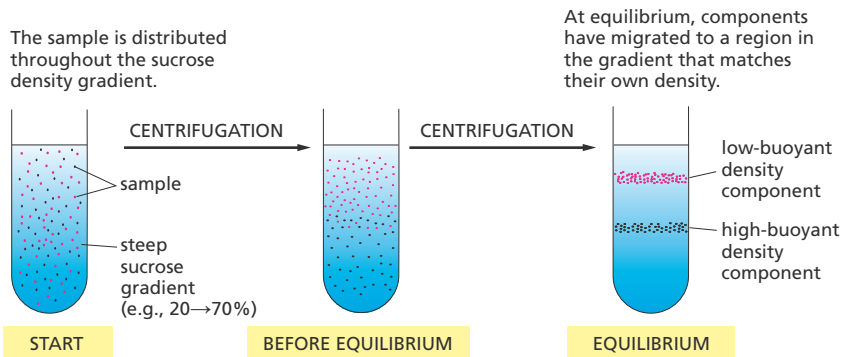
Subcellular components sediment at different rates according to their size after being carefully layered over a dilute salt solution and then centrifuged through it. In order to stabilize the sedimenting components against convective mixing in the tube, the solution contains a continuous shallow gradient of sucrose that increases in concentration toward the bottom of the tube. The gradient is typically 5→20% sucrose. When sedimented through such a dilute sucrose gradient, using a swinging-arm rotor, different cell components separate into distinct bands that can be collected individually.



After an appropriate centrifugation time, the bands may be collected, most simply by puncturing the plastic centrifuge tube and collecting drops from the bottom, as shown here.

EQUILIBRIUM SEDIMENTATION

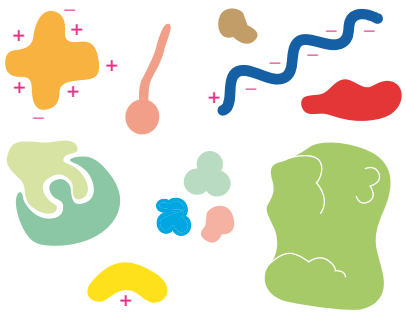
The ultracentrifuge can also be used to separate cell components on the basis of their buoyant density, independently of their size or shape. The sample is usually either layered on top of, or dispersed within, a steep density gradient that contains a very high concentration of sucrose or cesium chloride. Each subcellular component will move up or down when centrifuged until it reaches a position where its density matches its surroundings and then will move no further. A series of distinct bands will eventually be produced, with those nearest the bottom of the tube containing the components of highest buoyant density. The method is also called density gradient centrifugation.



A sucrose gradient is shown here, but denser gradients can be formed with cesium chloride that are particularly useful for separating nucleic acids (DNA and RNA).

The final bands can be collected from the base of the tube, as shown above for velocity sedimentation.

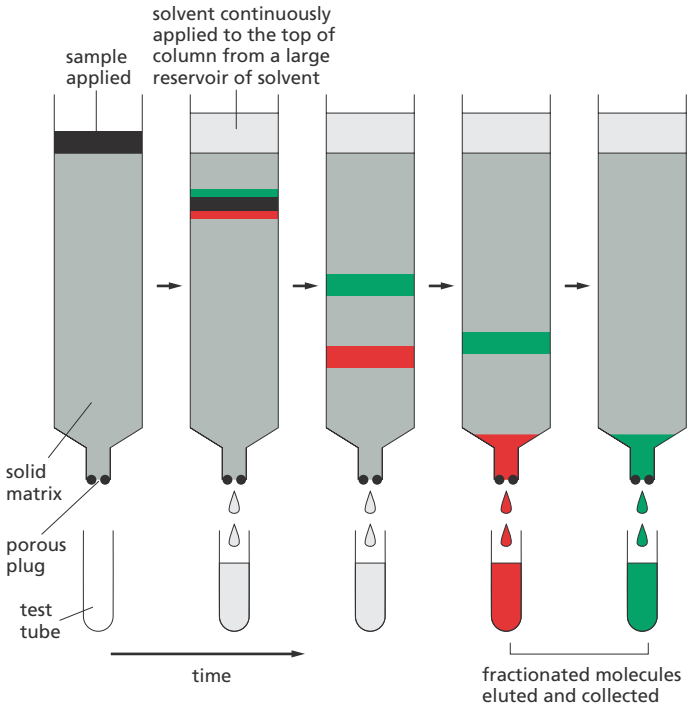
PROTEIN SEPARATION



Proteins are very diverse. They differ in size, shape, charge, hydrophobicity, and their affinity for other molecules. All of these properties can be exploited to separate them from one another so that they can be studied individually.

COLUMN CHROMATOGRAPHY

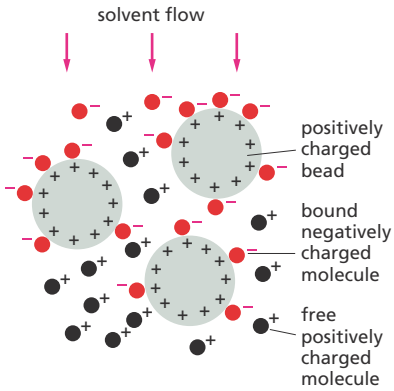
Proteins are often fractionated by column chromatography. A mixture of proteins in solution is applied to the top of a cylindrical column filled with a permeable solid matrix immersed in solvent. A large amount of solvent is then pumped through the column. Because different proteins are retarded to different extents by their interaction with the matrix, they can be collected separately as they flow out from the bottom. According to the choice of matrix, proteins can be separated according to their charge, hydrophobicity, size, or ability to bind to particular chemical groups (see below).



THREE KINDS OF CHROMATOGRAPHY

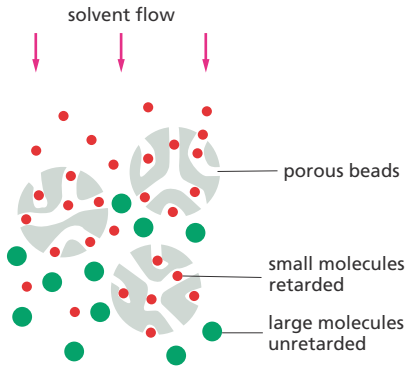
Although the material used to form the matrix for column chromatography varies, it is usually packed in the column in the form of small beads. A typical protein purification strategy might employ in turn each of the three kinds of matrix described below, with a final protein purification of up to 10,000-fold.

Purity can easily be assessed by gel electrophoresis (Panel 4–5).



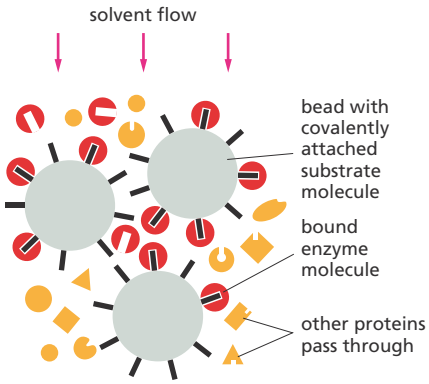
(A) ION-EXCHANGE CHROMATOGRAPHY

Ion-exchange columns are packed with small beads carrying either positive or negative charges that retard proteins of the opposite charge. The association between a protein and the matrix depends on the pH and ionic strength of the solution passing down the column. These can be varied in a controlled way to achieve an effective separation.



(B) GEL-FILTRATION CHROMATOGRAPHY

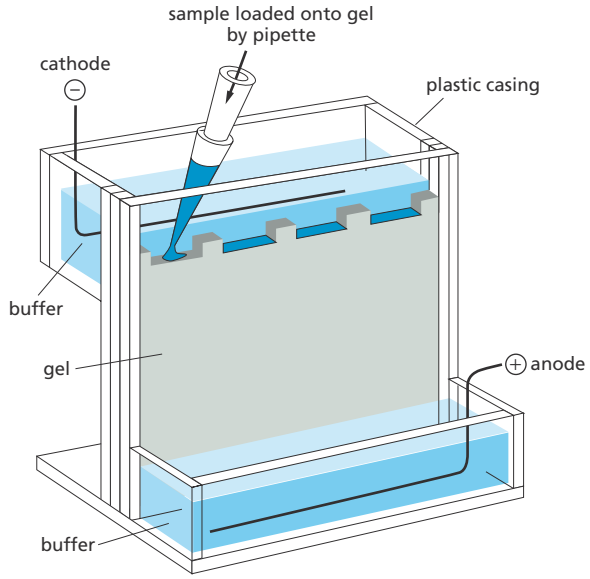
Gel-filtration columns separate proteins according to their size. The matrix consists of tiny porous beads. Protein molecules that are small enough to enter the holes in the beads are delayed and travel more slowly through the column. Proteins that cannot enter the beads are washed out of the column first. Such columns also allow an estimate of protein size.



(C) AFFINITY CHROMATOGRAPHY

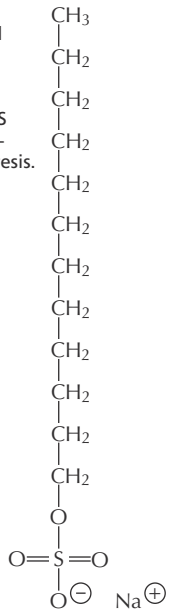
Affinity columns contain a matrix covalently coupled to a molecule that interacts specifically with the protein of interest (e.g., an antibody, or an enzyme substrate). Proteins that bind specifically to such a column can subsequently be released by a pH change or by concentrated salt solutions, and they emerge highly purified (see also Figure 4–48).

GEL ELECTROPHORESIS

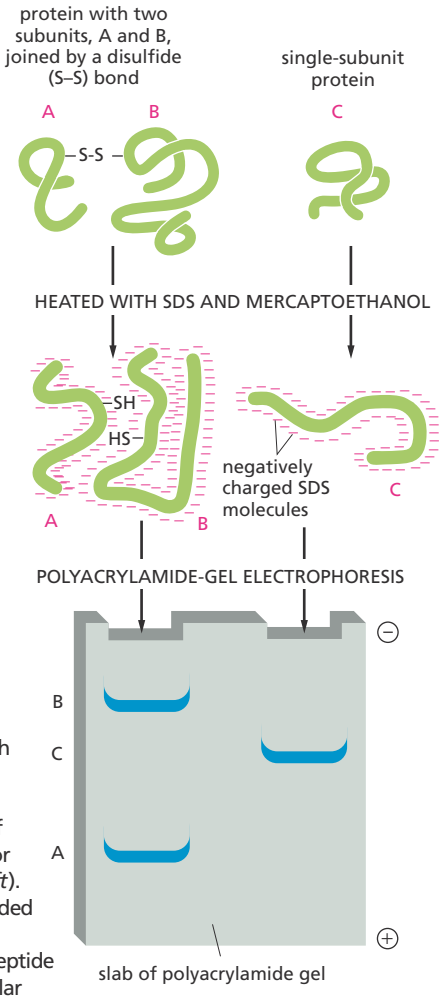


When an electric field is applied to a solution containing protein molecules, the molecules will migrate in a direction and at a speed that reflects their size and net charge. This forms the basis of the technique called electrophoresis.

The detergent sodium dodecyl sulfate (SDS) is used to solubilize proteins for SDS polyacrylamide-gel electrophoresis.



SDS

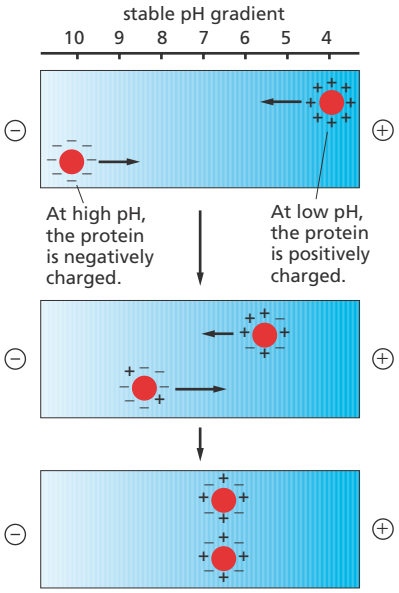


SDS polyacrylamide-gel electrophoresis (SDS-PAGE)

Individual polypeptide chains form a complex with negatively charged molecules of sodium dodecyl sulfate (SDS) and therefore migrate as negatively charged SDS–protein complexes through a slab of porous polyacrylamide gel. The apparatus used for this electrophoresis technique is shown above (*left*). A reducing agent (mercaptoethanol) is usually added to break any –S–S– linkages within or between proteins. Under these conditions, unfolded polypeptide chains migrate at a rate that reflects their molecular weight.

ISOELECTRIC FOCUSING

For any protein there is a characteristic pH, called the isoelectric point, at which the protein has no net charge and therefore will not move in an electric field. In isoelectric focusing, proteins are electrophoresed in a narrow tube of polyacrylamide gel in which a pH gradient is established by a mixture of special buffers. Each protein moves to a point in the pH gradient that corresponds to its isoelectric point and stays there.

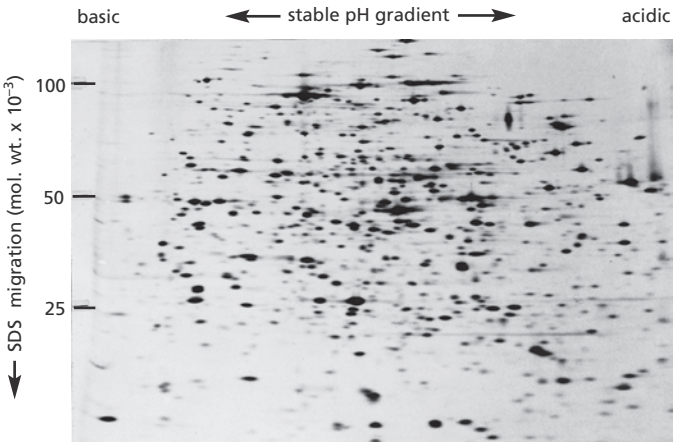


The protein shown here has an isoelectric pH of 6.5.

TWO-DIMENSIONAL POLYACRYLAMIDE-GEL ELECTROPHORESIS

Complex mixtures of proteins cannot be resolved well on one-dimensional gels, but two-dimensional gel electrophoresis, combining two different separation methods, can be used to resolve more than 1000 proteins in a two-dimensional protein map. In the first step, native proteins are separated in a narrow gel on the basis of their intrinsic charge using isoelectric focusing (see *left*). In the second step, this gel is placed on top of a gel slab, and the proteins are subjected to SDS-PAGE (see *above*) in a direction perpendicular to that used in the first step. Each protein migrates to form a discrete spot.

All the proteins in an *E. coli* bacterial cell are separated in this two-dimensional gel, in which each spot corresponds to a different polypeptide chain. They are separated according to their isoelectric point from left to right and to their molecular weight from top to bottom. (Courtesy of Patrick O'Farrell.)



ESSENTIAL CONCEPTS

- Living cells contain an enormously diverse set of protein molecules, each made as a linear chain of amino acids linked together by covalent peptide bonds.
- Each type of protein has a unique amino acid sequence, which determines both its three-dimensional shape and its biological activity.
- The folded structure of a protein is stabilized by multiple noncovalent interactions between different parts of the polypeptide chain.
- Hydrogen bonds between neighboring regions of the polypeptide backbone often give rise to regular folding patterns, known as α helices and β sheets.
- The structure of many proteins can be subdivided into smaller globular regions of compact three-dimensional structure, known as protein domains.
- The biological function of a protein depends on the detailed chemical properties of its surface and how it binds to other molecules, called ligands.
- When a protein catalyzes the formation or breakage of a specific covalent bond in a ligand, the protein is called an enzyme and the ligand is called a substrate.
- At the active site of an enzyme, the amino acid side chains of the folded protein are precisely positioned so that they favor the formation of the high-energy transition states that the substrates must pass through to be converted to product.
- The three-dimensional structure of many proteins has evolved so that the binding of a small ligand can induce a significant change in protein shape.
- Most enzymes are allosteric proteins that can exist in two conformations that differ in catalytic activity, and the enzyme can be turned on or off by ligands that bind to a distinct regulatory site to stabilize either the active or the inactive conformation.
- The activities of most enzymes within the cell are strictly regulated. One of the most common forms of regulation is feedback inhibition, in which an enzyme early in a metabolic pathway is inhibited by the binding of one of the pathway's end products.
- Many thousands of proteins in a typical eukaryotic cell are regulated by cycles of phosphorylation and dephosphorylation.
- GTP-binding proteins also regulate protein function in eukaryotes; they act as molecular switches that are active when GTP is bound and inactive when GDP is bound; turning themselves off by hydrolyzing their bound GTP to GDP.
- Motor proteins produce directed movement in eukaryotic cells through conformational changes linked to the hydrolysis of ATP to ADP.
- Highly efficient protein machines are formed by assemblies of allosteric proteins in which the various conformational changes are coordinated to perform complex functions.
- Covalent modifications added to a protein's amino acid side chains can control the location and function of the protein and can serve as docking sites for other proteins.
- Starting from crude cell or tissue homogenates, individual proteins can be obtained in pure form by using a series of chromatography steps.
- The function of a purified protein can be discovered by biochemical analyses, and its exact three-dimensional structure can be determined by X-ray crystallography or NMR spectroscopy.

KEY TERMS

active site
allosteric
α helix
amino acid sequence
antibody
antigen
β sheet
binding site
C-terminus
chromatography
coiled-coil
conformation
disulfide bond
electrophoresis
enzyme
feedback inhibition
fibrous protein
globular protein
GTP-binding protein
helix
intrinsically disordered sequence
ligand
lysozyme
mass spectrometry
motor protein
N-terminus
nuclear magnetic resonance (NMR) spectroscopy
peptide bond
polypeptide, polypeptide chain
polypeptide backbone
primary structure
protein
protein domain
protein family
protein kinase
protein machine
protein phosphatase
protein phosphorylation
quaternary structure
secondary structure
side chain
substrate
subunit
tertiary structure
transition state
X-ray crystallography

QUESTIONS

QUESTION 4–9

Look at the models of the protein in Figure 4–12. Is the red α helix right- or left-handed? Are the three strands that form the large β sheet parallel or antiparallel? Starting at the N-terminus (the *purple* end), trace your finger along the peptide backbone. Are there any knots? Why, or why not?

QUESTION 4–10

Which of the following statements are correct? Explain your answers.

A. The active site of an enzyme usually occupies only a small fraction of the enzyme surface.

B. Catalysis by some enzymes involves the formation of a covalent bond between an amino acid side chain and a substrate molecule.

C. A β sheet can contain up to five strands, but no more.

D. The specificity of an antibody molecule is contained exclusively in loops on the surface of the folded light-chain domain.

E. The possible linear arrangements of amino acids are so vast that new proteins almost never evolve by alteration of old ones.

F. Allosteric enzymes have two or more binding sites.

G. Noncovalent bonds are too weak to influence the three-dimensional structure of macromolecules.

H. Affinity chromatography separates molecules according to their intrinsic charge.

I. Upon centrifugation of a cell homogenate, smaller organelles experience less friction and thereby sediment faster than larger ones.

QUESTION 4–11

What common feature of α helices and β sheets makes them universal building blocks for proteins?

QUESTION 4–12

Protein structure is determined solely by a protein's amino acid sequence. Should a genetically engineered protein in which the original order of all amino acids is reversed have the same structure as the original protein?

QUESTION 4–13

Consider the following protein sequence as an α helix: Leu-Lys-Arg-Ile-Val-Asp-Ile-Leu-Ser-Arg-Leu-Phe-Lys-Val. How many turns does this helix make? Do you find anything remarkable about the arrangement of the amino acids in this sequence when folded into an α helix? (Hint: consult the properties of the amino acids in Figure 4–3.)

QUESTION 4–14

Simple enzyme reactions often conform to the equation

$$E + S \rightleftharpoons ES \rightarrow EP \rightleftharpoons E + P$$

where E, S, and P are enzyme, substrate, and product, respectively.

A. What does ES represent in this equation?

B. Why is the first step shown with bidirectional arrows and the second step as a unidirectional arrow?

C. Why does E appear at both ends of the equation?

D. One often finds that high concentrations of P inhibit the enzyme. Suggest why this might occur.

E. If compound X resembles S and binds to the active site of the enzyme but cannot undergo the reaction catalyzed by it, what effects would you expect the addition of X to the reaction to have? Compare the effects of X and of the accumulation of P.

QUESTION 4–15

Which of the following amino acids would you expect to find more often near the center of a folded globular protein? Which ones would you expect to find more often exposed to the outside? Explain your answers. Ser, Ser-P (a Ser residue that is phosphorylated), Leu, Lys, Gln, His, Phe, Val, Ile, Met, Cys–S–S–Cys (two cysteines that are disulfide-bonded), and Glu. Where would you expect to find the most N-terminal amino acid and the most C-terminal amino acid?

QUESTION 4–16

Assume you want to make and study fragments of a protein. Would you expect that any fragment of the polypeptide chain would fold the same way as it would in the intact protein? Consider the protein shown in Figure 4–19. Which fragments do you suppose are most likely to fold correctly?

QUESTION 4–17

Neurofilament proteins assemble into long, intermediate filaments (discussed in Chapter 17), found in abundance running along the length of nerve cell axons. The C-terminal region of these proteins is an unstructured polypeptide, hundreds of amino acids long and heavily modified by the addition of phosphate groups. The term "polymer brush" has been applied to this part of the neurofilament. Can you suggest why?

QUESTION 4–18

An enzyme isolated from a mutant bacterium grown at 20°C works in a test tube at 20°C but not at 37°C (37°C is the temperature of the gut, where this bacterium normally lives). Furthermore, once the enzyme has been exposed to the higher temperature, it no longer works at the lower one. The same enzyme isolated from the normal bacterium works at both temperatures. Can you suggest what happens (at the molecular level) to the mutant enzyme as the temperature increases?

QUESTION 4–19

A motor protein moves along protein filaments in the cell. Why are the elements shown in the illustration not sufficient to mediate directed movement (**Figure Q4–19**)? With reference to Figure 4–46, modify the illustration shown here to include other elements that are required to create a unidirectional motor, and justify each modification you make to the illustration.

Figure Q4–19

QUESTION 4–20

Gel-filtration chromatography separates molecules according to their size (see Panel 4–4, p. 166). Smaller molecules diffuse faster in solution than larger ones, yet smaller molecules migrate more slowly through a gel-filtration column than larger ones. Explain this paradox. What should happen at very rapid flow rates?

QUESTION 4–21

As shown in Figure 4–16, both α helices and the coiled-coil structures that can form from them are helical structures, but do they have the same handedness in the figure? Explain why?

QUESTION 4–22

How is it possible for a change in a single amino acid in a protein of 1000 amino acids to destroy its function, even when that amino acid is far away from any ligand-binding site?