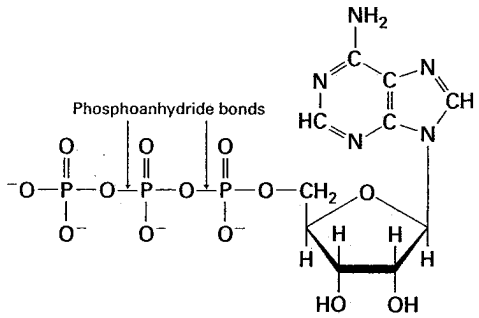


▲ **FIGURE 2-31 Adenosine triphosphate (ATP).** The two phosphoanhydride bonds (red) in ATP, which link the three phosphate groups, each has a ΔG° of about −7.3 kcal/mol for hydrolysis. Hydrolysis of these bonds, especially the terminal one, is the source of energy that drives many energy-requiring reactions in biological systems.

An ATP molecule has two key phosphoanhydride (also called phosphodiester) bonds (Figure 2-31). Hydrolysis of a phosphoanhydride bond (~) in each of the following reactions has a highly negative $\Delta G^{\circ\prime}$ of about −7.3 kcal/mol:

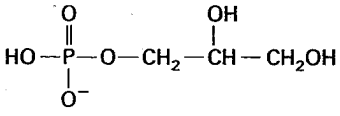
$$\underset{\text{(ATP)}}{\text{Ap~p~p}} + H_2O \longrightarrow \underset{\text{(ADP)}}{\text{Ap~p}} + P_i + H^+$$

$$\underset{\text{(ATP)}}{\text{Ap~p~p}} + H_2O \longrightarrow \underset{\text{(AMP)}}{\text{Ap}} + PP_i + H^+$$

$$\underset{\text{(ADP)}}{\text{Ap~p}} + H_2O \longrightarrow \underset{\text{(AMP)}}{\text{Ap}} + P_i + H^+$$

In these reactions, P_i stands for inorganic phosphate (PO_4^{3-}) and PP_i for inorganic pyrophosphate, two phosphate groups linked by a phosphoanhydride bond. As the top two reactions show, the removal of a phosphate or a pyrophosphate group from ATP leaves adenosine diphosphate (ADP) or adenosine monophosphate (AMP), respectively.

A phosphoanhydride bond or other **high-energy bond** (commonly denoted by ~) is not intrinsically different from other covalent bonds. High-energy bonds simply release especially large amounts of energy when broken by addition of water (hydrolyzed). For instance, the $\Delta G^{\circ\prime}$ for hydrolysis of a phosphoanhydride bond in ATP (−7.3 kcal/mol) is more than three times the $\Delta G^{\circ\prime}$ for hydrolysis of the phosphoester bond (red) in glycerol 3-phosphate (−2.2 kcal/mol):



A principal reason for this difference is that ATP and its hydrolysis products ADP and P_i are highly charged at neutral pH. During synthesis of ATP, a large input of energy is required to force the negative charges in ADP and P_i together. Conversely, much energy is released when ATP is hydrolyzed to ADP and P_i. In comparison, formation of the phosphoester bond between an uncharged hydroxyl in glycerol and P_i requires less energy, and less energy is released when this bond is hydrolyzed.

Cells have evolved protein-mediated mechanisms for transferring the free energy released by hydrolysis of phosphoanhydride bonds to other molecules, thereby driving reactions that would otherwise be energetically unfavorable. For example, if the ΔG for the reaction B + C → D is positive but less than the ΔG for hydrolysis of ATP, the reaction can be driven to the right by coupling it to hydrolysis of the terminal phosphoanhydride bond in ATP. In one common mechanism of such *energy coupling*, some of the energy stored in this phosphoanhydride bond is transferred to one of the reactants by breaking the bond in ATP and forming a covalent bond between the released phosphate group and one of the reactants. The phosphorylated intermediate generated in this way can then react with C to form D + P_i in a reaction that has a negative ΔG:

$$B + ATP \rightarrow B\sim p + ADP$$

$$B\sim p + C \rightarrow D + P_i$$

The overall reaction

$$B + C + ATP \rightleftharpoons D + ADP + P_i$$

is energetically favorable ($\Delta G < 0$).

An alternative mechanism of energy coupling is to use the energy released by ATP hydrolysis to change the conformation of the molecule to an "energy-rich" stressed state. In turn, the energy stored as conformational stress can be released as the molecule "relaxes" back into its unstressed conformation. If this relaxation process can be mechanistically coupled to another reaction, the released energy can be harnessed to drive important cellular processes.

As with many biosynthetic reactions, transport of molecules into or out of the cell often has a positive ΔG and thus requires an input of energy to proceed. Such simple transport reactions do not *directly* involve the making or breaking of covalent bonds; thus the $\Delta G^{\circ\prime}$ is 0. In the case of a substance moving into a cell, Equation 2-7 becomes

$$\Delta G = RT \ln \frac{[C_{in}]}{[C_{out}]} \qquad (2\text{-}10)$$

where $[C_{in}]$ is the initial concentration of the substance inside the cell and $[C_{out}]$ is its concentration outside the cell. We can see from Equation 2-10 that ΔG is positive for transport of a substance into a cell against its concentration gradient (when $[C_{in}] > [C_{out}]$); the energy to drive such "uphill" transport often is supplied by the hydrolysis of

ATP. Conversely, when a substance moves down its concentration gradient ($[C_{out}] > [C_{in}]$), ΔG is negative. Such "downhill" transport releases energy that can be coupled to an energy-requiring reaction, say, the movement of another substance uphill across a membrane or the synthesis of ATP itself (see Chapters 11 and 12).

ATP Is Generated During Photosynthesis and Respiration

Clearly, to continue functioning, cells must constantly replenish their ATP supply. In nearly all cells, the initial energy source whose energy is ultimately transformed into the phosphoanhydride bonds of ATP and bonds in other compounds is sunlight. In **photosynthesis,** plants and certain microorganisms can trap the energy in light and use it to synthesize ATP from ADP and P_i. Much of the ATP produced in photosynthesis is hydrolyzed to provide energy for the conversion of carbon dioxide to six-carbon sugars, a process called **carbon fixation:**

$$6\,CO_2 + 6\,H_2O \xrightarrow{ATP\ \ ADP + P_i} C_6H_{12}O_6 + 6\,O_2$$

In animals, the free energy in sugars and other molecules derived from food is released in the process of respiration. All synthesis of ATP in animal cells and in nonphotosynthetic microorganisms results from the chemical transformation of energy-rich compounds in the diet (e.g., glucose, starch). We discuss the mechanisms of photosynthesis and cellular respiration in Chapter 12.

The complete oxidation of glucose to yield carbon dioxide,

$$C_6H_{12}O_6 + 6\,O_2 \rightarrow 6\,CO_2 + 6\,H_2O$$

has a $\Delta G^{\circ\prime}$ of −686 kcal/mol and is the reverse of photosynthetic carbon fixation. Cells employ an elaborate set of protein-mediated reactions to couple the oxidation of 1 molecule of glucose to the synthesis of as many as 30 molecules of ATP from 30 molecules of ADP. This oxygen-dependent (**aerobic**) degradation (**catabolism**) of glucose is the major pathway for generating ATP in all animal cells, nonphotosynthetic plant cells, and many bacterial cells. Catabolism of fatty acids can also be an important source of ATP.

Light energy captured in photosynthesis is not the only source of chemical energy for all cells. Certain microorganisms that live in or around deep ocean vents, where adequate sunlight is unavailable, derive the energy for converting ADP and P_i into ATP from the oxidation of reduced inorganic compounds. These reduced compounds originate deep in the earth and are released at the vents.

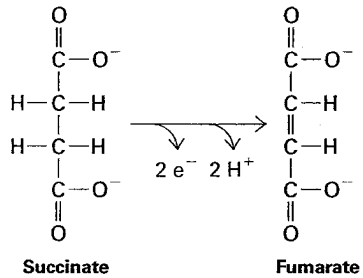
NAD^+ and FAD Couple Many Biological Oxidation and Reduction Reactions

In many chemical reactions, electrons are transferred from one atom or molecule to another; this transfer may or may not accompany the formation of new chemical bonds or the release of energy that can be coupled to other reactions. The loss of electrons from an atom or a molecule is called **oxidation,** and the gain of electrons by an atom or a molecule is called **reduction.** Because electrons are neither created nor destroyed in a chemical reaction, if one atom or molecule is oxidized, another must be reduced. For example, oxygen draws electrons from Fe^{2+} (ferrous) ions to form Fe^{3+} (ferric) ions, a reaction that occurs as part of the process by which carbohydrates are degraded in mitochondria. Each oxygen atom receives two electrons, one from each of two Fe^{2+} ions:

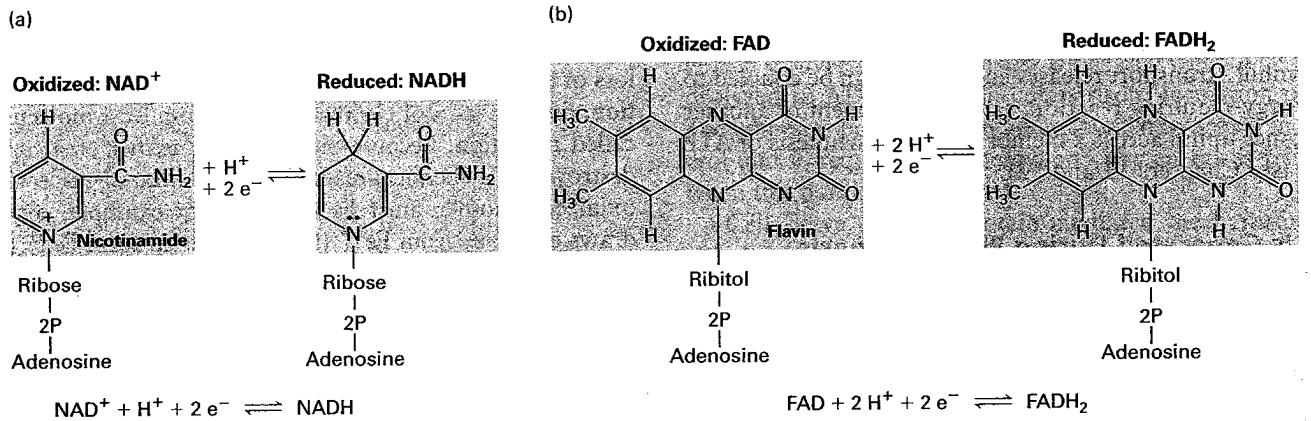
$$2\,Fe^{2+} + \tfrac{1}{2}\,O_2 \rightarrow 2\,Fe^{3+} + O^{2-}$$

Thus Fe^{2+} is oxidized, and O_2 is reduced. Such reactions in which one molecule is reduced and another oxidized often are referred to as **redox reactions.** Oxygen is an electron acceptor in many redox reactions in cells under aerobic conditions.

Many biologically important oxidation and reduction reactions involve the removal or the addition of hydrogen atoms (protons plus electrons) rather than the transfer of isolated electrons on their own. The oxidation of succinate to fumarate, which also occurs in mitochondria, is an example (Figure 2-32). Protons are soluble in aqueous solutions (as H_3O^+), but electrons are not and must be transferred directly from one atom or molecule to another without a water-dissolved intermediate. In this type of oxidation reaction, electrons often are transferred to small electron-carrying molecules, sometimes referred to as coenzymes. The most common of these electron carriers are **NAD^+ (nicotinamide adenine dinucleotide),** which is reduced to NADH, and **FAD (flavin adenine dinucleotide),** which is reduced to $FADH_2$ (Figure 2-33). The reduced forms of these coenzymes can transfer protons and electrons to other molecules, thereby reducing them.



▲ **FIGURE 2-32 Conversion of succinate to fumarate.** In this oxidation reaction, which occurs in mitochondria as part of the citric acid cycle, succinate loses two electrons and two protons. These are transferred to FAD, reducing it to $FADH_2$.



$NAD^+ + H^+ + 2\,e^- \rightleftharpoons NADH$

$FAD + 2\,H^+ + 2\,e^- \rightleftharpoons FADH_2$

▲ FIGURE 2-33 The electron-carrying coenzymes NAD^+ and FAD. (a) NAD^+ (nicotinamide adenine dinucleotide) is reduced to NADH by the addition of two electrons and one proton simultaneously. In many biological redox reactions, a pair of hydrogen atoms (two protons and two electrons) are removed from a molecule. In some cases, one of the protons and both electrons are transferred to NAD^+; the other proton is released into solution. (b) FAD (flavin adenine dinucleotide) is reduced to $FADH_2$ by the addition of two electrons and two protons, as occurs when succinate is converted to fumarate (see Figure 2-32). In this two-step reaction, addition of one electron together with one proton first generates a short-lived semiquinone intermediate (not shown), which then accepts a second electron and proton.

To describe redox reactions, such as the reaction of ferrous ion (Fe^{2+}) and oxygen (O_2), it is easiest to divide them into two half-reactions:

$$\textit{Oxidation of } Fe^{2+}\text{:}\quad 2\,Fe^{2+} \rightarrow 2\,Fe^{3+} + 2\,e^-$$
$$\textit{Reduction of } O_2\text{:}\quad 2\,e^- + \tfrac{1}{2}\,O_2 \rightarrow O^{2-}$$

In this case, the reduced oxygen (O^{2-}) readily reacts with two protons to form one water molecule (H_2O). The readiness with which an atom or a molecule *gains* an electron is its **reduction potential** E. The tendency to *lose* electrons, the **oxidation potential,** has the same magnitude but opposite sign as the reduction potential for the reverse reaction.

Reduction potentials are measured in volts (V) from an arbitrary zero point set at the reduction potential of the following half-reaction under standard conditions (25 °C, 1 atm, and reactants at 1 M):

$$H^+ + e^- \underset{\text{oxidation}}{\overset{\text{reduction}}{\rightleftharpoons}} \tfrac{1}{2}\,H_2$$

The value of E for a molecule or an atom under standard conditions is its standard reduction potential, E'_0. A molecule or an ion with a positive E'_0 has a higher affinity for electrons than the H^+ ion does under standard conditions. Conversely, a molecule or ion with a negative E'_0 has a lower affinity for electrons than the H^+ ion does under standard conditions. Like the values of $\Delta G^{\circ\prime}$, standard reduction potentials may differ somewhat from those found under the conditions in a cell because the concentrations of reactants in a cell are not 1 M.

In a redox reaction, electrons move spontaneously toward atoms or molecules having *more positive* reduction potentials. In other words, a compound having a more negative reduction potential can transfer electrons spontaneously to (i.e., reduce) a compound with a more positive reduction potential. In this type of reaction, the change in electric potential ΔE is the sum of the reduction and oxidation potentials for the two half-reactions. The ΔE for a redox reaction is related to the change in free energy ΔG by the following expression:

$$\Delta G\ (\text{cal/mol}) = -n\ (23{,}064)\ \Delta E\ (\text{volts}) \qquad (2\text{-}11)$$

where n is the number of electrons transferred. Note that a redox reaction with a positive ΔE value will have a negative ΔG and thus will tend to proceed spontaneously from left to right.

KEY CONCEPTS OF SECTION 2.4

Biochemical Energetics

- The change in free energy ΔG is the most useful measure for predicting the direction of chemical reactions in biological systems. Chemical reactions tend to proceed spontaneously in the direction for which ΔG is negative. The magnitude of ΔG is independent of the reaction rate.
- The chemical free-energy change $\Delta G^{\circ\prime}$ equals $-2.3\ RT \log K_{eq}$. Thus the value of $\Delta G^{\circ\prime}$ can be calculated from the experimentally determined concentrations of reactants and products at equilibrium.
- The rate of a reaction depends on the activation energy needed to energize reactants to a transition state. Catalysts such as enzymes speed up reactions by lowering the activation energy of the transition state.
- A chemical reaction having a positive ΔG can proceed if it is coupled with a reaction having a negative ΔG of larger magnitude.
- Many otherwise energetically unfavorable cellular processes are driven by the hydrolysis of phosphoanhydride bonds in ATP (see Figure 2-31).

- Directly or indirectly, light energy captured by photosynthesis in plants and photosynthetic bacteria is the ultimate source of chemical energy for almost all cells.
- An oxidation reaction (loss of electrons) is always coupled with a reduction reaction (gain of electrons).
- Biological oxidation and reduction reactions often are coupled by electron-carrying coenzymes such as NAD^+ and FAD (see Figure 2-33).
- Oxidation-reduction reactions with a positive ΔE have a negative ΔG and thus tend to proceed spontaneously.

Key Terms

acid *52*
α carbon atom (C_α) *41*
amino acids *41*
amphipathic *31*
base *52*
buffers *52*
chemical potential energy *54*
covalent bond *32*
dehydration reaction *40*
ΔG (free-energy change) *55*
disulfide bond *43*
endergonic *55*
endothermic *56*
energy coupling *58*
enthalpy (H) *55*
entropy (S) *55*
equilibrium constant *49*
exergonic *55*
exothermic *56*
fatty acids *47*
hydrogen bond *37*
hydrophilic *31*
hydrophobic *31*
hydrophobic effect *38*
ionic interactions *36*
molecular complementarity *39*
monosaccharides *44*
nucleosides *44*
nucleotides *44*
oxidation *59*
pH *51*
phosphoanhydride bonds *57*
phospholipid bilayers *41*
polar *34*
polymer *40*
redox reaction *59*
reduction *59*
saturated *47*
steady state *50*
stereoisomers *33*
unsaturated *47*
van der Waals interactions *37*

Review the Concepts

1. The gecko is a reptile with an amazing ability to climb smooth surfaces, including glass. Recent discoveries indicate that geckos stick to smooth surfaces via van der Waals interactions between septae on their feet and the smooth surface. How is this method of stickiness advantageous over covalent interactions? Given that van der Waals forces are among the weakest molecular interactions, how can the gecko's feet stick so effectively?

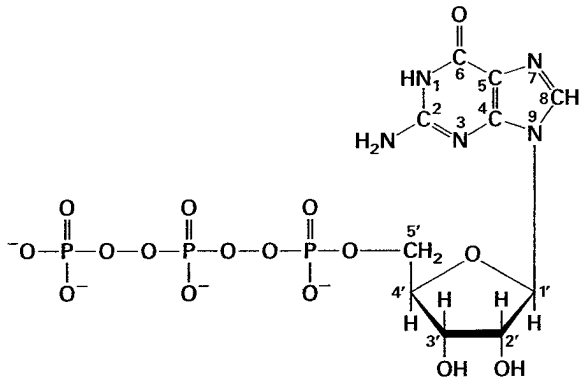
2. The K^+ channel is an example of a transmembrane protein (a protein that spans the phospholipid bilayer of the plasma membrane). What types of amino acids are likely to be found (a) lining the channel through which K^+ passes, (b) in contact with the hydrophobic core of the phospholipid bilayer containing fatty acyl groups, (c) in the cytosolic domain of the protein, and (d) in the extracellular domain of the protein?

3. V-M-Y-F-E-N: This is the single-letter amino acid abbreviation for a peptide. What is the net charge of this peptide at pH 7.0? An enzyme called a protein tyrosine kinase can attach phosphates to the hydroxyl groups of tyrosine. What is the net charge of the peptide at pH 7.0 after it has been phosphorylated by a tyrosine kinase? What is the likely source of phosphate utilized by the kinase for this reaction?

4. Disulfide bonds help to stabilize the three-dimensional structure of proteins. What amino acids are involved in the formation of disulfide bonds? Does the formation of a disulfide bond increase or decrease entropy (ΔS)?

5. In the 1960s, the drug thalidomide was prescribed to pregnant women to treat morning sickness. However, thalidomide caused severe limb defects in the children of some women who took the drug, and its use for morning sickness was discontinued. It is now known that thalidomide was administered as a mixture of two stereoisomeric compounds, one of which relieved morning sickness and the other of which was responsible for the birth defects. What are stereoisomers? Why might two such closely related compounds have such different physiologic effects?

6. Name the compound shown below.



Is this nucleotide a component of DNA, RNA, or both? Name one other function of this compound.

7. The chemical basis of blood-group specificity resides in the carbohydrates displayed on the surface of red blood cells. Carbohydrates have the potential for great structural diversity. Indeed, the structural complexity of the oligosaccharides that can be formed from four sugars is greater than that for oligopeptides from four amino acids. What properties of carbohydrates make this great structural diversity possible?

8. Ammonia (NH_3) is a weak base that under acidic conditions becomes protonated to the ammonium ion in the following reaction:

$$NH_3 + H^+ \rightarrow NH_4^+$$

NH_3 freely permeates biological membranes, including those of lysosomes. The lysosome is a subcellular organelle with a pH of about 4.5–5.0; the pH of cytoplasm is ~7.0. What is the effect on the pH of the fluid content of lysosomes

when cells are exposed to ammonia? *Note:* Protonated ammonia does not diffuse freely across membranes.

9. Consider the binding reaction L + R → LR, where L is a ligand and R is its receptor. When 1×10^{-3} M L is added to a solution containing 5×10^{-2} M R, 90% of the L binds to form LR. What is the K_{eq} of this reaction? How will the K_{eq} be affected by the addition of a protein that catalyzes this binding reaction? What is the K_d?

10. What is the ionization state of phosphoric acid in the cytoplasm? Why is phosphoric acid such a physiologically important compound?

11. The $\Delta G^{\circ\prime}$ for the reaction X + Y → XY is −1000 cal/mol. What is the ΔG at 25 °C (298 Kelvin) starting with 0.01 M each X, Y, and XY? Suggest two ways one could make this reaction energetically favorable.

12. According to health experts, saturated fatty acids, which come from animal fats, are a major factor contributing to coronary heart disease. What distinguishes a saturated fatty acid from an unsaturated fatty acid, and to what does the term *saturated* refer? Recently, trans unsaturated fatty acids, or trans fats, which raise total cholesterol levels in the body, have also been implicated in heart disease. How does the cis stereoisomer differ from the trans configuration, and what effect does the cis configuration have on the structure of the fatty acid chain?

13. Chemical modifications to amino acids contribute to the diversity and function of proteins. For instance, γ-carboxylation of specific amino acids is required to make some proteins biologically active. What particular amino acid undergoes this modification, and what is the biological relevance? Warfarin, a derivative of coumarin, which is present in many plants, inhibits γ-carboxylation of this amino acid and was used in the past as a rat poison. At present, it is also used clinically in humans. What patients might be prescribed warfarin and why?

References

Alberty, R. A., and R. J. Silbey. 2005. *Physical Chemistry,* 4th ed. Wiley.

Atkins, P., and J. de Paula. 2005. *The Elements of Physical Chemistry,* 4th ed. W. H. Freeman and Company.

Berg, J. M., J. L. Tymoczko, and L. Stryer. 2007. *Biochemistry,* 6th ed. W. H. Freeman and Company.

Cantor, P. R., and C. R. Schimmel. 1980. *Biophysical Chemistry.* W. H. Freeman and Company.

Davenport, H. W. 1974. *ABC of Acid-Base Chemistry,* 6th ed. University of Chicago Press.

Eisenberg, D., and D. Crothers. 1979. *Physical Chemistry with Applications to the Life Sciences.* Benjamin-Cummings.

Guyton, A. C., and J. E. Hall. 2000. *Textbook of Medical Physiology,* 10th ed. Saunders.

Hill, T. J. 1977. *Free Energy Transduction in Biology.* Academic Press.

Klotz, I. M. 1978. *Energy Changes in Biochemical Reactions.* Academic Press.

Murray, R. K., et al. 1999. *Harper's Biochemistry,* 25th ed. Lange.

Nicholls, D. G., and S. J. Ferguson. 1992. *Bioenergetics 2.* Academic Press.

Oxtoby, D., H. Gillis, and N. Nachtrieb. 2003. *Principles of Modern Chemistry,* 5th ed. Saunders.

Sharon, N. 1980. Carbohydrates. *Sci. Am.* 243(5):90–116.

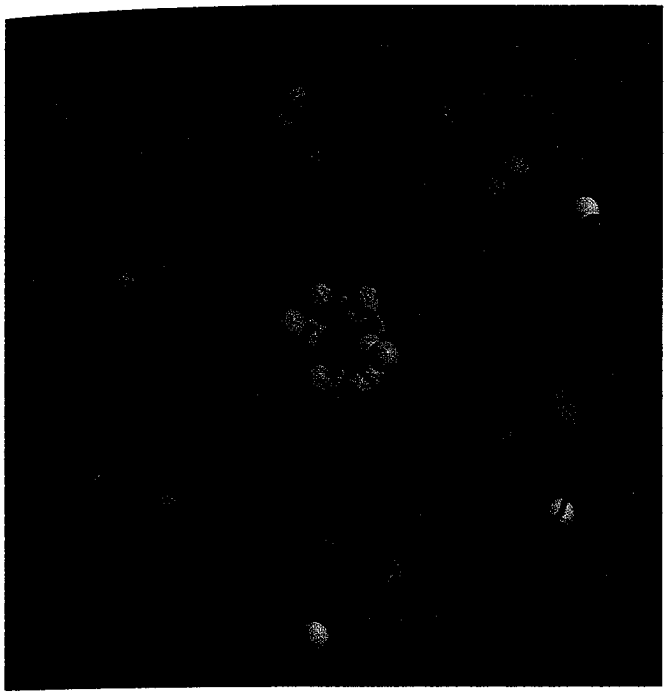
Tanford, C. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes,* 2d ed. Wiley.

Tinoco, I., K. Sauer, and J. Wang. 2001. *Physical Chemistry—Principles and Applications in Biological Sciences,* 4th ed. Prentice Hall.

Van Holde, K., W. Johnson, and P. Ho. 1998. *Principles of Physical Biochemistry.* Prentice Hall.

Voet, D., and J. Voet. 2004. *Biochemistry,* 3d ed. Wiley.

Wood, W. B., et al. 1981. *Biochemistry: A Problems Approach,* 2d ed. Benjamin-Cummings.



Ribbon diagram of a beta propeller domain from the human signaling protein Keapl. Ten water molecules (spheres) are bound to each of the six blades of the propeller. Many proteins are built from multiple, independently stable protein domains. [From L. J. Beamer, X. Li, C. A. Bottoms, and M. Hannink, 2005, *Acta Crystallogr. D: Biol. Crystallogr.* **61**(10):1335–1342.] Credit: Courtesy of Robert Huber, Martinsried.

CHAPTER

3

PROTEIN STRUCTURE AND FUNCTION

Proteins, which are polymers of amino acids, come in many sizes and shapes. Their three-dimensional diversity reflects underlying structural differences: principally variations in their lengths and amino acid sequences, and in some cases, differences also in the number of disulfide bonds or the attachment of small molecules or ions to their amino acid side chains. In general, the linear, unbranched polymer of amino acids composing any protein will fold into only one or a few closely related three-dimensional shapes—called **conformations.** The conformation of a protein together with the distinctive chemical properties of its amino acid side chains determines its function. As a consequence, proteins can perform a dazzling array of distinct functions inside and outside of cells that either are essential for life or provide selective evolutionary advantage to the cell or organism that contains them. It is, therefore, not surprising that characterizing the structures and activities of proteins is a fundamental prerequisite for understanding how cells work. Much of this textbook is devoted to examining how proteins act together to enable cells to live and function properly.

Many proteins can be grouped into just a few broad functional classes. *Structural proteins,* for example, determine the shapes of cells and their extracellular environments, and serve as guide wires or rails to direct the intracellular movement of molecules and organelles. They usually are formed by the assembly of multiple protein subunits into very large, long structures. *Scaffold proteins* bring other proteins together into ordered arrays to perform specific functions more efficiently than if those proteins were not assembled together. *Enzymes* are proteins that catalyze chemical reactions. *Membrane transport proteins* permit the flow of ions and molecules across cellular membranes. *Regulatory proteins* act as signals, sensors, and switches to control the activities of cells by altering the functions of other proteins and genes. These include *signaling proteins,* such as hormones and cell-surface receptors that transmit extracellular signals to the cell interior.

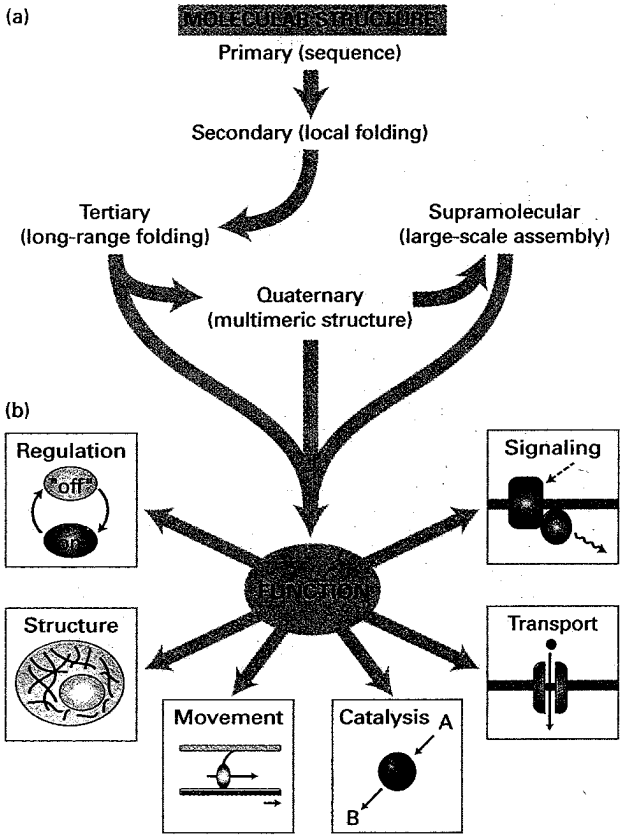
OUTLINE

Motor proteins are responsible for moving other proteins, organelles, cells—even whole organisms. Any one protein can be a member of more than one protein class, as is the case of some cell-surface signaling receptors that are both enzymes and regulator proteins because they transmit signals from outside to inside cells by catalyzing chemical reactions. To accomplish efficiently their diverse missions some proteins assemble into large complexes, often called *molecular machines*.

How do proteins mediate so many diverse functions? They do this by exploiting a few simple activities. Most fundamentally, proteins *bind*—to one another, to other macromolecules, such as DNA, and to small molecules and ions. In many cases such binding can induce a conformational change in the protein and thus influence its activity. Binding is based on molecular complementarity between a protein and its binding partner, as described in Chapter 2. A second key activity is enzymatic *catalysis*. Appropriate folding of a protein will place some amino acid side chains and carboxyl and amino groups of the backbone into positions that permit the catalysis of covalent bond rearrangements. A third activity involves *folding* into a channel or pore within a membrane through which molecules and ions flow. Although these are especially crucial protein activities, they are not the only ones. For example, fish that live in frigid waters—the Antarctic borchs and Arctic cods—have antifreeze proteins in their circulatory systems to prevent water crystallization at subzero temperatures.

A complete understanding of how proteins permit cells to live and thrive requires the identification and characterization of all the proteins used by a cell. In a sense, molecular cell biologists want to compile a complete protein 'parts list' and construct an all-inclusive "users manual" that describes how these proteins work. Compiling a comprehensive protein parts list has become feasible in recent years with the sequencing of entire **genomes**—complete sets of genes—of more and more organisms. From a computer analysis of genome sequences, researchers can deduce the number of amino acids and their sequence of most of the encoded proteins (Chapter 5). The term **proteome** was coined to refer to the entire protein complement of an organism. The human genome contains only 25,000 genes that encode proteins. However, variations in mRNA production (e.g., alternative splicing (Chapter 8)) and more than 100 types of protein modifications may generate hundreds of thousands of distinct human proteins. Remarkably, the human proteome is only about five times as large, comprising about 33,000 different proteins. By comparing protein sequences and structures of proteins of unknown function to those of known function, scientists can often deduce much about their functions. In the past, characterization of protein function by genetic, biochemical, or physiological methods often preceded the identification of particular proteins. In the modern genomic and proteomic era, a protein is usually identified prior to determining its function.

In this chapter, we begin our study of how the structure of a protein gives rise to its function, a theme that recurs throughout this book (Figure 3-1). The first section examines how chains of amino acid building blocks are arranged in a three-dimensional structural hierarchy. The next section discusses

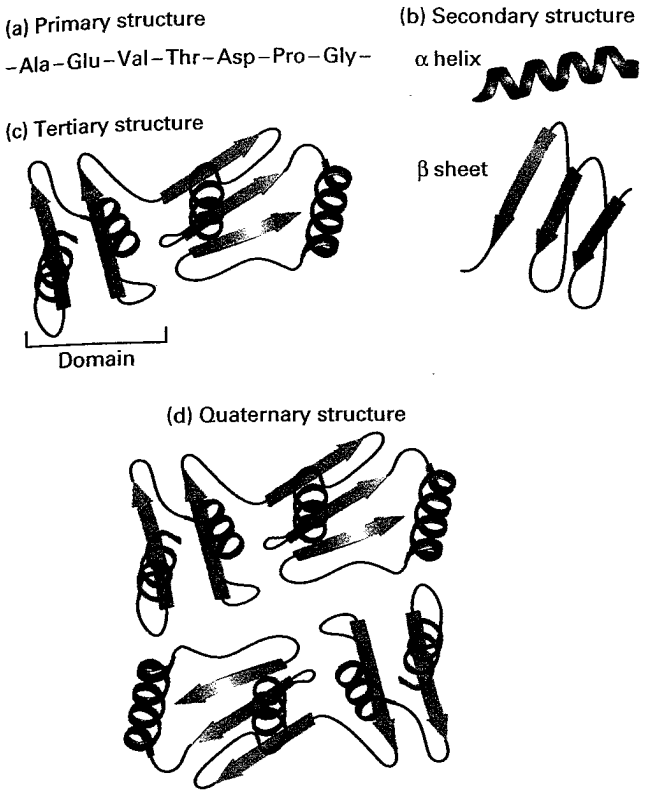


▲ **FIGURE 3-1 Overview of protein structure and function.** (a) Proteins are assembled according to a hierarchy of structures. A polypeptide's linear sequence of amino acids linked by peptide bonds (primary structure) folds into local helices or sheets (secondary structure) that pack into large (longer-range) complex three-dimensional structures (tertiary structure). Some individual polypeptides associate into multichain complexes (quaternary structure), which in some cases can be very large, consisting of tens to hundreds of subunits (supramolecular assemblies). (b) Protein function includes organization of the genome, other proteins, lipid bilayer membranes, and cytoplasm (structure); control of protein activity (regulation), monitoring of the environment and transmitting resultant information (signaling), flow of small molecules and ions across membranes (transport); catalysis of chemical reactions (via enzymes); and generation of force for movement (via motor proteins). These functions and others arise from specific binding interactions and conformational changes in the structure of a properly folded protein.

how proteins fold into these structures. We then turn to protein function, focusing on enzymes, the special class of proteins that catalyze chemical reactions. Various mechanisms that cells use to control the activities and life spans of proteins are covered in the next two sections. Next comes a section on commonly used techniques in the biologist's tool kit for isolating proteins and characterizing their properties. The chapter concludes with a discussion of the burgeoning field of proteomics.

3.1 Hierarchical Structure of Proteins

A protein chain folds into a distinct three-dimensional shape that is stabilized by noncovalent interactions between regions in the linear sequence of amino acids. A key concept in

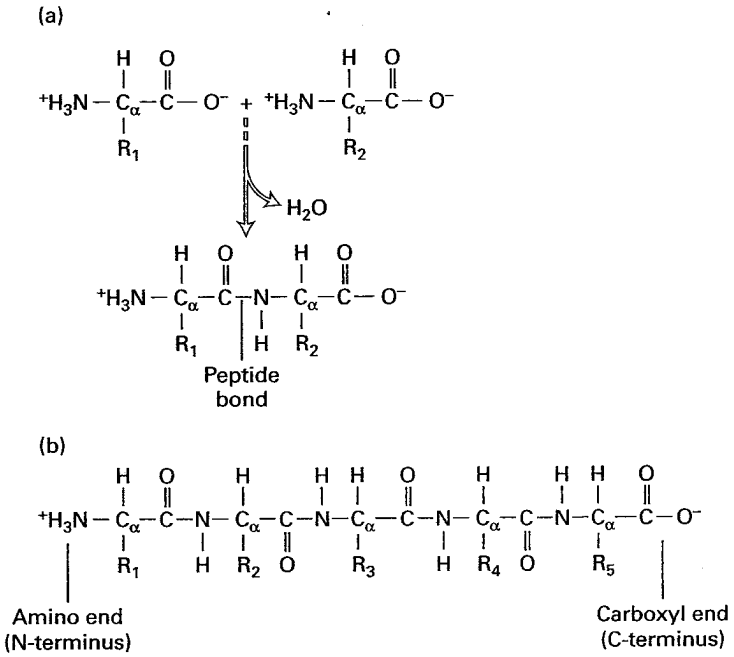


▲ **FIGURE 3-2 Four levels of protein hierarchy.** (a) The linear sequence of amino acids linked together by peptide bonds is the primary structure. (b) Folding of the polypeptide chain into local α helices or β sheets represents secondary structure. (c) Secondary structural elements together with various loops and turns in a single polypeptide chain pack into a larger independently stable structure, which may include distinct domains; this is tertiary structure. (d) Some individual polypeptides with their own tertiary structures can associate into a quaternary structure defining a multichain complex.

understanding how proteins work is that *function is derived from three-dimensional structure, and three-dimensional structure, which is determined primarily by noncovalent interactions between regions in the linear sequence of amino acids, is specified by amino acid sequence.* Indeed, principles relating biological structure and function initially were formulated by the biologists Johann von Goethe (1749–1832), Ernst Haeckel (1834–1919), and D'Arcy Thompson (1860–1948). They greatly influenced the school of "organic" architecture pioneered in the early twentieth century that is epitomized by the dicta "form follows function" (Louis Sullivan) and "form is function" (Frank Lloyd Wright). Here, we consider the architecture of proteins at four levels of organization: primary, secondary, tertiary, and quaternary (Figure 3-2).

The Primary Structure of a Protein Is Its Linear Arrangement of Amino Acids

As discussed in Chapter 2, proteins are constructed by the polymerization of 20 different types of amino acids. Individual amino acids are linked together in linear, unbranched chains by covalent amide bonds, called **peptide bonds,** with occasional disulfide bonds covalently linking side chains together. Peptide bond formation between the amino group of one amino acid and the carboxyl group of another results in the net release of a water molecule (dehydration) (Figure 3-3a). The repeated amide N, α carbon (C_α), carbonyl C and oxygen atoms of each amino acid residue form the backbone of a protein molecule from which the various side-chain groups project (Figure 3-3b, c). As a consequence of the peptide linkage, the backbone exhibits directionality because all the amino groups are located on the same side of the C_α atoms. Thus one end of a protein has a free (unlinked) amino group (the *N-terminus*), and the other end has a free carboxyl group (the *C-terminus*). The sequence of a protein chain is conventionally written with its



▲ **FIGURE 3-3 Structure of a polypeptide.** (a) Individual amino acids are linked together by peptide bonds, which form via reactions that result in a loss of water (dehydration). R_1, R_2, etc., represent the side chains ("R groups") of amino acids. (b) Linear polymers of peptide bond–linked amino acids are called *polypeptides,* which have a free amino end (N-terminus) and a free carboxyl end (C-terminus). (c) A ball-and-stick model shows peptide bonds (yellow) linking the amino nitrogen atom (blue) of one amino acid (aa) with the carbonyl carbon atom (gray) of an adjacent one in the chain. The R groups (green) extend from the α carbon atoms (black) of the amino acids. These side chains largely determine the distinct properties of individual proteins.

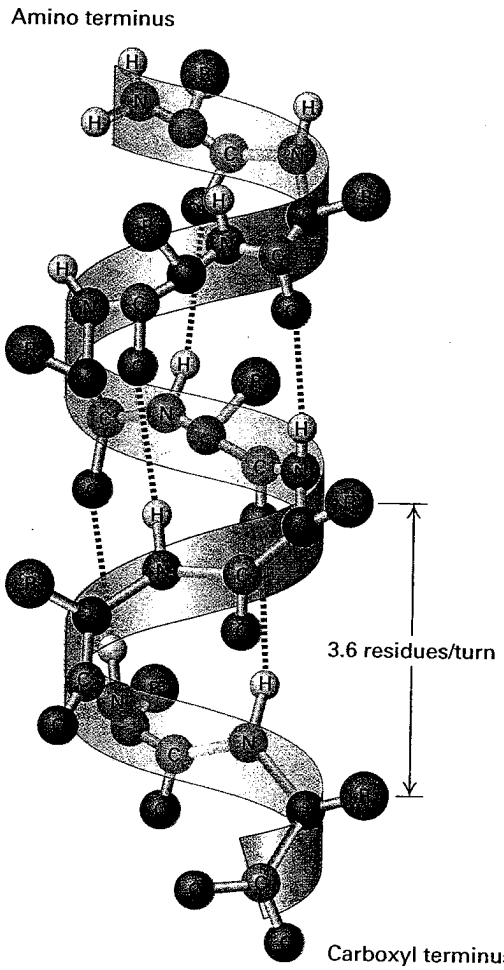
N-terminal amino acid on the left and its C-terminal amino acid on the right, and the amino acids are numbered sequentially starting from the amino terminus (number 1).

The **primary structure** of a protein is simply the linear arrangement, or sequence, of the amino acid residues that compose it. Many terms are used to denote the chains formed by the polymerization of amino acids. A short chain of amino acids linked by peptide bonds and having a defined sequence is called an **oligopeptide**, or just **peptide**; longer chains are referred to as **polypeptides**. Peptides generally contain fewer than 20–30 amino acid residues, whereas polypeptides are often 200–500 residues long. The longest protein described to date is the muscle protein titin with 26,926 residues. We generally reserve the term **protein** for a polypeptide (or complex of polypeptides) that has a well-defined three-dimensional structure. It is implied that proteins and peptides are the natural products of a cell.

The size of a protein or a polypeptide is reported as its mass in **daltons** (a dalton is 1 atomic mass unit) or as its molecular weight (MW), which is a dimensionless number. For example, a 10,000-MW protein has a mass of 10,000 daltons (Da), or 10 kilodaltons (kDa). In the penultimate section of this chapter, we will consider different methods for measuring the sizes and other physical characteristics of proteins. The known and predicted proteins encoded by the yeast genome have an average molecular weight of 52,728 and contain, on average, 466 amino acid residues. The average molecular weight of amino acids in proteins is 113, taking into account their average relative abundances. This value can be used to estimate the number of residues in a protein from its molecular weight or, conversely, its molecular weight from the number of residues.

Secondary Structures Are the Core Elements of Protein Architecture

The second level in the hierarchy of protein structure is **secondary structure.** Secondary structures are stable spatial arrangements of segments of a polypeptide chain held together by hydrogen bonds between backbone amide and carbonyl groups and often involving repeating structural patterns. A single polypeptide may contain multiple types of secondary structure in various portions of the chain, depending on its sequence. The principal secondary structures are the **alpha (α) helix,** the **beta (β) sheet,** and a short U-shaped **beta (β) turn.** Portions of the polypeptide that don't form these structures, but nevertheless have a well-defined, stable shape, are said to have an *irregular* structure. The term *random coil* applies to highly flexible portions of a polypeptide chain that have no fixed three-dimensional structure. In an average protein, 60 percent of the polypeptide chain exists as α helices and β sheets; the remainder of the molecule is in coils and turns. Thus, α helices and β sheets are the major internal supportive elements in most proteins. In this section, we explore the shapes of secondary structures and the forces that favor their formation. In later sections, we examine how linear arrays of secondary structure fold together into larger, more complex arrangements called tertiary structure.



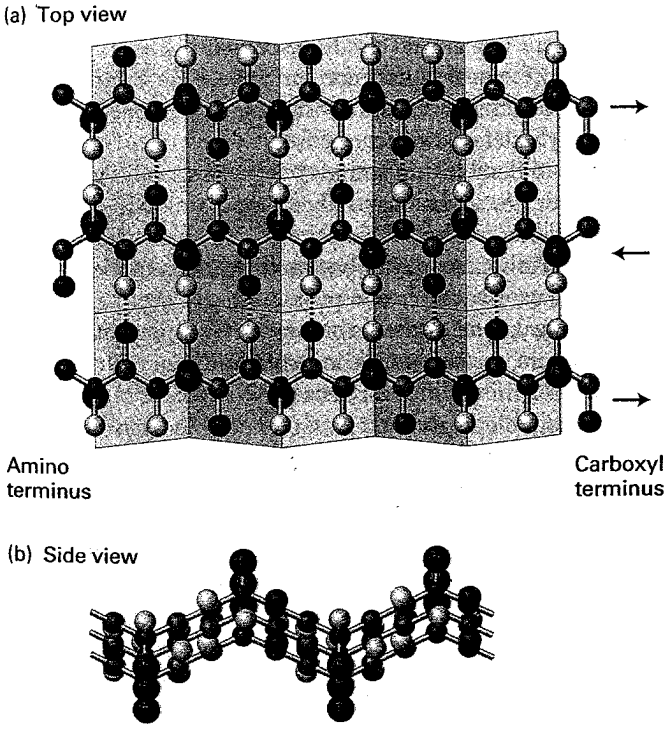
▲ **FIGURE 3-4 The α helix, a common secondary structure in proteins.** The polypeptide backbone (seen as a ribbon) is folded into a spiral that is held in place by hydrogen bonds between backbone oxygen and hydrogen atoms. Only hydrogens involved in bonding are shown. The outer surface of the helix is covered by the side-chain R groups (green).

The α Helix In a polypeptide segment folded into an α helix, the backbone forms a spiral structure in which the carbonyl oxygen atom of each peptide bond is hydrogen-bonded to the amide hydrogen atom of the amino acid four residues farther along the chain (in the direction of the C-terminus) (Figure 3-4). Within an α helix, all the backbone amino and carboxyl groups are hydrogen-bonded to one another, except at the very beginning and end of the helix. This periodic arrangement of bonds confers an amino-to-carboxy-terminal directionality on the helix because all the hydrogen bond acceptors (e.g., the carbonyl groups) have the same orientation (pointing in the downward direction in Figure 3-4) and results in a structure in which there is a complete turn of the spiral every 3.6 residues. An α helix 36 amino acids long has 10 turns of the helix and is 5.4 nm long (0.54 nm/turn).

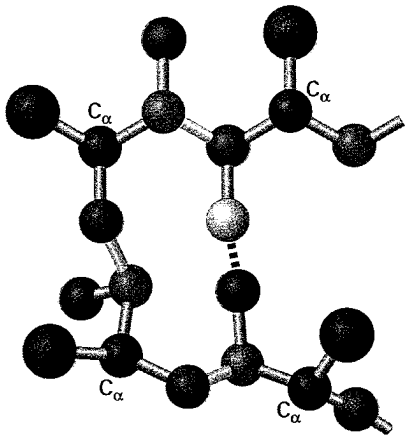
The stable arrangement of hydrogen-bonded amino acids in the α helix holds the backbone in a straight, rod-like cylinder from which the side chains point outward. The relative hydrophobic or hydrophilic quality of a particular helix within a protein is determined entirely by the

characteristics of the side chains, because all the polar amino and carboxyl groups of the peptide backbone are engaged in hydrogen bonding with one another in the helix. In water-soluble proteins, the hydrophilic helices tend to be found on the outside surfaces, where they can interact with the aqueous environment, whereas hydrophobic helices tend to be buried within the core of the folded protein. The amino acid proline is usually not found in α helices, because the covalent bonding of its amino group with a carbon in the side chain prevents its participation in stabilizing the backbone through normal hydrogen bonding. While the classic α helix is the most intrinsically stable, and most common helical form in proteins, there are variations, such as more tightly or loosely twisted helices. For example, in a specialized helix called a coiled coil (described several sections farther on), the helix is more tightly wound (3.5 residues and 0.51 nm per turn).

The β Sheet Another type of secondary structure, the β sheet, consists of laterally packed β strands. Each β strand is a short (5- to 8-residue), nearly fully extended polypeptide segment. Unlike in the α helix (where hydrogen bonding between the amino and carboxyl groups in the backbone occurs between nearly adjacent residues), hydrogen bonding in the β sheet occurs between backbone atoms in separate, but adjacent, β strands (Figure 3-5a). These distinct β strands may be either within a single polypeptide chain, with short or long loops between the β strand segments, or on different polypeptide chains. Figure 3-5b shows how two or more β strands align into adjacent rows, forming a nearly two-dimensional β pleated sheet (or simply *pleated sheet*), in which hydrogen bonds within the plane of the sheet hold the β strands together as the side chains stick out above and below the plane. Like α helices, β strands have a directionality defined by the orientation of the peptide bond. Therefore, in a pleated sheet, adjacent β strands can be oriented in the same (parallel) or opposite (antiparallel) directions with respect to each other. In some proteins, β sheets form the floor of a binding pocket or a hydrophobic core; in other proteins embedded in membranes the β sheets curve around and form a hydrophilic central pore through which ions and small molecules may flow (Chapter 11).



▲ **FIGURE 3-5 The β sheet, another common secondary structure in proteins.** (a) Top view of a simple three-stranded β sheet with antiparallel β strands. The stabilizing hydrogen bonds between the β strands are indicated by green dashed lines. (b) Side view of a β sheet. The projection of the R groups (green) above and below the plane of the sheet is obvious in this view. The fixed bond angles in the polypeptide backbone produce a pleated contour.



▲ **FIGURE 3-6 Structure of a β turn.** Composed of four residues, β turns reverse the direction of a polypeptide chain (≈180° U-turn). The C_α carbons of the first and fourth residues are usually <0.7 nm apart, and those residues are often linked by a hydrogen bond. β turns facilitate the folding of long polypeptides into compact structures.

β Turns Composed of four residues, β turns are located on the surface of a protein, forming sharp bends that reverse the direction of the polypeptide backbone, often toward the protein's interior. These short, U-shaped secondary structures are often stabilized by a hydrogen bond between their end residues (Figure 3-6). Glycine and proline are commonly present in turns. The lack of a large side chain in glycine and the presence of a built-in bend in proline allow the polypeptide backbone to fold into a tight U shape. β turns help large proteins to fold into highly compact structures. There are six types of well-defined turns, their detailed structures depending on the arrangement of H-bonding interactions. A polypeptide backbone also may contain longer bends, or loops. In contrast with tight β turns, which exhibit just a few well-defined conformations, longer loops can have many different conformations.

Overall Folding of a Polypeptide Chain Yields Its Tertiary Structure

Tertiary structure refers to the overall conformation of a polypeptide chain—that is, the three-dimensional arrangement of all its amino acid residues. In contrast with secondary